A STUDY ON GENETIC ALGORITHM BASED HYBRID SOFTCOMPUTING MODEL FOR BENIGNANCY/MALIGNANCY DETECTION OF MASSES USING DIGITAL MAMMOGRAM

MAHUA BHATTACHARYA*, NAVEEN SHARMA, VAIBHAV GOYAL and SAGAR BHATIA
Indian Institute of Information Technology & Management
Morena Link Road, Gwalior-474003, India
*bmahua@hotmail.com
mb@iiitm.ac.in

ARPITA DAS
Institute of Radio Physics & Electronics
University of Calcutta, 92, A.P.C. Road
Kolkata-700009

Received 15 April 2010
Revised 24 January 2011

In present works authors have developed a computerized classification procedure for tumor mass in breasts using digital mammogram. The process implements genetic algorithm and hybrid neuro-fuzzy approaches to classify tumor masses into benign and malignant group in order to assist the physicians for treatment planning. The classification process is based on accurate analysis of shape and margin of tumor mass appearing in breast. The shape features using Fourier descriptors introduce a large number of feature vectors. Thus, to classify different boundaries, a standard multilayer preceptor needs large number of inputs. Simultaneously, to train the network, a large number of training cycles and huge memory are also required. It is obvious that a complicated structure invites the problem of over learning and misclassification. In proposed methodology genetic algorithm (GA) has been used for the searching of effective input feature vectors. Adaptive neuro-fuzzy model has been used for final classification of different boundaries of tumor masses. The proposed technique is an innovative soft computing approach that removes the limitation of conventional neural networks and indicates a promising direction of adaptation in a changing environment. The classification system utilizes a Euclidean distance function to detect the belongingness of masses in benign and in malignant classes along with degree of benignancy/malignancy. Presently 200 digitized mammograms from MIAS and other databases have been considered for the experiment and which have shown an average of approximately 86% correct classification as compared with clinical data with a highest rate of 88.9%.

Keywords: Breast cancer diagnostic system; benignancy and malignancy; Fourier descriptors; Fuzzy c-means clustering; genetic algorithm; adaptive neuro-fuzzy; digital mammography.

*Corresponding author.
1. Introduction

Every year, millions of women develop new cases of breast cancer around the world. The detection procedures are based on clinical examination, breast imaging like mammography, MRI and core biopsy. An early detection of breast cancer gives the patient a good chance of survival. Mammographic images show signs of obstruction and many direct and indirect radiographic signs due to space occupying lesions in the tissue region of breast.\textsuperscript{1–3} The obstruction causes the dilation of mammary vein. The conventional method is the localization of the lesion by needle\textsuperscript{2} which is painful and costly. Since the failure to detect any abnormal lesion at an early stage may lead to disastrous consequences, improvement of mammographic image quality is essential for breast cancer screening. Many researchers have investigated the finer details of micro-calcifications. Verma \textit{et al.}\textsuperscript{38} has reported the classification of micro-calcification in benign and malignant group by finding best feature sets and suitable NN architecture. Researchers also depict the computer aided diagnostics and detection procedures of masses/calcification using digital mammogram.\textsuperscript{4–13,37,39–42,44,45}

The aim of most studies of computer aided detection of breast cancer is to help the physicians to determine the treatment procedure for breast cancer identification. Masses in digital mammograms appear as 3-dimensional lesions and represent a localizing sign of breast cancer. They are described by their shape, margin, textural characteristics, and effect on surrounding tissues. The margin of a mass is one of the most important criteria in determining whether the mass is likely to be benign or malignant. Benign tumors\textsuperscript{14} have clearly defined round or round to oval shaped edges and contain cells that look healthy, just like normal cells. They tend to grow slowly. These tumors may cause harm if they start to interfere with normal breast functions, whereas malignant masses spread quickly into surrounding breast tissue. They have irregular borders, multiple protrusions and made up of abnormally shaped cells.\textsuperscript{8,14–16} The present work is a continuation of our earlier work based on theory of shape related to gradation of benignancy/malignancy of tumor in tissue region. We have suggested \textit{shape similarity measure}\textsuperscript{8,16,17} to find out the prognosis of diseases where the idea of shape similarity measure has been implemented by computing distance function $D$ between the contours of tumor lesions and the model. Bruce \textit{et al.}\textsuperscript{18} reported the results of applying multi-resolution techniques to the problem of tumor mass classification. Kupinski \textit{et al.}\textsuperscript{19} developed two seeded lesion segmentation techniques, one based on a single feature called the radial gradient index (RGI) and the other based on simple probabilistic models. Mudigonda \textit{et al.}\textsuperscript{20} proposed an algorithm to segment masses by establishing intensity links, and to analyze oriented flow-like textural information in the ribbons of pixels across the margins of masses to determine if the segmented regions are true mass regions or false positives. Sahiner \textit{et al.}\textsuperscript{21} developed an automated, three-stage segmentation method including clustering, active contour, and detection stages. The method also carried out classification of the segmented masses as benign and malignant using speculation measures and morphological features. Morphologic spectrum of
mammographic masses is shown in Fig. 1. An efficient classification method of suspected region in digital mammogram has been presented by Verma et al.\textsuperscript{41} which presents a novel soft cluster neural network technique for the classification of suspicious areas in digital mammograms. The technique introduces the concept of soft clusters within a neural network layer and combines them with least squares for optimizing neural network weights.

In the present paper, authors have described a classification process of tumor mass in breast using genetic algorithm and adaptive neuro-fuzzy techniques considering contour-based features. To describe the contour/margin of masses very precisely, we introduce Fourier descriptors as shape-based features.\textsuperscript{22} In the next stage, classifier has been designed using neuro-fuzzy techniques\textsuperscript{5, 22–28, 36} to classify the tumor growths into benignancy and malignancy. This method also incorporates the automated false positive reduction of mass boundaries.

2. Morphology of Tumor Mass
Masses in mammogram are compact areas that appear brighter than the tissue in which they are embedded because of higher attenuation of X-rays.\textsuperscript{26} However, when a mass is buried in dense fibro-glandular tissue, it is very difficult to identify. The primary features that indicate malignancy are related to the mass shape, margin and texture. Using mass morphology, we have used shape or contour as features to classify malignant and benign masses.

3. Proposed Method for Classification of Tumor Masses
Tumor masses have been extracted from surrounding normal breast tissues by fuzzy based segmentation technique. Significant shape-based boundary features are selected for the purpose of classification. The genetic algorithm GA based optimization technique has been implemented for reduction of feature space. Finally, features are fed to an adaptive neuro-fuzzy classifier for a decision of the space occupying tumor lesions whether the masses belong to benign group or malignant group. The overview of the proposed method is presented in Fig. 2.
3.1. Segmentation of tumor mass

In the proposed technique, fuzzy $c$-means clustering algorithm has been used for intensity based segmentation of tumor mass. Total number of fuzzy cluster centers chosen is three as shown in Fig. 3. Cluster center $A$ represents the healthy breast tissue. Second cluster $B$ represents the false presence of mass region (due to dense fibroglandular tissue) and $C$ represents the actual mass region.
A Study on Genetic Algorithm Based Hybrid Softcomputing Model

The ultimate Fuzzy partition membership functions have been shown in Fig. 3, which show that there is an overlapping between the membership functions \( A, B \) and \( C \). Presently, proper decision has been made on the basis, which is described below:

If the possibility of a breast region regarding its belongingness to the calcification part is greater than 50% that is, the membership value of curve \( C > 0.5 \), decision may be taken that the particular region is belonging within the calcified lesion.

According to the decision rule, the shaded region in Fig. 3 indicates the region of interest (ROI).

\textbf{Algorithm:} Let \( X = \{x_1, x_2, \ldots, x_n\} \) be a set of given data. A fuzzy \( c \)-partition of \( X \) is a family of fuzzy subsets of \( X \), denoted by \( P = \{A_1, A_2, \ldots, A_c\} \), which satisfies

\[
\sum_{i=1}^{c} A_i(x_k) = 1.
\] (1)

The performance index of a fuzzy partition \( P, J_m(P) \), is defined in terms of the cluster centers by the formula

\[
J_m(A, v_1, \ldots, v_c) = \sum_{k=1}^{n} \sum_{i=1}^{c} [A_i(x_k)]^m \|x_k - v_i\|^2,
\] (2)

where \( \|x_k - v_i\|^2 \) represents the distance between \( x_k \) and \( v_i \) (\( v_i \) is the cluster centers). Clearly, the smaller the value of \( J_m(P) \), the better the fuzzy partition \( P \). Thus, the goal of fuzzy \( c \)-means clustering method is to find a fuzzy partition \( P \) that minimizes the performance index \( J_m(P) \), which offers

\[
v_i = \frac{\sum_{k=1}^{N} [A_i(x_k)]^m x_k}{\sum_{k=1}^{N} [A_i(x_k)]^m}
\] (3)

and

\[
A_i(x_k) = \frac{1}{\sum_{j=1}^{C} \left( \frac{d_{ik}}{d_{jk}}\right)^{m-1}}
\] (4)

Thus fuzzy \( c \)-means clustering algorithm is an iterated procedure.

3.2. \textit{Extraction of boundary as feature using Fourier descriptors}

Feature selection is the choice of descriptors in a particular application. Presently, we introduce \textit{Fourier descriptors} as the boundary features having the information about the shape and margin of the segmented masses. Among different tumor masses, the stellate mass contains much more high frequency components than the nodular shape masses while the round shaped, smooth boundary masses only contain low frequency components.

\textbf{Algorithm:} Let us consider a figure that describes \( k \)-points digital boundary in the \( x-y \) plane, starting at an arbitrary point \((x_0, y_0)\) to the coordinate pairs
\((x_1, y_1), (x_2, y_2), \ldots, (x_{k-1}, y_{k-1})\) along the boundary. These co-ordinates are represented by the form \(x(k) = x_k\) and \(y(k) = y_k\).

Thus the boundary can be represented as

\[ s(k) = x(k), y(k) \quad \text{for} \quad k = 0, 1, 2, \ldots, k - 1. \]  

Each co-ordinate pair can be treated as a complex number so that

\[ s(k) = x(k) + j*y(k) \quad \text{for} \quad k = 0, 1, 2, \ldots, k - 1. \]  

The \(x\)-axis is treated as the real axis and \(y\)-axis as the imaginary one. The Discrete Fourier Transform (DFT) of \(s(k)\) is given below

\[ a(u) = \frac{1}{K} \times \sum_{k=0}^{K-1} S(k) \times e^{-j2\pi uk/K} \quad \text{for} \quad u = 0, 1, 2, \ldots, K - 1. \]  

The complex coefficient \(a(u)\) is the Fourier descriptor of the edge points along the boundary. The compactness of a particular shape is another important feature used for boundary descriptor. It is defined as \((\text{perimeter})^2/\text{area}\). Compactness is a dimensionless quantity and is minimal (nearly equal to one) for round shaped region. Compactness is also insensitive to the orientation of the images. In the present work, the computation of compactness has been considered as an important shape feature.

### 3.3. Introduction to genetic algorithm for reduction of feature subspace

The feature selection problem (FSP) is an important issue in machine learning which consists of finding a feature subset of input training as well as test patterns that enable one to describe all the information required for classifying them. The boundary or margin detection of masses based on Fourier descriptor introduces large number of feature vectors. Thus classifying different boundaries a standard classifier needs a large number of inputs which may introduce the problem of over learning and chances of misclassification. In proposed method genetic algorithm (GA)\(^{29-33}\) has been used to search two significant shape descriptors which are able to represent the particular class of masses. Genetic algorithms (GAs) are search algorithms based on the mechanics of natural selection and natural genetics. Basically, GA uses three operators --- selection (or reproduction), crossover and mutation to achieve the goal of evolution. Genetic algorithms efficiently exploit the information to speculate on new search points with expected improved performance. (This method allows making escapes from local optima and the chromosomes will approach the global optimum.) The steps of GAs are summarized in Fig. 4.

In the present work, different image boundaries of tumor masses have been recognized on the basis of Fourier descriptors, which play the role of payoff values (objective function) associated with individual strings. We encode the magnitude of Fourier descriptors as binary strings and optimize these strings to achieve the best two features that are responsible for representing the class of the objects.
3.3.1. Selection/reproduction

The single piece of information that a GA receives from the environment is a scalar indicator (called fitness function) that evaluates the performance of each chromosome. The GA then uses that evaluation to bias the selection of chromosomes so that those with better scores tend to reproduce more often than those with worse scores.

3.3.2. Crossover

Reproduction alone does nothing to promote exploration of new search space, since no new points are searched. If we only copy the old structures without change, we...
will never achieve anything new when crossover is required. Crossover is a structured yet randomized information exchange between strings. Crossover creates new structures with a minimum of disruption to the allocated strategy dictated by reproduction. In a standard GA, a population of chromosomes undergo the crossover operator to generate offspring—with greater chance of selection being given to the chromosomes with higher fitness. The combination of reproduction and crossover give genetic algorithms much power.

3.3.3. Mutation

It is the well-known ‘diversity’ concept of the beauty-of-natural phenomenon that helps a population to survive under changing environmental condition. Mutation makes a small random change to a single solution. If the population does not contain all encoded information needed to survive, only crossover can produce a satisfactory solution. A mutation operator can prevent any single bit from converging to a value throughout the entire population and more importantly, it can prevent the population from converging and stagnating at any local optima.

3.3.4. Computational approach using GA

Suppose at a given time step $t$, there are $m$ examples of schema (chromosome) $H$ within the population $A(t)$, where we may write,

$$m = m(H, t).$$

During reproduction, a chromosome will mate according to its fitness, given by

$$p_i = f_i / \sum f.$$

where $f_i$—fitness value of $i$th chromosome

After picking a non-overlapping population of size ($n$) with replacement from the population $A(t)$, we obtain $m(H, t + 1)$ examples in the population at time $t + 1$. Hence

$$(H, t + 1) = m(H, t)^* n^* f(H) / \sum f.$$ (10)

where $f(H)$ is the average fitness of the population representing the chromosome $H$ at time $t$.

If we write the average fitness of the entire population as

$$f' = \sum f / n$$

Then

$$m(H, t + 1) = m(H, t)^* f(H) / f'$$ (12)

In words chromosome with fitness value above the population average will receive an increasing number of samples in the next generation while chromosome
with fitness value below the population average will receive a decreasing number of samples. Thus, the effect of reproduction on the number of chromosome is clear that above average chromosomes grow and below-average chromosomes die off. On the other hand, reproduction alone does nothing to promote the exploration of new regions of the search space. Since no new points are searched, we copy old structures without change and crossover comes into play and results in an exponential decrease or increase of chromosomes in a population.

Let us consider two different chromosomes $H_1$ and $H_2$ with length ($l$) 7, exchanging information between them.

$$H_1 = *1* | ***0$$
$$H_2 = ***| 10***$$

The asterisk is a “don’t care” symbol which matches either 0 or 1 at a particular position.

Here the crossing site has been marked with separator symbol $|$. If the crossover site is selected uniformly at random among the $l - 1 = 7 - 1 = 6$ possible sites, then chromosome $H_1$ is destroyed with probability

$$p_d = \delta(H_1)/(l - 1) = 5/6.$$  \hspace{1cm} (13)

Or survival probability $1 - p_d = 1/6$ (the defining length $\delta(H_1)$ of chromosome $H_1$ is the first and last specific string position). Similarly, the chromosome $H_2$ will be destroyed with probability,

$$p_d = \delta(H_2)/(l - 1) = 1/6.$$  \hspace{1cm} (14)

The survival probability of a chromosome under simple crossover is

$$p_s = 1 - \delta(H)/(l - 1).$$  \hspace{1cm} (15)

Since the chromosome is likely to be disrupted whenever a crossover site within the length $l - 1$.

If the crossover is itself performed by uniformly random way, say with probability $p_c$ at particular mating, the survival probability may be given by the expression

$$p_s \geq 1 - p_s \cdot \delta(H)/(l - 1).$$  \hspace{1cm} (16)

Considering the combined effect of reproduction and crossover, we obtain the estimate:

$$m(H, t + 1) \geq m(H, t) \cdot f(H)/f'[1 - p_s \cdot \delta(H)/(l - 1)].$$  \hspace{1cm} (17)

Both crossover and reproduction chromosome growing or decaying depend on two factors: whether the chromosome is above or below the population average and whether the chromosome has relatively short or long defining length. Those chromosomes with both above-average fitness and short defining length are growing at an exponential rate.
The last operator to consider is mutation. Mutation is the random alteration of a single position with probability $p_m$. Therefore, since a single bit survives with probability $(1 - p_m)$, and since each of the mutations is statistically independent, a particular chromosome survives when each of the $o(H)$ (order of chromosome denoted as $o(H) \rightarrow$ states the number of fixed positions) fixed positions within the chromosome survives. Multiplying the survival probability $(1 - p_m)$ by itself $o(H)$ times, we obtain the probability of surviving mutation, $(1 - p_m)^{o(H)}$. For small values of $p_m$ ($p_m \ll 1$), the chromosome survival probability may be approximated by $1 - o(H) * p_m$.

We therefore conclude that a particular chromosome $H$ receives an expected number of copies in the next generation under reproduction, crossover and mutation as given by the following equation,

$$m(H, t + 1) \geq m(H, t) \ast f(H)/f'[1 - p_s \ast \delta(H)/(l - 1) - (H) \ast p_m] \quad (18)$$

Thus with the addition of mutation, our final conclusion on the survival of chromosome is: low order, above average chromosomes receives exponentially increasing trials in subsequent generations.

In the present study, image boundaries of different tumor masses have been recognized on the basis of Fourier descriptors which play the role of payoff values (objective function) associated with individual binary coded chromosomes or strings. Genetic algorithm (GA) has been used to search the two most significant shape descriptors that are capable to represent the particular class of masses.

3.4. Classification of significant features using ANFIS network

The proposed method uses adaptive neuro-fuzzy network\textsuperscript{22–24,27,28,43} for classification of features into benign and malignant groups. The decision making by the computerized expert diagnosis system is closer to reality. The adaptive neuro-fuzzy model removes the limitations of conventional back propagation neural networks. To adapt the network with ever-changing environments, hybrid-learning rule has been used. ANFIS topology and the learning method that has been used for this neuro-fuzzy network are presented in this section. Both neural network and fuzzy logic are model-free estimators and share the common ability to deal with uncertainties and noise. Both of them encode the information in parallel and distribute architectures in a numerical framework. Hence, it is possible to convert fuzzy logic architecture to a neural network and vice versa. This makes it possible to combine the advantages of neural network and fuzzy logic. The ANFIS is composed of two parts. The first is the antecedent part and the second is the conclusion part, which are connected to each other by fuzzy rules based in network form. The ANFIS structure shown in, is a five-layer network. Figure 5 illustrates the structure of the
adaptive neuro-fuzzy architecture for boundary detection of tumor masses where nodes of the same layer have similar functions as described below.

**Layer 1:** Every node I in this layer is an adaptive node with a node function

$$O_{1,i} = \mu_{A_i}(x), \text{ for } i = 1, 2$$

$$O_{1,i} = \mu_{B_i}(y),$$

where $x$ (or $y$) is the input to node $i$ and $A_i$ (or $B_i$) is a linguistic label (such as large or small) associated with this node. In other words $O_{1,i}$ is the membership grade of fuzzy set $A$ (or $B$) and the membership function for $A$ can be any appropriate parameterized membership function, such as generalized bell function:

$$\mu_A = \frac{1}{1 + \frac{|x - c_i|}{a_i b_i}}$$

(19)

where $\{a_i, b_i, c_i\}$ is the parameter set. As the values of these parameters change, the bell-shaped function varies accordingly. Parameters of this layer are referred to as *premise parameters*.

Fig. 5. Adaptive neuro-fuzzy model for final classification.
Layer 2: Every node in this layer is a fixed node labeled Π, whose output is the product of all the incoming signals

\[ O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad \text{for } i = 1, 2. \]  

(20)

In general, any T-norm operator that performs fuzzy AND can be used as the node function in this layer.

Layer 3: Every node in this layer is a fixed node labeled N. The ith node calculates the ratio of the rule’s firing strength to the sum of all rules’ firing strengths:

\[ O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 - w_2}. \]  

(21)

For convenience, outputs of this layer are called normalized firing strengths.

Layer 4: Every node i in this layer is an adaptive node with a node function

\[ O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i). \]  

(22)

where \( w_i \) is a normalized firing strength from layer 3 and \( \{ p_i, q_i, r_i \} \) is the parameter set of this node. Parameters of this layer are referred to as consequent parameters.

Layer 5: The single node in this layer is fixed node labeled \( \sum \), which computes the overall output as the summation of all incoming signals:

\[ O_{5,i} = \sum \overline{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i}. \]  

(23)

3.4.1. Hybrid learning rule: training algorithm of adaptive neuro-fuzzy model

Hybrid learning rule combines the steepest decent method and least-squares estimator for the fast identification of parameters in adaptive neuro-fuzzy model. Hybrid learning is to be applied in a batch mode where each epoch is composed of a forward pass and a backward pass. In the forward pass after an input vector is presented, node outputs go forward until layer 4 and consequent parameters are identified by the least squares method. In the backward pass, the error signals propagate backward and the premise parameters are updated by the gradient descent. The hybrid method converges much faster than a single approach since it reduces the search space dimensions of the original pure back propagation learning. Also, the hybridization of neuro-fuzzy approaches is robust and adaptive even in noisy, uncertain environments. Table 1 summarizes the activities in each pass.

Limitation of adaptive neuro-fuzzy model is that the architecture is learned well only when number of inputs is very small (three to four only). In the present paper we set the input feature vector size to three only, and there are two bell-shaped membership functions assigned for each input variable. Thus, the number of fuzzy if-then rules for adaptive neuro-fuzzy learning is \( 2^3 = 8 \) only.
Table 1. Two passes in the hybrid learning procedure adaptive neuro-fuzzy inference.

<table>
<thead>
<tr>
<th>#</th>
<th># Forward Pass</th>
<th># Backward Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premise parameters</td>
<td>Fixed</td>
<td>Gradient descent</td>
</tr>
<tr>
<td>Consequent parameters</td>
<td>Least squares estimator</td>
<td>Fixed</td>
</tr>
<tr>
<td>Signals</td>
<td>Node outputs</td>
<td>Error signals</td>
</tr>
</tbody>
</table>

3.5. Decision making logic

An important step for successful pattern recognition problem is to design an appropriate decision rule for accurate classification of patterns. A set of 200 different mammograms has been used for the implementation of the proposed algorithm. The proposed neuro-fuzzy model has been trained with round-shaped benign masses specified by the radiologists. Our objective is to classify each of the tests mass, whether they belong to the benign or malignant stage. A distance function ($\mu_1$) has been defined for the determination of roundness deviation of the contour of mass. The degree of malignancy is higher for the higher value of $\mu_1$.

The distance function ($\mu_1$) is the Euclidean distance and corresponding decisions are given below:

$$\mu_1 = \|D_1 - O_1\|$$  \hspace{1cm} (24)

where, $D_1 = $ Desired output value for benign micro calcification, $O_1 = $ output value of the test mammograms

Hence, $\mu_1 = [(D_1 - O_1)^T(D_1 - O_1)]^{0.5}[(D_1 - O_1)^2]^{0.5}$  \hspace{1cm} (25)

The decision on the Boundary of test masses is defined below:

If $\mu_1 \leq 20$, the shape & margin of test masses are considered as: *Almost Round or Round to Oval Shape & Smooth Boundary — Benign Stage.*

If $20 < \mu_1 \leq 40$, the shape & margin of test masses are considered as: *Ovulated Shape & Non-Circumscribed Boundary — Tendency towards Malignancy.*

If $\mu_1 > 40$, the shape & margin of the test masses are considered as: *Irregular Shape & ill-defined Boundary — Possibly in Malignant Stage.*

4. Experimental Results

The methodology developed for classification of tumor lesion in breast has been implemented on digital images of mammographic database.

4.1. Image database

We have applied the proposed algorithm to the databases from Medical Image Analysis Society (MIAS) and others, consisting of 200 images. The classifier was first trained with obvious benign masses as identified by the expert radiologists. The other non-obvious cases have been tested and classified in benign and malignant
groups using the proposed methodology. Benign masses from the digital mammogram database are identified by radiologists. All required information related to extracted tumor mass is further used to design the decision support system for classification of tumor growth.

The training database consists of almost round or round to oval benign masses with circumscribed margins as specified by the radiologists. On the basis of that test image, database is evaluated.

4.2. Feature extraction

A set of features has been computed for each suspicious area identified from the mammogram in this research. The features include area, mean, standard deviation, variance, skew, entropy, energy, mod, median and RMS.

The formulae for every feature are described below.

For each of the formulae, \( T \) is the total number of pixels, \( g \) is an index value of image \( I \), \( K \) is the total number of gray levels, \( j \) the gray level value, \( I(g) \) is the gray level value of pixel \( g \) is in image \( I \), \( N(j) \) is the number of pixels with gray level \( j \) in image \( I \), \( P(I(g)) \) is the probability of gray level value \( I(g) \) occurring in image \( I \), \( P(j) = N(I(g))/T \), and \( P(j) \) is the probability of gray level value \( j \) occurring in image \( I \), \( P(j) = N(j)/T \).

Notations of the features are as follows:

- average gray level = \( \frac{1}{T} \sum_{g=0}^{T-1} I(g) \)
- energy = \( -\sum_{j=0}^{K-1} [P(j)]^2 \)
- entropy = \( -\sum_{j=0}^{K-1} P(j) \log_2 [P(j)] \)
- variance(\( \sigma^2 \)) = \( \sum_{g=0}^{T-1} (j - \text{AvgGray})^2 \)
- standard deviation(\( \sigma \)) = \( \sqrt{\sum_{g=0}^{T-1} (j - \text{AvgGray})^2} \)
- skew = \( \frac{1}{\sigma^3} \sum_{j=0}^{K-1} (j - \text{AvgGray})^3 \)
- rms value = \( \sqrt{\frac{1}{T} \sum_{g=0}^{T-1} I(g)^2} \)
- Mode = \( \max(N(j)) \)

Computation of the features of suspected patterns has been shown in Table 2.
4.3. Feature selection using GA

To conceive the intended body part using the GA evolutionary process, the feature of this particular chromosome must be specified. The program will proceed and each of the generated chromosomes will be checked according to this ideally specified parts. The measure of this checking mechanism represents the fitness function. This can be a combination of the features, such as the shape of the body, the color of the body and the size of the given part. In the genetic algorithm, feature selection involves many generations. The results are based on the random initialization to the weights of every individual in the population with the following parameters:

In present experiment: population size: 40, number of generation: 55, probability of crossover: 0.7, probability of mutation: 0.3.

Genetic algorithm - Pseudo-code algorithm

step 1 start
step 2 Choose initial population
step 3 extract the features from the data (images) such as mean, median, energy etc.
step 4 select the features using genetic algorithm (using crossover and mutation)
step 5 test the classification rates by the new generation
step 6 repeat from step 4.
step 7 compare classification rates of different output sets
step 8 repeat from step 4.
step 9 determine the best combination of features
step 10 end

The experiments have been presented with two micro-calcification areas (both benign) for training and 17 micro-calcification areas (11 benign, 6 malignant) were used for testing in this paper. Many experiments using different parameters were run to find the feature or combination of features that best classifies a micro-calcification area into benign and malignant.

The training database consists of almost round or round to oval benign masses with circumscribed margins. On the basis of the boundary/margin information acquired from train data, classification process has been accomplished for test database in benign and malignant category as demonstrated in Table 3. The Table 3 presents results regarding decision of benignancy/malignancy of few test data base where the example of training data base has been demonstrated by train data1 and train data2. The decision making parameter for the classification process is the distance function ($\mu_1$) which computes the deviation of any mass from the roundness and give a measure for degree of malignancy. In the present experiment, we
Table 2. Computation of features of suspected patterns.

<table>
<thead>
<tr>
<th>Data</th>
<th>Area</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Variance (2nd Moment)</th>
<th>Skew (3rd Moment)</th>
<th>Entropy</th>
<th>Energy</th>
<th>Mode</th>
<th>Median</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>372644</td>
<td>252.475</td>
<td>8.8653</td>
<td>78.594</td>
<td>−1469473.587</td>
<td>0.04211</td>
<td>0.9858</td>
<td>254.00</td>
<td>255.00</td>
<td>253.23</td>
</tr>
<tr>
<td>D2</td>
<td>368174</td>
<td>253.883</td>
<td>3.7514</td>
<td>14.0736</td>
<td>−1928244.52</td>
<td>0.03070</td>
<td>0.9962</td>
<td>254.00</td>
<td>255.00</td>
<td>253.91</td>
</tr>
<tr>
<td>D3</td>
<td>355070</td>
<td>250.885</td>
<td>18.8314</td>
<td>354.623</td>
<td>−149500.71</td>
<td>0.50468</td>
<td>0.9265</td>
<td>254.00</td>
<td>255.00</td>
<td>251.59</td>
</tr>
<tr>
<td>D4</td>
<td>368260</td>
<td>253.957</td>
<td>1.5272</td>
<td>2.3324</td>
<td>−29422259.353</td>
<td>0.02064</td>
<td>0.9976</td>
<td>254.00</td>
<td>255.00</td>
<td>253.96</td>
</tr>
<tr>
<td>D5</td>
<td>367800</td>
<td>253.697</td>
<td>6.75608</td>
<td>45.6446</td>
<td>−3384805.09</td>
<td>0.04106</td>
<td>0.9951</td>
<td>254.00</td>
<td>255.00</td>
<td>253.78</td>
</tr>
<tr>
<td>D6</td>
<td>367965</td>
<td>253.77</td>
<td>5.8958</td>
<td>33.7077</td>
<td>−5340480.52</td>
<td>0.03812</td>
<td>0.9956</td>
<td>254.00</td>
<td>255.00</td>
<td>253.84</td>
</tr>
<tr>
<td>D7</td>
<td>366918</td>
<td>253.72</td>
<td>5.5298</td>
<td>30.5796</td>
<td>−6175306.44</td>
<td>0.07913</td>
<td>0.9896</td>
<td>254.00</td>
<td>255.00</td>
<td>253.78</td>
</tr>
<tr>
<td>D8</td>
<td>368406</td>
<td>253.94</td>
<td>2.6354</td>
<td>6.9453</td>
<td>−5724580.725</td>
<td>0.01400</td>
<td>0.9984</td>
<td>254.00</td>
<td>255.00</td>
<td>253.95</td>
</tr>
<tr>
<td>D9</td>
<td>367986</td>
<td>253.82</td>
<td>4.1247</td>
<td>17.037</td>
<td>−1490265.964</td>
<td>0.04917</td>
<td>0.9941</td>
<td>254.00</td>
<td>255.00</td>
<td>253.85</td>
</tr>
<tr>
<td>D10</td>
<td>368001</td>
<td>253.73</td>
<td>6.1709</td>
<td>38.0806</td>
<td>−4444491.37</td>
<td>0.04858</td>
<td>0.9943</td>
<td>254.00</td>
<td>255.00</td>
<td>253.81</td>
</tr>
<tr>
<td>D11</td>
<td>367384</td>
<td>253.33</td>
<td>10.2572</td>
<td>105.2104</td>
<td>−961745.68</td>
<td>0.14012</td>
<td>0.9815</td>
<td>254.00</td>
<td>255.00</td>
<td>253.54</td>
</tr>
<tr>
<td>D12</td>
<td>363558</td>
<td>252.16</td>
<td>15.8095</td>
<td>249.9433</td>
<td>−257862.04</td>
<td>0.19789</td>
<td>0.9715</td>
<td>254.00</td>
<td>255.00</td>
<td>252.81</td>
</tr>
<tr>
<td>D13</td>
<td>367350</td>
<td>253.67</td>
<td>5.5917</td>
<td>31.2680</td>
<td>−5966006.29</td>
<td>0.05867</td>
<td>0.9926</td>
<td>254.00</td>
<td>255.00</td>
<td>253.73</td>
</tr>
<tr>
<td>D14</td>
<td>367830</td>
<td>253.64</td>
<td>7.0089</td>
<td>49.1258</td>
<td>−3028881.47</td>
<td>0.06603</td>
<td>0.9922</td>
<td>254.00</td>
<td>255.00</td>
<td>253.73</td>
</tr>
<tr>
<td>D15</td>
<td>366438</td>
<td>253.58</td>
<td>6.5275</td>
<td>42.6095</td>
<td>−3746199.62</td>
<td>0.09637</td>
<td>0.9879</td>
<td>254.00</td>
<td>255.00</td>
<td>253.66</td>
</tr>
<tr>
<td>D16</td>
<td>364140</td>
<td>252.31</td>
<td>15.9117</td>
<td>253.1846</td>
<td>−253522.15</td>
<td>0.18904</td>
<td>0.9750</td>
<td>254.00</td>
<td>255.00</td>
<td>252.81</td>
</tr>
<tr>
<td>D17</td>
<td>366666</td>
<td>253.54</td>
<td>6.4414</td>
<td>41.4919</td>
<td>−3986226.72</td>
<td>0.08516</td>
<td>0.9889</td>
<td>254.00</td>
<td>255.00</td>
<td>253.62</td>
</tr>
</tbody>
</table>
Table 3. Decision on degree and tendency of benignity/malignancy of some test database looking at $\mu_1$ value.

<table>
<thead>
<tr>
<th>Database of Masses</th>
<th>Obtained Value $O_1$</th>
<th>Decision on Margin of the Mass</th>
<th>Degree of Benignity/ Malignancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Train data 1)</td>
<td>50.9258</td>
<td>0.9258 Round to oval &amp; smooth</td>
<td>Benign Stage</td>
</tr>
<tr>
<td>(Train data 2)</td>
<td>50.1647</td>
<td>0.1647 Round to oval &amp; smooth</td>
<td>Benign Stage</td>
</tr>
<tr>
<td>Sample 1</td>
<td>17.4536</td>
<td>32.5464 Lobulated &amp; non-circumscribed</td>
<td>Transformation towards malignancy</td>
</tr>
<tr>
<td>Sample 2</td>
<td>5.8825</td>
<td>44.1175 Irregular &amp; ill-defined</td>
<td>Malignant Stage</td>
</tr>
<tr>
<td>Sample 3</td>
<td>6.2617</td>
<td>43.7383 Irregular &amp; ill-defined</td>
<td>Malignant Stage</td>
</tr>
<tr>
<td>Sample 4</td>
<td>4.8751</td>
<td>45.1249 Irregular &amp; ill-defined</td>
<td>Malignant stage</td>
</tr>
<tr>
<td>Sample 5</td>
<td>6.7601</td>
<td>43.2399 Irregular &amp; ill-defined</td>
<td>Malignant stage</td>
</tr>
<tr>
<td>Sample 6</td>
<td>23.4251</td>
<td>26.5749 Lobulated &amp; non-circumscribed</td>
<td>Marginally benign</td>
</tr>
<tr>
<td>Sample 7</td>
<td>20.2598</td>
<td>29.7402 Lobulated &amp; non-circumscribed</td>
<td>Marginally benign with a tendency towards malignant</td>
</tr>
<tr>
<td>Sample 8</td>
<td>10.5575</td>
<td>39.4425 Lobulated &amp; non-circumscribed</td>
<td>Possibility towards malignant stage</td>
</tr>
<tr>
<td>Sample 9</td>
<td>4.8944</td>
<td>45.1056 Irregular &amp; ill-defined</td>
<td>Malignant stage</td>
</tr>
<tr>
<td>Sample 10</td>
<td>10.5601</td>
<td>39.4399* Lobulated &amp; non-circumscribed</td>
<td>Possibility towards malignant stage</td>
</tr>
<tr>
<td>Sample 11</td>
<td>19.8101</td>
<td>30.1899 Lobulated &amp; non-circumscribed</td>
<td>Marginally benign with tendency towards malignancy</td>
</tr>
<tr>
<td>Sample 12</td>
<td>25.0830</td>
<td>24.9170 Lobulated &amp; non-circumscribed</td>
<td>Marginally benign</td>
</tr>
<tr>
<td>Sample 13</td>
<td>8.8083</td>
<td>41.1917 Irregular &amp; ill-defined</td>
<td>Possibly in malignant stage</td>
</tr>
<tr>
<td>Sample 14</td>
<td>5.4698</td>
<td>44.5302 Irregular &amp; ill-defined</td>
<td>Malignant stage</td>
</tr>
<tr>
<td>Sample 15</td>
<td>34.8931</td>
<td>15.1069 Round to oval &amp; smooth</td>
<td>Benign Stage</td>
</tr>
<tr>
<td>Sample 16</td>
<td>11.3161</td>
<td>38.6839 Irregular &amp; ill-defined</td>
<td>Possibility in malignant stage</td>
</tr>
<tr>
<td>Sample 17</td>
<td>25.6223</td>
<td>24.3777 Lobulated &amp; non-circumscribed</td>
<td>Marginally benign</td>
</tr>
<tr>
<td>Sample 18</td>
<td>21.6041</td>
<td>28.3959 Lobulated &amp; non-circumscribed</td>
<td>Marginally benign, having some tendency of malignancy</td>
</tr>
<tr>
<td>Sample 19</td>
<td>41.5803</td>
<td>8.4197 Round to oval &amp; smooth</td>
<td>Benign Stage</td>
</tr>
</tbody>
</table>

have used contour/margin based feature of the segmented tumor mass. The decision regarding the belongingness of the suspected region in benign or in malignant group has been taken according to the computed parameters for each case. Figure 6 shows some mammogram having train patterns and test patterns.

Superiority of classification process using GA based neural net approach has been supported by Verma et al. in their work. This has been accomplished by varying average population, population size and generation count.

The experiments were conducted by the classification rate of testing set to calculate the fitness for reproduction of Genetic feature selection. In all the tables,
Some examples of few of the non-obvious case studies and the corresponding final decision on benignancy/malignancy are given below:

Fig. 6. Some examples of training data and test data, the suspected pattern and contour of the pattern.
Fig. 6. (Continued)
Table 4. Highest classification rate.

<table>
<thead>
<tr>
<th>Features (1-Selected)</th>
<th>Training Set</th>
<th>Testing Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-E</td>
<td>M-E</td>
</tr>
<tr>
<td>1000010100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01011110100</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>10100101100</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1110100001</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1110011111</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>001111110000</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>001110100000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1000011000</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1100011100</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5. The feature selection reached classification rate >80% using threshold 0.4.

<table>
<thead>
<tr>
<th>Features (1-Selected)</th>
<th>Training Set</th>
<th>Testing Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-E</td>
<td>M-E</td>
</tr>
<tr>
<td>0000011001</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1001100111</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1011000011</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0010100111</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1000010111</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>001110100000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1110110000</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>000111110000</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0101100011</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0110011000</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0101011011</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6. The feature selections reached classification rate ≥80% using output threshold 0.3.

<table>
<thead>
<tr>
<th>Features (1-Selected)</th>
<th>Training Set</th>
<th>Testing Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-E</td>
<td>M-E</td>
</tr>
<tr>
<td>1110111011</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>1110100111</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>1011100001</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>1100111011</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>1011100000</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>1111110111</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>0001101001</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>0101011011</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>0100010111</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>0100111011</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>1000111100</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>1101110001</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>0011110000</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>1100111110</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>
Table 7. The feature selections reached classification rate 85% using different thresholds.

<table>
<thead>
<tr>
<th>Features (1-Selected)</th>
<th>Threshold</th>
<th>B-E</th>
<th>M-E</th>
<th>T-E</th>
<th>T-Rate (%)</th>
<th>B-E</th>
<th>M-E</th>
<th>T-E</th>
<th>T-Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1101111101</td>
<td>0.3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>88.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0101001111</td>
<td>0.3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td>83.33</td>
</tr>
<tr>
<td>0010011010</td>
<td>0.3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>88.89</td>
</tr>
<tr>
<td>1011110001</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td>88.89</td>
</tr>
<tr>
<td>0101101111</td>
<td>0.4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td>88.89</td>
</tr>
<tr>
<td>0110110101</td>
<td>0.4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td>88.89</td>
</tr>
</tbody>
</table>

the column features was described by using the values 0s and 1s. The 0 means the responded feature is not selected and the 1 means the feature is selected.

The sequence of the features is: (1) area, (2) mean, (3) standard deviation, (4) variance, (5) skew, (6) entropy, (7) energy, (8) mod, (9) median, and (10) RMS.

Benign-Error (“B-E”) is used to represent the number of classification errors for benign micro-calculations of the whole training set or testing set. Malignant-Error (“M-E”) represents the number of classification errors for malignant micro-calculations of the whole training set or testing set. Total-Error (“T-E”) refers to the number of classification errors for all the micro-calculations of the whole training set or testing set. “T-rate” is the abbreviation of Total-classification Rate. It is calculated by the following formula: “T-rate” = ((total number of samples) - “T-E”)/(total number of samples).

Tables 4 to 7 depict the classification results of the suspected areas at different threshold levels with percentage T rate.

5. Discussion

In proposed methodology for tumor classification using mammogram, authors have attempted to develop a technique based on GA based adaptive neuro-fuzzy model using boundary of lesion or region of interest ROI. This classification concerns the prediction regarding the prognosis of the disease either towards benignancy or malignancy using machine intelligence, taking shape and margin into account. The proposed classification network has been trained by hybrid-learning rule with maximum fifty epochs. The learning process is continued till the performance goal is reached. It has been noted that there is no sharp demarcation among the three stages — benign stage, tendency towards malignancy and malignant stage. Under this circumstance, the adaptive neuro-fuzzy is a robust soft-computing approach to take decision even in the presence of uncertainty as it appears in the process of classification of tumor masses in subsequent groups of benignancy and malignancy. The experimental results show that the proposed technique is able to detect the potential abnormalities, and may alert the radiologists during the therapeutic planning.
Proposed method also introduces GA for the searching of two of the most significant \textit{Fourier shape descriptors} that are able to represent a particular class of masses. \textit{Compactness measure} along with two \textit{Fourier descriptors} introduces very precise shape representation of a particular class of the tumor lesion. We have designed a classifier based on \textit{genetic algorithm} and \textit{adaptive neuro-fuzzy} structure to overcome the problem of over-learning and chance of misclassification for breast cancer screening. In the present paper, the classification process is based on boundary/margin feature of tumor lesion and we have considered adaptive neuro-fuzzy based approach for contour detection. We understand from the recommendation of radiologists/physicians, the accuracy level for classification will be enhanced if we include other features related to tumor biology and imaging properties.

6. Conclusion
The achieved classification rate and feature sets are promising. The highest classification rate achieved for testing set was 88.89%. Five features are considered to be the most significant features of a digital mammogram for microcalcification classification. They are mean, standard deviation, skew, energy and entropy. Our genetic algorithm based feature selection approach is effective. It has got the best classification rate than any other randomly selected feature combinations.

Acknowledgment
The authors would like to thank to Dr. S. K. Sharma of EKO X-ray and Imaging Institute, Kolkata. The authors also acknowledge CSIR, India for financial support to continue this research work.

References


