Egyptian Informatics Journal (2016) **17**, 11–20

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

FULL-LENGTH ARTICLE

Classification of Clustered Microcalcifications using MLFFBP-ANN and SVM

Baljit Singh Khehra [a,*], **Amar Partap Singh Pharwaha** [b,1]

[a] *Department of Computer Science & Engineering, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib 140407, Punjab, India*
[b] *Department of Electronics and Communication Engineering, Sant Longowal Institute of Engineering and Technology, Longowal 148106, Sangrur, Punjab, India*

Received 26 March 2015; revised 26 July 2015; accepted 16 August 2015
Available online 14 September 2015

KEYWORDS

Computer-Aided Diagnosis;
Clustered Microcalcifications;
LM-MLFFBP-ANN;
SMO-SVM;
Confusion matrix and ROC analysis

Abstract The classifier is the last phase of Computer-Aided Diagnosis (CAD) system that is aimed at classifying Clustered Microcalcifications (MCCs). Classifier classifies MCCs into two classes. One class is benign and other is malignant. This classification is done based on some meaningful features that are extracted from enhanced mammogram. A number of classifiers have been proposed for CAD system to classify MCCs as benign or malignant. Recently, researchers have used Artificial Neural Networks (ANNs) as classifiers for many applications. Multilayer Feed-Forward Backpropagation (MLFFB) is the most important ANN that has been successfully used by researchers to solve various problems. Similarly, Support Vector Machines (SVMs) belong to another category of classifiers that researchers have recently given considerable attention. So, to explore MLFFB and SVM classifiers for MCCs classification problem, in this paper, Levenberg-Marquardt Multilayer Feed-Forward Backpropagation ANN (LM-MLFFB-ANN) and Sequential Minimal Optimization (SMO) based SVM (SMO-SVM) are used for the classification of MCCs. Thus, a comparative evaluation of the relative performance of LM-MLFFBP-ANN and SMO-SVM is investigated to classify MCCs as benign or malignant. For this comparative evaluation, first suitable features are extracted from mammogram images of DDSM database. After this, suitable features are selected using Particle Swarm Optimization (PSO). At the end, MCCs are classified using LM-MLFFBP-ANN and SMO-SVM classifiers based on the selected features. Confusion matrix and ROC analysis are used to

* Corresponding author. Tel.: +91 9463446505.
E-mail addresses: baljitkhehra@rediffmail.com (B.S. Khehra), amarpartapsingh@yahoo.com (A.P.S. Pharwaha).
[1] Tel.: +91 9463122255.

Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.eij.2015.08.001

measure the performance of LM-MLFFBP-ANN and SMO-SVM classifiers. Experimental results indicate that the performance of SMO-SVM is better than that of LM-MLFFBP-ANN.

1. Introduction

Breast cancer that occurs among women in both developed and developing countries is one of the most dangerous diseases. It is difficult to prevent it but early detection is the key for reducing the mortality rate. Mammography is one of the most effective imaging techniques for early detection of breast cancer [1]. Clusters of Microcalcifications (MCCs), mass lesions, distortion in breast architecture and asymmetry between breasts are various types of breast abnormalities that are partially detected from mammograms. Clusters of Microcalcifications (MCCs) are the most frequent symptoms of Ductal Carcinoma in Situ (DCIS). DCIS is one of various types of breast cancers [2]. Although mammography is frequently used in both developed and developing countries for breast cancer detection, but un-correct reading of mammogram is a problem. This type of problem is occurred due to human error. Un-correct readings of mammogram are called false positive and false negative readings of mammogram. Due to false positive detection, the need of unnecessary biopsy occurs while due to false negative detection, an actual tumor remains undetected. Thus, false positive and false negative readings of mammogram are main causes of unnecessary biopsy and missing the best treatment time. In fact, the need of the hour is to develop Computer-Aided Diagnosis (CAD) system for improving diagnosis accuracy of early breast carcinoma which would prevent unnecessary biopsy and not miss the best treatment time. The classifier is the last phase of CAD system that is aimed to classify MCCs as benign or malignant. For this, first mammogram images are enhanced. After this, features are extracted from enhanced mammogram. Then, a suitable set of features is searched from extracted features. At the end, classifiers classify MCCs based on suitable set of features. Recently, various researchers have applied a variety of classifiers for CAD system to classify MCCs as benign or malignant. Kramer and Aghdasi [3] used multi-scale statistical texture features to classify MCs in digitized mammograms using K-Nearest Neighbor (KNN) classifier. Bruce and Adhami [4] classified mammographic masses into stellate, nodular and round using Linear Discriminant Analysis (LDA) classifier. Bottema and Slavotinek [5] classified lobular and DCIS (small cell) MCs in digital mammograms using decision trees. In 2007, Bayesian network classifiers are used by Nicandro et al. [6] for the diagnosis of breast cancer. Fuzzy rough sets hybrid scheme is used by Hassanien [1] for breast cancer detection. For classification of MCs as benign or malignant, Artificial Neural Networks (ANNs) have been widely used [7–9]. Multilayer Feed-Forward Backpropagation (MLFFB) is the most important ANN that has been applied successfully to solve many problems [10–12]. Support Vector Machines (SVMs) have also been recently used to solve many problems [2,13–16]. SVMs are based on statistical learning theory. In the proposed research work, LM-MLFFBP-ANN and SMO-SVM are explored to classify MCCs as benign or malignant. Experiments are performed on mammogram images of DDSM database [17].

2. Artificial Neural Network

An ANN is a computational model that is commonly used in situations when the knowledge is not properly defined and there is a need to solve non-linear complex problems. Multilayer Feed-Forward Neural Network trained with Backpropagation algorithm is widely used for non-linear classification problems [18]. Multilayer Feed-Forward Backpropagation Artificial Neural Network (MLFFBP-ANN) is treated as a nested sigmoid scheme. Therefore, the following equation is used to represent the output function of ANN [19]:

$$F(x_i) = F_N(W_N * (F_{N-1}(\ldots F_2(W_2 * F_1(W_1 * x_1 + B_1) + B_2)\ldots) + B_{N-1}) + B_N) \quad (1)$$

where N is total layers of Artificial Neural Network; B_1, B_2, …, B_N are the bias vectors; W_1, W_2, …, W_N are the weight vectors and F_1, F_2, …, F_N are activation transfer functions of layers.

In fact, studies on ANNs [20,21] have highlighted that the most often used ANN architecture is feed-forward network designed around multilayer topologies with Backpropagation learning algorithm. In the feed-forward ANN, a neuron transfers data to the neuron of the next layer through a function called activation function that may be the sigmoid or linear one [22]. When sigmoid activation function is used, the output of the jth neuron is calculated as

$$z_j = \frac{1}{1 + e^{-\sigma(z\_in_j)}} \quad (2)$$

where z_in_j is the input of the jth neuron received from the neurons of the previous layer calculated as

$$z\_in_j = b_j + \sum_{i=1}^{n} x_i w_{ij} \quad (3)$$

where x_i is the output of ith neuron of the previous layer; n is the total number of neurons in the previous layer; w_{ij} is the connection weight of the jth neuron with ith neuron of the previous layer; b_j is the jth neuron bias and σ defines the steepness of the sigmoid activation function.

In the learning phase, w_{ij} and b_j values are updated using the following equations [23]:

$$w_{ij}(new) = w_{ij}(old) + \Delta w_{ij} \quad (4)$$

$$b_j(new) = b_j(old) + \Delta b_j \quad (5)$$

$$\Delta w_{ij} = \alpha \delta_j x_i \quad (6)$$

$$\Delta b_j = \alpha \delta_j \quad (7)$$

where α is the learning rate and δ is correction factor. The performance of ANN is evaluated by Mean Square Error (MSE) that is defined as

$$\text{MSE} = \frac{1}{L}\sum_{k=1}^{L}\sum_{j=1}^{m}(t_j^k - y_j^k)^2 \tag{8}$$

where L is the number of training pairs; m is the number of neurons in the output layer; y_j^k and t_j^k are the actual and target outputs at jth neuron for kth training pair.

2.1. MLFFBP-ANN for classifying MCCs

Classification of MCCs as benign or malignant classes is a two-class pattern classification problem. Let $X_j \in [x_1, x_2, x_3, \ldots, x_i, \ldots, x_n]$ be a set of features extracted from mammograms that acts as input vector for MLFFBP-ANN and $Y_j \in [0, 1]$ be output vector of MLFFBP-ANN. '0' represents benign MCCs and '1' represents malignant MCCs. Let M^j be one training set of L samples, i.e. $M^j = [(X_j, Y_j)],\ j = 1, 2, 3, \ldots, L$.

Number of neurons in input layer (n) depends upon input vector and number of hidden neurons (p) is chosen experimentally. Number of neurons in output depends upon output vector. Features are used as inputs of the neurons of input layer. Thus, n is number of features extracted from MCCs. Classification of MCCs is a two-class pattern classification problem. So, output layer has only one neuron to represent binary output. Neurons are interconnected with each other and a weight is assigned to each link for representing the link-strength between the neurons. In order to classify MCCs as benign or malignant, an attempt is made here to implement a MLFFBP-ANN based classifier to classify MCCs as benign or malignant. An algorithm is written using MATLAB programming. The implementation neural model comprises of a hidden layer of sigmoidal neurons that receives numeric values of features and broadcasts output values to a layer of linear neurons, which finally computes the network output. Using Eq. (2), the output of jth neuron of hidden layer for the proposed model is computed by

$$z_j = \tan sig\left(v_{oj} + \sum_{i=1}^{n} x_i v_{ij}\right) \tag{9}$$

where n is the total number of neurons in input layer; v_{oj} is the bias of jth neuron of hidden layer; v_{ij} is the weight between ith neuron of input layer and jth neuron of hidden layer and tan sig(ṡ) is a hyper-tangent sigmoid activation transfer function. The output of the proposed MLFFBP-ANN model is computed as

$$F_{\text{ANN}}(X_j) = purelin\left(w_{ok} + \sum_{j=1}^{p} z_j w_{jk}\right) \tag{10}$$

where p is the total number of neurons in hidden layer; w_{ok} is the bias of kth neuron of the output layer (in the proposed model, $k = 1$ because output layer has only one neuron); w_{jk} is the weight between jth neuron of hidden layer and kth neuron of the output layer; and $purelin$(ṡ) is linear activation transfer function.

The proposed MLFFBP-ANN model updates weights and biases values by means of an adaptive process which minimizes the output neurons errors using Eqs. (4)–(7) as follows:

$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij} \tag{11}$$

$$\Delta v_{ij} = \alpha \delta_j x_i \tag{12}$$

$$v_{oj}(new) = v_{oj}(old) + \Delta v_{oj} \tag{13}$$

$$\Delta v_{oj} = \alpha \delta_j \tag{14}$$

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk} \tag{15}$$

$$\Delta w_{jk} = \alpha \delta_k z_j \tag{16}$$

$$w_{ok}(new) = w_{ok}(old) + \Delta w_{ok} \tag{17}$$

$$\Delta w_{ok} = \alpha \delta_k \tag{18}$$

where δ_k is factor that is used to update weights w_{jk} and δ_j is a factor that is used to update weights v_{ij}. Using Eq. (8), MSE for the proposed model is calculated as

$$\text{MSE} = \frac{1}{L}\sum_{j=1}^{L}[Y_j - F_{\text{ANN}}(X_j)]^2 \tag{19}$$

During training, a set of numeric values of features of MCCs corresponding to the MCCs category is used to update the weights and biases of the neurons to minimize the output neuron error. However, the best ANN structure is not known in advance [18]. The best ANN structure depends upon the number of hidden layers, number of neurons in each hidden layer, activation function, learning algorithm and training parameters. To train MLFFBP-ANN, various training algorithms are available [24]. In the proposed research work, Levenberg-Marquardt training algorithm is considered to train MLFFBP-ANN for characterization of MCCs as benign or malignant.

3. Support Vector Machine

SVM is a two-class classifier developed by Vapnik [25]. Learning of SVM is supervising learning that is based on statistical learning theory. Basic principal of SVM is structural risk minimization. Structural risk minimization means to get a low error rate on unseen data set (outside training data set). For non-linear classification problems, kernel function based SVM is used. Kernel function converts non-linear classification problem to linear classification problem through mapping the input feature space to higher dimensional feature. After this, an optimal separating hyperplane is used to separate the two classes of the two-class pattern classification problem.

For classification of MCCs as benign or malignant, a set of L training data samples is considered. Such set is denoted as $\{(X_j, Y_j),\ j = 1, 2, \ldots, L\}$, where X_j is input data sample that belongs to class $Y_j \in \{+1, -1\}$. Input data sample is represented by a vector $X \in \{x_i, i = 1, 2, \ldots, n\}$ in which $\{x_i, i = 1, 2, \ldots, n\}$ is a set of n features of the cluster of MCs and output is represented as $Y \in \{+1, -1\}$, where '+1' is for malignant cluster of MCs and '−1' is for benign cluster of MCs. For separating the positive and negative classes, a separating hyperplane is used. Separating hyperplane should be optimal for correct classification of positive and negative classes. From the above discussion, the formulation of optimization problem to find optimal separating hyperplane can be stated as

$$\text{Target}: \ \text{Minimize}\ \frac{w^T w}{2}$$

$$\text{Constraints}: \ Y_j(w^T X_j + b) \geqslant 1 \quad \text{for } \forall j$$

where w is the norm to the hyperplane; $\frac{|b|}{\|w\|}$ is the perpendicular distance from the origin to hyperplane to the origin.

Thus, the main objective is to find w and b for minimization of $\frac{w^T w}{2}$ along with the satisfaction of constraints. The optimal values w^* and b^* are used to classify a test example Z as follows:

$$class(Z) = sign(w^{*T} Z + b^*) \tag{20}$$

The above defined problem belongs to Quadratic Programming (QP) optimization problem with linear constraints. Lagrangian formulation of the above said problem [26] is required to solve it. In Lagrangian formulation, objective function is defined as

$$L_P = \frac{1}{2} w^T w - \sum_{j=1}^{L} \alpha_j Y_j (w^T X_j + b) + \sum_{j=1}^{L} \alpha_j \tag{21}$$

Target: Minimize L_P w.r.t w, b

Constraints:

(a) Derivatives of L_P with respect to all α_j vanish
(b) $\alpha_j \geqslant 0$

Dual formulation of the above primal problem is written as

Target: Maximize L_P

Constraints:

(a) Gradient of L_P w.r.t w and b vanishes
(b) $\alpha_j \geqslant 0$

When gradient of L_P w.r.t w and b is vanished, then the following conditions are occured:

$$w = \sum_{j=1}^{L} \alpha_j Y_j X_j \tag{22}$$

$$\sum_{j=1}^{L} \alpha_j Y_j = 0 \tag{23}$$

From Eqs. (21)–(23), the following equation is obtained:

$$L_D = \sum_{j=1}^{L} \alpha_j - \frac{1}{2} \sum_{j=1}^{L} \sum_{i=1}^{L} \alpha_j \alpha_i Y_j Y_i X_j^T X_i \tag{24}$$

Thus, new formulation of the problem is obtained as

Target: Maximize L_D

Constraints:

(a) $\sum_{j=1}^{L} \alpha_j Y_j = 0$
(b) $\alpha_j \geqslant 0$

Thus, the main objective is to find $\alpha_1, \alpha_2, \ldots, \alpha_L$ for maximization of L_D along with the satisfaction of constraints. For non-zero α_j^*, $j = 1, 2, \ldots, L_s$, the optimal values of w^* and b^* are obtained as follows:

$$w^* = \sum_{j=1}^{L_s} \alpha_j^* Y_j X_j \tag{25}$$

$$b^* = Y_j - \sum_{j=1}^{L_s} \alpha_j^* Y_j X_j^T X_j \tag{26}$$

where non-zero Lagrange multipliers, α_j^*, $j = 1, 2, \ldots, L_s$, indicate their corresponding support vectors $S_j \in (X_j, Y_j)$. Thus, the following equation is used to classify a test example Z as

$$class(Z) = sign\left(\sum_{j=1}^{L_s} \alpha_j^* Y_j X_j^T Z + b^* \right) \tag{27}$$

3.1. Karush–Kuhn–Tucker condition

Lagrangian formulation of the problem (L_P) is a convex minimum QP optimization problem. Property of convex minimization problem is as follows: if a local minimum exists, then it is a global minimum [27]. Karush–Kuhn–Tucker condition [28] is sufficient when objective function is convex and solution space is also convex. According to this condition, gradient of the objective function of Lagrangian problem w.r.t w and b vanishes and multiplication of each Lagrangian multiplier with corresponding constraint is also zero for all Lagrangian multipliers greater than or equal to zero [29,30].

$$\frac{\partial L_P}{\partial w} = 0 \Rightarrow w = \sum_{j=1}^{L} \alpha_j Y_j X_i \tag{28}$$

$$\frac{\partial L_P}{\partial b} = 0 \Rightarrow \sum_{j=1}^{L} \alpha_j Y_j = 0 \tag{29}$$

$$\alpha_j (Y_j (w^T X_j + b) - 1) = 0 \quad \forall j \tag{30}$$

$$\alpha_j \geqslant 0 \quad \forall j \tag{31}$$

The above equations are used to obtain optimal values of w^* and b^*.

3.2. Non-linear SVM classifier

For solving non-linear classification problems, non-linear SVM classifiers are used through kernel functions. Kernel function maps training input data of input space R^d onto a higher dimensional feature space H using transformation operator $\phi(\cdot)$. This is done to separate training data points into two classes by a hyperplane [29].

$$\phi : \ R^d \rightarrow H \tag{32}$$

Relation between kernel function $K(X_j, X_i)$ and mapping operator $\phi(\cdot)$ [31] is shown as

$$K(X_j, X_i) = \phi(X_j)^T \phi(X_i) \quad \forall X_j, X_i \in R^d \tag{33}$$

Thus, dual form of the problem can be formulated as follows:

Find α_1, α_2, …, α_L such that $L_D = \sum_{j=1}^{L} \alpha_j - \frac{1}{2} \sum_{j=1}^{L} \sum_{i=1}^{L} \alpha_j \alpha_i Y_j Y_i K(X_j, X_i)$ is maximized and

(a) $\sum_{j=1}^{L} \alpha_j Y_j = 0$
(b) $0 \leqslant \alpha_j \leqslant C$ for $\forall \alpha_j$

For non-zero α_j^*, $j = 1, 2, \ldots, L_s$, w^* is calculated from Eq. (25) and b^* is obtained as follows:

$$b^* = Y_j - \sum_{j=1}^{L_s} \alpha_j^* Y_j K(X_j, X_j) \tag{34}$$

Thus, a test example Z is classified as

$$class(Z) = sign\left(\sum_{j=1}^{L_s}\alpha_j^* Y_j K(X_j, Z) + b^*\right) \quad (35)$$

A most commonly used kernel function in SVM [32,33] is linear that is defined as follows:

$$K(x, y) = x^T y \quad (36)$$

3.3. Sequential Minimal Optimization for SVM

Sequential Minimal Optimization (SMO) [34] decomposes SVM-QP problem into QP sub-problems. At each step, the smallest possible optimization problem is selected for solving. For this, at each step, SMO selects two Lagrange multipliers. This is done to find the optimal values for Lagrange multipliers and SVM is updated to reflect the new optimal values. First, a Lagrange multiplier (α_1) is selected that violates the Karush–Kuhn–Tucker condition [28] for the optimization problem. After this, second Lagrange multiplier (α_2) is selected and optimizes the pair (α_2, α_2). This process is repeated to achieve the convergence. The main advantage of SMO is that two Lagrange multipliers are solved analytically instead of entirely numerical QP optimization. In addition, no extra matrix is required for storage at all.

Table 1 Confusion matrix.

	Target	
	Positive	Negative
Decision by classifier		
Positive	*True positive*	*False positive*
Negative	*False negative*	*True negative*

4. Measures for classifier accuracy

Confusion matrix [35] and ROC analysis [36] are two measures that are commonly used to find the accuracy of classifier to classify MCCs as benign and malignant. A confusion matrix evaluates the accuracy of classifier based on the actual and predicted classifications done by a classifier. For a classifier and an instance, possible outcomes are four: true positive, true negative, false positive and false negative. True positive is a correct judgment of classifier about a malignant cluster of MCs while true negative is a correct judgment of classifier about a benign cluster of MCs. Similarly, false positive is a wrong judgment of classifier about a benign cluster of MCs while false negative is a wrong judgment of classifier about a malignant cluster of MCs. These possible outcomes of a classifier are shown in Table 1. Such table is called confusion matrix.

ROC analysis is another measure that is used to find the accuracy of classifier related to medical decision. In ROC analysis, ROC curve is plotted to measure the accuracy of classifier. To plot ROC curve, 1-*Specificity* is taken along x-axis while *Sensitivity* is taken along y-axis and at various threshold settings, the curve is generated by plotting the *Sensitivity* against the 1-*Specificity*. The meaning of 1-*Specificity* is *False Positive Rate* while *Sensitivity* is *True Positive Rate*.

In case of confusion matrix, accuracy of classifier is found by using the following equation:

True Positive Rate and *False Positive Rate* are defined as

$$True\ Positive\ Rate\ (Sensitivity) = \frac{TPs}{TPs + FNs} \quad (37)$$

$$False\ Positive\ Rate\ (1\text{-}Specificity) = \frac{FPs}{TNs + FPs} \quad (38)$$

where TPs are number of true positive decisions taken by classifier; TNs are number of true negative decisions taken

Table 2 Performance of LM-MLFFBP to classify MCCs in 10 random experimental trials.

Trial no.	Confusion matrix		Accuracy from confusion matrix	Area under ROC curve (A_Z)	Sensitivity	Specificity
1	98	10	0.8466	0.9003	0.8376	0.8611
	19	62				
2	94	9	0.8307	0.9027	0.8034	0.8750
	23	63				
3	106	28	0.7937	0.8383	0.9060	0.6111
	11	44				
4	98	18	0.8042	0.8371	0.8376	0.7500
	19	54				
5	104	21	0.8201	0.8370	0.8889	0.7083
	13	51				
6	101	10	0.8624	0.9098	0.8632	0.8611
	16	62				
7	99	17	0.8148	0.8782	0.8462	0.7639
	18	55				
8	108	12	0.8889	0.9082	0.9231	0.8333
	9	60				
9	104	24	0.8042	0.8800	0.8889	0.6667
	13	48				
10	101	16	0.8307	0.8460	0.8632	0.7778
	16	56				
Average	–	–	0.8296	0.8738	0.8658	0.7708
Standard deviation	–	–	0.0294	0.0313	0.0363	0.0893

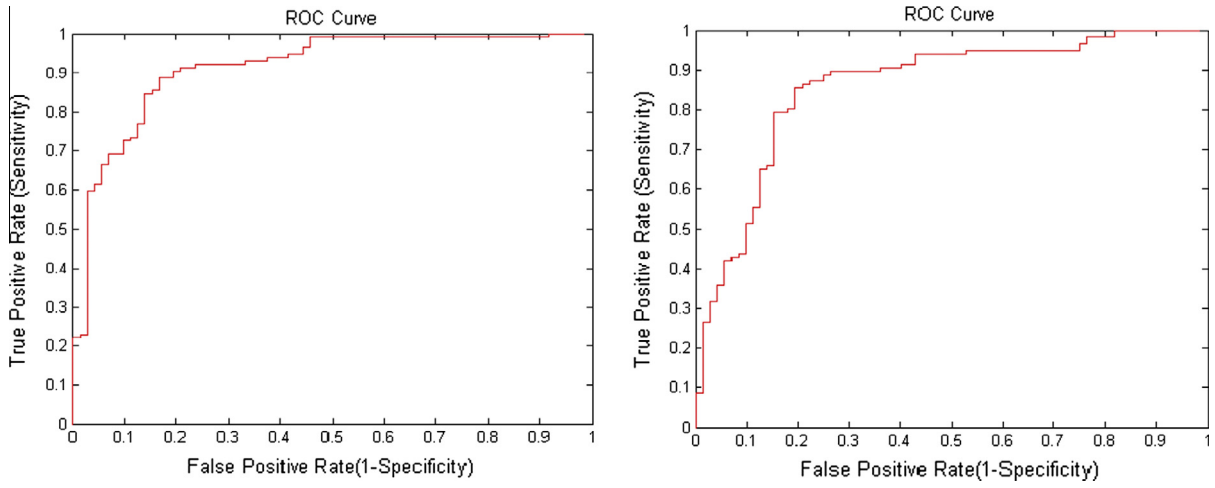

Figure 1 ROC curves of 1st and 10th random experimental trials of LM-MLFFBP classifier for classifying MCCs as benign and malignant.

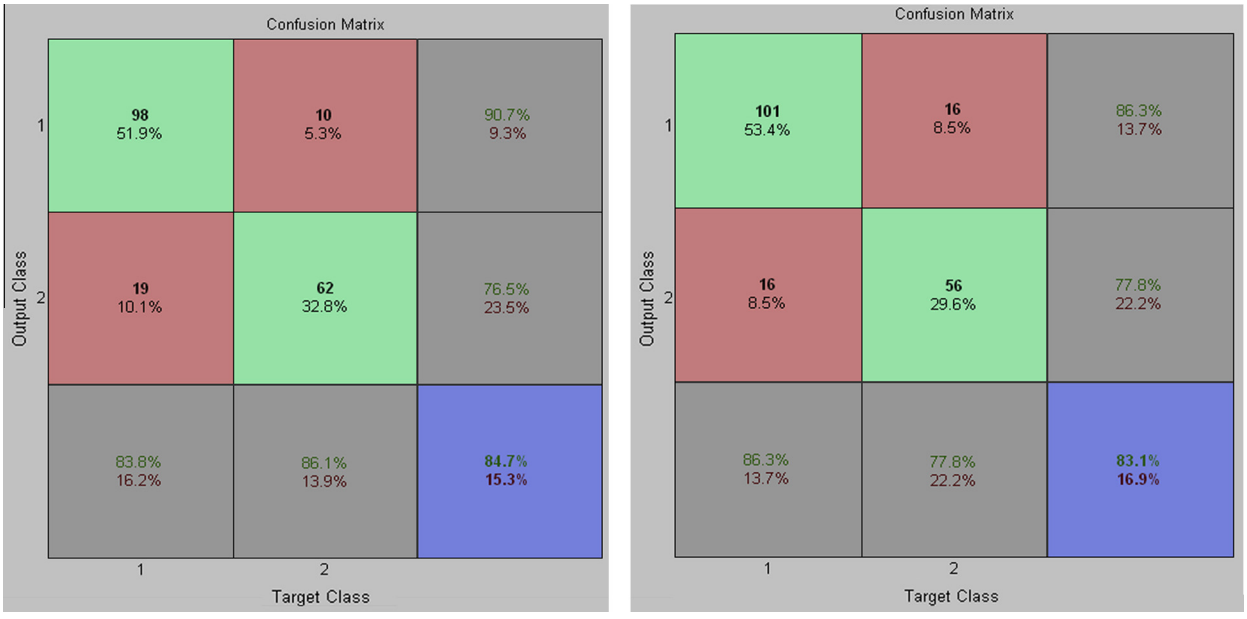

Figure 2 Confusion matrices of 1st and 10th random experimental trials of LM-MLFFBP classifier for classifying MCCs as benign and malignant.

by classifier; *FNs* are number of false negative decisions taken by classifier and *FPs* are number of false positive decisions taken by classifier.

In case of confusion matrix, accuracy of classifier is found by using the following equation:

$$Accuracy = \frac{TPs + TNs}{TPs + FPs + TNs + FNs} \tag{39}$$

In case of ROC analysis, the area under the ROC curve (A_Z) is used to measure the accuracy of classifier [37]. A_Z is in the range between 0.0 and 1.0. So, A_Z lies between 0.0 and 1.0. For 100% accuracy, A_Z should be 1.0. Trapezoidal rule or Simpson's rule can be used to compute A_Z.

According to Hosmer and Lemeshow [38], classifiers are divided into the following four categories based on the *Accuracy*:

- If $0.5 \leqslant Accuracy < 0.6$, then classifier is called fail classifier
- If $0.6 \leqslant Accuracy < 0.7$, then classifier is called poor classifier
- If $0.7 \leqslant Accuracy < 0.8$, then classifier is called fair classifier
- If $0.8 \leqslant Accuracy < 0.9$, then classifier is called good classifier
- If $0.9 \leqslant Accuracy \leqslant 1.0$, then classifier is called excellent classifier

5. Experimental results and discussion

In order to explore LM-MLFFBP-ANN and SMO-SVM to classify MCCs as benign or malignant, experiments are performed on data extracted from mammogram images of DDSM database [17]. For comparative evaluation, confusion matrix and ROC analysis are used. MATLAB 7.7 software is used for simulation.

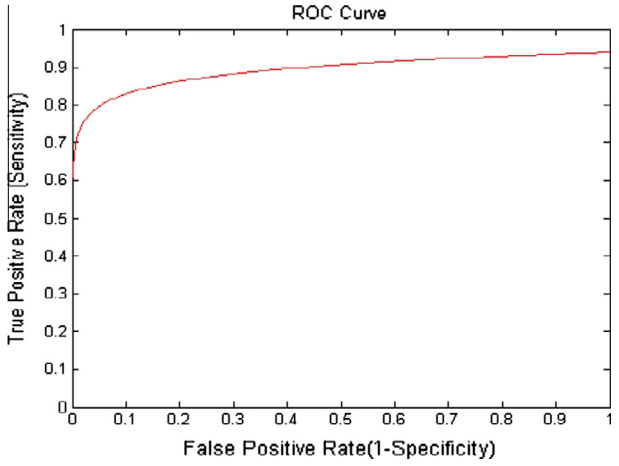

Figure 3 Common ROC curve of 10 random experimental trials of LM-MLFFBP classifier for classifying MCCs as benign and malignant.

Table 3 Average accuracy of LM-MLFFBP in terms of confusion matrix and ROC analysis.

Average accuracy from confusion matrices	Average accuracy from ROC curves	Accuracy from common ROC curve	Overall accuracy
0.8296	0.8738	0.8918	0.8651

5.1. Results of LM-MLFFBP-ANN

In order to find the performance of LM-MLFFBP for classifying MCCs as benign or malignant, different types of mammogram images are taken from standard benchmark digital database for screening mammography (DDSM) [17]. From mammogram images of DDSM database, a total of 380 suspicious regions are selected. From these samples, malignant samples are 235 and benign samples are 145. A set of 50 features is extracted from suspicious regions [39]. After this, Particle Swarm Optimization (PSO) is used to select an optimal subset of 23 most suitable features from 50 extracted. Such optimal subset of features is used in LM-MLFFBP. In the architecture of LM-MLFFBP, one hidden layer is taken with 15 hidden units. The activation function between input layer and hidden layer is sigmoid while that between hidden layer and output layer is linear. Default values of different parameters are set according to MATLAB 7.7 environment. For training purpose, 191 samples are selected from 380 samples and the remaining samples (189) are used for testing purpose. For finding the performance of the classifier to classify MCCs as benign or malignant, 10 random experiment trials are performed. Confusion matrix and ROC analysis are used to measure the performance of the trained classifier for classifying Clusters of MCs. Tabular results of 10 random experimental trials of LM-MLFFBP for classifying MCCs as benign or malignant in the form of accuracy calculated from confusion matrix and ROC analysis are shown in Table 2. Fig. 1 illustrates ROC curves of first and last random experimental trials while Fig. 2 illustrates confusion matrices of first and last random experimental trials.

Table 4 Performance of SMO-SVM with linear kernel function to classify MCCs in 10 random experimental trials.

Trial no.	Confusion matrix		Accuracy from confusion matrix	Area under ROC curve (A_Z)	Sensitivity	Specificity
1	107	14	0.8730	0.8686	0.9145	0.8056
	10	58				
2	110	13	0.8942	0.8941	0.9402	0.8194
	7	59				
3	106	15	0.8624	0.8571	0.9060	0.7917
	11	57				
4	108	16	0.8677	0.8663	0.9231	0.7778
	9	56				
5	105	11	0.8783	0.8704	0.8974	0.8472
	12	61				
6	107	14	0.8730	0.8686	0.9145	0.8056
	10	58				
7	111	19	0.8677	0.8761	0.9487	0.7361
	6	56				
8	104	8	0.8889	0.8799	0.8889	0.8889
	13	64				
9	107	12	0.8836	0.8782	0.9145	0.8333
	10	60				
10	107	9	0.8995	0.8927	0.9145	0.8750
	10	63				
Average	–	–	0.8788	0.8752	0.9162	0.8181
Standard deviation	–	–	0.0124	0.0116	0.0179	0.0456

From confusion matrices, it is observed that the average accuracy is 0.8296 while from ROC analysis (average of all areas under ROC curves) the average accuracy is 0.8738. In ROC analysis, Simpson's rule is used to find area under ROC curve. Logarithmic function is used to plot a common ROC curve of 10 random experimental trials. Common ROC curve is shown in Fig. 3. Accuracy from common ROC curve is 0.8918. The overall accuracy of LM-MLFFBP is calculated through the average of the three accuracy measures (average accuracy from confusion matrices, average accuracy from ROC curve and accuracy from common ROC curve). Thus, the overall accuracy of LM-MLFFBP is 0.8651 that is shown in Table 3.

5.2. Results of SMO-SVM

Secondly, in order to explore the performance of SMO-SVM for classifying MCCs as benign or malignant, the same samples that have been used for LM-MLFFBP are considered. The same 23 features are used that have been selected from 50 features by PSO for LM-MLFFBP. In this study, linear kernel function is considered. In the same way as used in LM-MLFFBP, the same 191 samples as used in LM-MLFFBP are used for training and the same 189 samples are used for testing purpose. Similarly, as in LM-MLFFBP, 10 random experimental trials are performed. Results of 10 random experimental trials in the form of confusion matrix and area under ROC curve are shown in Table 4. Fig. 4 is used to show ROC curves of first and last random experimental trials while Fig. 5 illustrates confusion matrices of first and last experimental trials. From confusion matrix, the average accuracy of SMO-SVM classifier for classifying Clusters of MCs is 0.8788 while average accuracy of SMO-SVM classifier for classifying Clusters of MCs from ROC curves is 0.8752. Fig. 6 illustrates common ROC curve obtained from 10 random experimental trials.

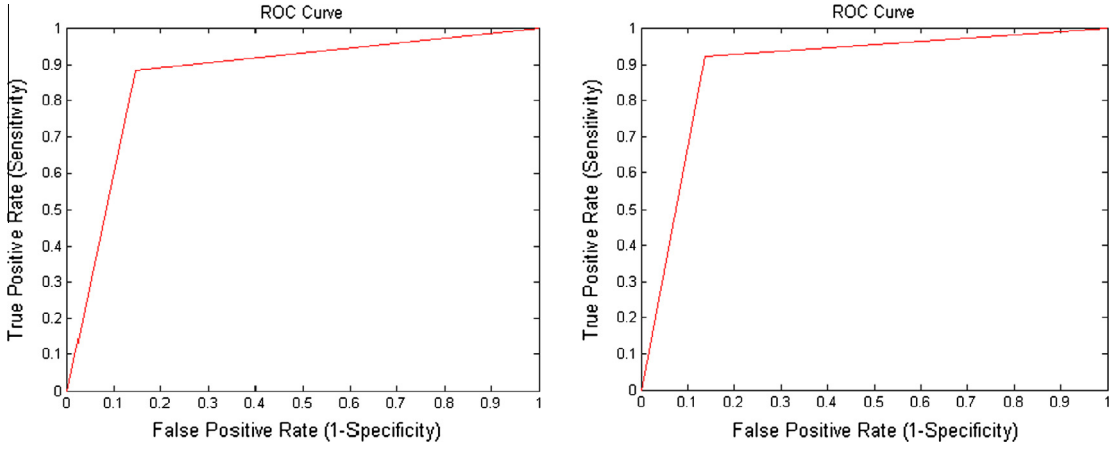

Figure 4 ROC curves of 1st and 10th random experimental trials of SMO-SVM with linear kernel function classifier for classifying MCCs as benign and malignant.

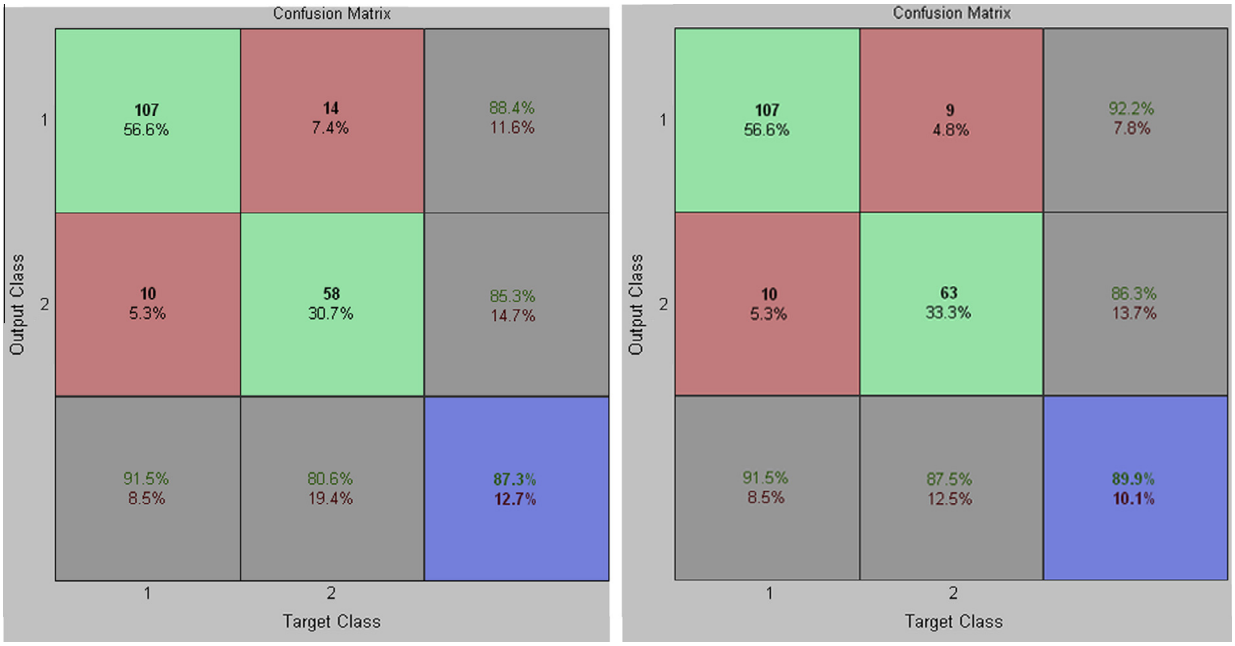

Figure 5 Confusion matrices of 1st and 10th random experimental trials of SMO-SVM with linear kernel function classifier for classifying MCCs as benign and malignant.

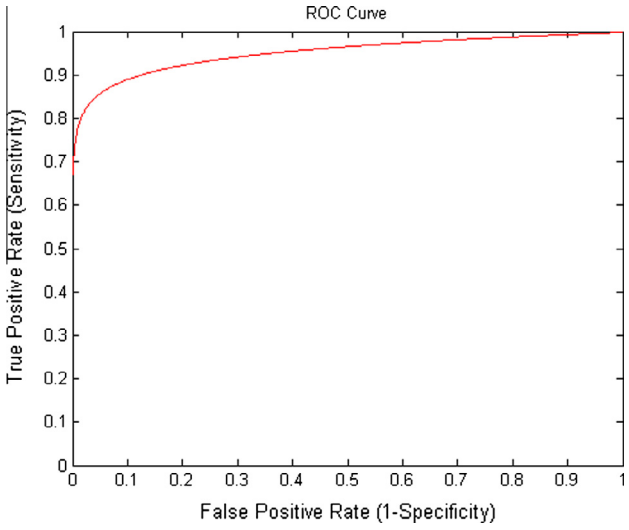

Figure 6 Common ROC curve of 10 random experimental trials of SMO-SVM with linear kernel function for classifying MCCs as benign and malignant.

Table 5 Average accuracy of SMO-SVM with linear kernel function in terms of confusion matrix and ROC analysis.

Average accuracy from confusion matrices	Average accuracy from ROC curves	Accuracy from common ROC curve	Overall accuracy
0.8788	0.8752	0.9509	0.9016

Accuracy of SMO-SVM from ROC curve in terms of area under ROC curve is 0.9509. Thus, the overall accuracy of SVM with linear kernel function and SMO hyperplane finding method for classifying MCCs as benign or malignant is 0.9016 that is shown in Table 5.

6. Conclusion and future work

In this paper, an attempt is made to compare MLFFB-ANN and SVM classifiers to classify MCCs as benign or malignant. For this purpose, Levenberg-Marquardt training algorithm for MLFFB-ANN and Sequential Minimal Optimization hyperplane finding method with linear kernel function for SVM are investigated. For this investigation, 10 random experiment trials are performed for LM-MLFFBP-ANN and SMO-SVM classifiers to classify MCCs as benign or malignant. From these experimental results, it is observed that LM-MLFFBP-ANN classifier belongs to good classifier category according to Hosmer and Lemeshow's rule, while linear kernel function with SMO method based SVM classifier belongs to the excellent classifier category. Results of this study are quite promising for selecting a suitable classifier to classify MCCs as benign or malignant. Based on the results of simulation studies and experiments performed in this study, it is concluded that linear kernel function with SMO method based SVM classifier can be used as a classifier to classify MCCs as benign or malignant for achieving highest accuracy. This research work is very useful for radiologists to characterize clusters of MCs in mammogram.

The results of the mentioned classifiers are encouraging and show good accuracy within experimental errors. But, in future to achieve above 91% overall accuracy, metaheuristic approaches can also be used to find the optimal hyperplane along with different kernel functions in SVM.

References

[1] Hassanien AE. Fuzzy rough sets hybrid scheme for breast cancer detection. Image Vis Comput 2007;25(2):172–83.
[2] Fu JC, Lee SK, Wong ST, Yeh JY, Wang AH, Wu HK. Image segmentation features selection and pattern classification for mammographic microcalcifications. Comput Med Imag Graph 2005;29(6):419–29.
[3] Kramer D, Aghdasi F. Classifications of microcalcifications in digitized mammograms using multiscale statistical texture analysis. In: Proc. South African symposium on communications and signal processing (COMSIG-98), Rondebosch, South African, 7–8 Sep.; 1998. p. 121–6.
[4] Bruce LM, Adhami RR. Classifying mammographic mass shapes using the wavelet transform modulus-maxima method. IEEE Trans Med Imag 1999;18(12):1170–7.
[5] Bottema MJ, Slavotinek JP. Detection and classification of lobular and DCIS (small cell) microcalcifications in digital mammograms. Pattern Recogn Lett 2000;21(13–14):1209–14.
[6] Nicandro C-R, Hector GA-M, Humberto C-C, Luis AN-F, Rocio EB-M. Diagnosis of breast cancer using Bayesian networks: a case study. Comput Biol Med 2007;37(11):1553–64.
[7] Christoyianni I, Koutras A, Dermatas E, Kokkinakis G. Computer aided diagnosis of breast cancer in digitized mammograms. Comput Med Imag Graph 2002;26(5):309–19.
[8] Halkiotis S, Botsis T, Rangoussi M. Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks. Signal Process 2007;87():1559–68.
[9] Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA, Tourassi GD. Training neural network classifiers for medical decision making: the effect of imbalanced datasets on classification performance. Neural Netw 2008;21(2–3):427–36.
[10] Delen D, Walker D, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med 2005;34(2):113–27.
[11] Singh AP, Kamal TS, Kumar S. Virtual curve tracer for estimation of static response characteristics of transducers. Measurement 2005;38(2):166–75.
[12] Varela C, Tahoces PG, Mendez AJ, Souto M, Vidal JJ. Computerized detection of breast masses in digitized mammograms. Comput Biol Med 2007;37(2):214–26.
[13] Arodz T, Kurdziel M, Sevre EOD, Yuen DA. Pattern recognition techniques for automatic detection of suspicious-looking anomalies in mammograms. Comput Methods Programs Biomed 2005;79(2):135–49.
[14] El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. IEEE Trans Med Imag 2002;21(12):1552–63.
[15] Mavroforakis ME, Georgiou HE, Dimitropoulos N, Cavouras D, Theodoridis S. Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers. Artif Intell Med 2006;37():145–62.
[16] Wei L, Yang Y, Nishikawa RM. Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. Pattern Recogn 2009;42(6):1126–32.
[17] http://marathon.csee.usf.edu/Mammography/Database.html.

[18] Widrow B, Lehr MA. 30 Years of adaptive neural networks: perceptron, machine and backpropagation. Proc IEEE 1990;78 ():1415–42.
[19] Haykin S. Neural networks: a comprehensive foundation. 2nd ed. Delhi, India: Pearson Education, Inc.; 2004, p. 202.
[20] Bhowmick B, Pal NR, Pal S, Patel SK, Das J. Detection of microcalcification with neural networks. In: Proc. IEEE international conference on engineering of intelligent systems (ICEIS-06), Islamabad, Pakistan, 22–23 Apr.; 2006. p. 281–6.
[21] Cheng HD, Cui M. Mass lesion detection with a fuzzy neural network. Pattern Recogn 2004;37(6):1189–200.
[22] Haykin S. Neural networks and learning machines. 3rd ed. New Delhi: PHI; 2010, p. 691.
[23] Fausett L. Fundamentals of neural networks: architectures, algorithms, and applications. Englewood Cliffs (NJ, USA): Prentice-Hall, Inc.; 1994, p. 289.
[24] Hagan MT, Demuth HB, Beale M. Neural network design. USA: Thomson Learning; 1996, p. 2–11.
[25] Vapnik VN. The nature of statistical learning theory. 2nd ed. NY (USA): Springer-Verlag; 2000, p. 131.
[26] Rardin RL. Optimization in operation research. 2nd ed. Delhi, India: Pearson Education, Inc.; 2003, p. 810.
[27] Deb K. Optimization for engineering design: algorithms and examples. New Delhi, India: PHI; 2003, p. 77.
[28] Taha HA. Operation research: an introduction. 7th ed. New Delhi, India: Pearson Education, Inc.; 2006, p. 765.
[29] Burges CJC. A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 1998;2(2):121–67.
[30] Fletcher R. Practical methods of optimization. 2nd ed. NY (USA): John Wiley & Sons, Inc.; 1987, p. 219.
[31] Ayat NE, Cheriet M, Suen CY. Automatic model selection for the optimization of SVM kernels. Pattern Recogn 2005;38(10):1733–45.
[32] Karatzoglou A, Meyer D, Hornik K. Support vector machines in R. J Stat Softw 2006;15(9):1–28.
[33] Malon C, Uchida S, Suzuki M. Mathematical symbol recognition with support vector machines. Pattern Recogn Lett 2008;29 ():1326–32.
[34] Platt JC. Fast training of support vector machines using sequential minimal optimization. In: Scholkopf Bernhard, Burges Christopher JC, Smola Alexander J, editors. Advances in Kernel methods: support vector learning. Cambridge (MA, USA): MIT Press; 1999. p. 185–208.
[35] Kohavi R, Provost F. Glossary of terms. J Mach Learn-Spec Issue Appl Mach Learn Knowl Discov Process 1998;30(2–3):271–4.
[36] Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett 2006;27(8):861–74.
[37] Karnan M, Thangavel K. Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications. Comput Methods Programs Biomed 2007;87(1):12–20.
[38] Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. NJ (USA): John Wiley & Sons, Inc.; 2000, p. 164.
[39] Khehra BS, Pharwaha APS. Least-squares support vector machine for characterization of clusters of microcalcifications. World Acad Sci, Eng Technol Int J Comput, Inform Sci Eng 2013;7(12):932–41.