International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:12, 2013

Least-Squares Support Vector Machine for Characterization of Clusters of Microcalcifications

Baljit Singh Khehra, Amar Partap Singh Pharwaha

***Abstract*—**Clusters of Microcalcifications (MCCs) are most frequent symptoms of Ductal Carcinoma in Situ (DCIS) recognized by mammography. Least-Square Support Vector Machine (LS-SVM) is a variant of the standard SVM. In the paper, LS-SVM is proposed as a classifier for classifying MCCs as benign or malignant based on relevant extracted features from enhanced mammogram. To establish the credibility of LS-SVM classifier for classifying MCCs, a comparative evaluation of the relative performance of LS-SVM classifier for different kernel functions is made. For comparative evaluation, confusion matrix and ROC analysis are used. Experiments are performed on data extracted from mammogram images of DDSM database. A total of 380 suspicious areas are collected, which contain 235 malignant and 145 benign samples, from mammogram images of DDSM database. A set of 50 features is calculated for each suspicious area. After this, an optimal subset of 23 most suitable features is selected from 50 features by Particle Swarm Optimization (PSO). The results of proposed study are quite promising.

***Keywords*—**Clusters of Microcalcifications, Ductal Carcinoma in Situ, Least-Square Support Vector Machine, Particle Swarm Optimization.

I. Introduction

BREAST cancer is the major occurrence of cancer among women in both developed and developing countries. Still, there is no known way of preventing it but early detection of breast cancer is the key to improving breast cancer prognosis. Mammography is one of the most effective tools in early detection of breast cancer [1]. It is reliable, low cost and highly sensitive method. Mammography offers high quality images at low x-rays radiation doses. Mammography uses low energy x-rays that pass through the compressed breast of patient and are absorbed by film during an examination. Mammography is the only widely accepted imaging method for routine breast cancer screening. It is recommended that women at the ages of 40 or above should have a mammogram every one to two years [2]. Although mammography is widely used around the world for breast cancer detection, there are some difficulties when mammograms are searched for signs of abnormality by expert radiologists. One of the difficulties is that some anomalies may be missed due to human error or as a result of fatigue. This is one of the main reasons of false positive and false negative readings of mammogram. False positive detection causes unnecessary biopsy. It has been estimated that only 15-30% of breast biopsy cases are proved to be cancerous [3]. On the other hand, in a false negative detection, an actual tumor remains undetected. Retrospective studies [4] have shown that 10-30% of the visible cancers are undetected. So, false positive and false negative have caused a high proportion of women without cancer to undergo breast biopsies or miss the best treatment time. Thus, there is a significant necessity to improve the correct diagnosis rate of cancer. Several solutions were proposed in the past to increase accuracy and sensitivity of mammography and reduce unnecessary biopsies. Independent double reading of mammograms by two radiologists is one of the solutions and has proved effective in significantly increasing the sensitivity of mammographic screening [5]. The basic idea of independent double reading is for the mammograms to be read by two radiologists independently. However, this solution is both highly costly and time consuming. Instead of double reading, radiologists have an opportunity to improve their diagnosis with the aid of computer system. It might provide a useful second opinion to radiologists during mammographic interpretation.

The classifier plays an important role in the implementation of Computer-Aided Diagnosis (CAD) of mammography. It is last phase of a CAD scheme that is aimed at assisting radiologists in making more accurate diagnosis of breast cancer on mammograms [6]. The classifier makes the decision regarding the region of suspicion. The aim of the classification stage of CAD system is the characterization of each cluster as malignant or benign using the selected features. To evaluate the performance of classifier for classifying MCCs as benign and malignant, confusion matrix and Receiving Operating Characteristics (ROC) analysis are used.

A number of classifiers have been proposed for CAD system to classify MCCs as benign or malignant. Kramer and Aghdasi [7] used K-Nearest Neighbor (KNN) classifier to classify MCs in digitized mammograms using multi-scale statistical texture analysis. Bruce and Adhami [8] used Linear Discriminant Analysis (LDA) to classify mammographic masses into three classes: stellate, nodular and round. Bottema and Slavotinek [9] used decision trees for classification of lobular and DCIS (small cell) MCs in digital mammograms. In 2007, Nicandro et al. [10] used Bayesian network classifiers for the diagnosis of breast cancer. Hassanien [11] used fuzzy rough sets hybrid scheme for breast cancer detection. Artificial Neural Networks (ANNs) have also been widely used for classification of MCs as benign or malignant [12]–[14]. SVM

Baljit Singh Khehra is with the Department of Computer Science & Engineering, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib-140407, Punjab, India (phone: +919463446505; e-mail: baljitkhehra@rediffmail.com, baljitkhehra74@gmail.com).

Amar Partap Singh Pharwaha is with the Department of Electronics and communications Engineering, Sant Longowal Institute of Engineering and Technology, Longowal-148106, Sangrur, Punjab, India (phone: +919463122255; e-mail: amarpartapsingh@yahoo.com).

is a supervised learning two-class classifier derived from statistical learning theory. It is developed by Vapnik [15]. It is based on the principle of structural risk minimization. SVM has been most recently used in many applications [16]–[20]. LS-SVM is a variant of the standard SVM. It is proposed by Suykens and Vandewalle [21]. In the proposed research work, LS-SVM is explored for classifying MCCs as benign or malignant task from various kernel function point of view.

II. Description of LS-SVM for Classification of MCCs

When SVM is used for classification of the two-class pattern classification problem, the aim is to find the optimal separating hyperplane that separates these two classes [22]. Classification of MCCs as benign or malignant is a two-class pattern classification problem. To begin, let $\{(X_j, Y_j), j=1, 2, \ldots\ldots\ldots, L\}$ be a set of L training data samples, where $X_j \in R^d$ are L data points, each of which belong to class $Y_j \in \{+1, -1\}$. In classification of MCCs as benign or malignant problem, vector $X \in \{x_i, i=1, 2,\ldots\ldots\ldots,n\}$ denotes a cluster of MCs to be classified, where $\{x_i, i=1, 2,\ldots\ldots\ldots,n\}$ denotes a set of n features of the cluster of MCs, and $Y \in \{+1, -1\}$ denotes its class label i.e. '+1' denotes malignant cluster of MCs and '-1' denotes benign cluster of MCs. An optimal separating hyperplane is a hyperplane that correctly separates the positive and negative classes. An optimization problem for optimal separating hyperplane can be formulated as

Objective function: Minimize $\frac{w^T w}{2}$

Subject to the constraints:

$$Y_j \left(w^T X_j + b \right) \geq 1 \qquad for \qquad \forall j$$

where $w^T X$ denotes dot product between w and X; parameter w is the norm to the hyperplane; $\frac{|b|}{\|w\|}$ is the perpendicular distance from the hyperplane to the origin and $\|w\|$ is the Euclidean norm of w. Hence, the goal is to find w and b such that $\frac{w^T w}{2}$ is minimized and $Y_j \left(w^T X_j + b \right)$ is greater than or equal to 1 for all j. The optimal solution w^*, b^* enables classification of a test example Z as follows:

$$class\ (Z) = sign\ \left(w^{*T} Z + b^* \right) \tag{1}$$

The above said optimization problem is a Quadratic Programming (QP) optimization problem with linear constraints. To solve this, it is necessary to switch on Lagrangian formulation of the problem [23] because when constraints are replaced by constraints on the Lagrange multipliers, then it is much easier to handle. To obtain Lagrangian formulation of the problem, first take positive Lagrange multipliers α_j, $j=1,2,\ldots\ldots\ldots,L$. After this, the constraint equations are multiplied by positive Lagrange multipliers and subtracted from the objective function. Thus, the following objective function is obtained:

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

Objective function: Minimize L_P w. r. t w, b
Subject to the constraints:
(a) Derivatives of L_P w. r. t all α_j vanish
(b) $\alpha_j \geq 0$

Dual problem can be formulated from the above primal problem as follows:
Objective function: Maximize L_P
Subject to the constraints:
(a) Gradient of L_P w. r. t w and b vanish
(b) $\alpha_j \geq 0$

Now, Gradient of L_P w. r. t w and b vanish gives the following conditions:

$$w = \sum_{j=1}^{L} \alpha_j Y_j X_j \tag{3}$$

$$\sum_{j=1}^{L} \alpha_j Y_j = 0 \tag{4}$$

By substituting (3) and (4) into (2), the following equation is obtained:

$$L_D = \sum_{j=1}^{L} \alpha_j - \frac{1}{2} \sum_{j=1}^{L} \sum_{i=1}^{L} \alpha_j \alpha_i Y_j Y_i X_j^T X_i \tag{5}$$

In this case, the problem is formulated as
Objective function: Maximize L_D
Subject to the constraints:

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

(b). $\alpha_j \geq 0$

Thus, the goal is to find $\alpha_1, \alpha_2, \ldots\ldots,\alpha_L$ such that L_D is maximized along with $\sum_{j=1}^{L} \alpha_j Y_j$ is equal to zero and α_j is greater than or equal to 0 for all j. For non-zero α_j^*, $j = 1,2,\ldots\ldots\ldots, L_s$, the optimal values of w^* and b^* are obtained as follows:

$$w^* = \sum_{j=1}^{L_s} \alpha_j^* Y_j X_j \tag{6}$$

$$b^* = Y_j - \sum_{j=1}^{L_s} \alpha_j^* Y_j X_j^T X_j \tag{7}$$

where non-zero Lagrange multipliers, α_j^*, $j = 1,2,\ldots\ldots\ldots, L_s$, indicate their corresponding

support vectors $S_j \in (X_j, Y_j)$. Thus, classification of a test example Z is done as

$$class\ (Z) = sign\left(\sum_{j=1}^{L_s} \alpha_j^* Y_j X_j^T Z + b^*\right) \quad (8)$$

A. Soft-Margin SVM

For soft-margin classification [24], slack variables ζ_j can be added to allow misclassification of difficult or noisy examples. Thus, modified formulation of the primal problem is as follows:

Objective function: Minimize $\frac{1}{2} w^T w + C \sum_{j=1}^{L} \zeta_j$

Subject to the constraints:

$$Y_j\left(w^T X_j + b\right) \geq 1 - \zeta_j \quad for \quad \forall j\ and \quad \zeta_j \geq 0$$

where C is a soft-margin parameter that controls the penalty for misclassifying the training samples. Now, modified formulation of the dual problem is as follows:

Objective function: Maximize L_D

Subject to the constraints:

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

(b) $0 \leq \alpha_j \leq C \quad for \quad \forall \alpha_j$

Similarly as mentioned in the dual problem, now, the goal is to find $\alpha_1, \alpha_2, \ldots\ldots, \alpha_L$ such that L_D is maximized along with $\sum_{j=1}^{L} \alpha_j Y_j$ is equal to zero and $0 \leq \alpha_j \leq C \quad for \quad \forall \alpha_j$.

For non-zero α_j^*, $j = 1, 2, \ldots\ldots\ldots, L_s$, w^* is obtained from (6) and b^* is calculated as follows:

$$b^* = Y_j\left(1 - \zeta_j\right) - \sum_{j=1}^{L_s} \alpha_j^* Y_j X_j^T X_j \quad (9)$$

B. Non-Linear SVM

Non-linear SVM classifier is used to solve non-linear classification problems through a kernel function. Kernel function uses transformation operator $\phi(.)$ to map two classes of training data points in an input space R^d onto a higher dimensional feature space H so that the two classes of training data points can be separated by a hyperplane [25].

$$\phi : R^d \rightarrow H \quad (10)$$

Mainly, kernel function is used to convert non-linear classification problems into linear classification problems. Relation between kernel function $K(X_j, X_i)$ and mapping operator $\phi(.)$ [26] is shown as

$$K\left(X_j, X_i\right) = \phi\left(X_j\right)^T \phi\left(X_i\right) \qquad \forall X_j, X_i \in R^n \quad (11)$$

Thus, dual formulation of problem is as follows:

Find $\alpha_1, \alpha_2, \ldots\ldots, \alpha_L$ such that

$L_D = \sum_{j=1}^{L} \alpha_j - \frac{1}{2} \sum_{j=1}^{L} \sum_{i=1}^{L} \alpha_j \alpha_i Y_j Y_i K\left(X_j, X_i\right)$ is maximized and

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

(b) $0 \leq \alpha_j \leq C \quad for \quad \forall \alpha_j$

For non-zero α_j^*, $j = 1, 2, \ldots\ldots\ldots, L_s$, w^* is obtained from (6) and b^* is calculated as follows:

$$b^* = Y_j - \sum_{j=1}^{L_s} \alpha_j^* Y_j K\left(X_j, X_j\right) \quad (12)$$

Thus, classification of a test example Z is done as

$$class\,(Z) = sign\left(\sum_{j=1}^{L_s} \alpha_j^* Y_j K\left(X_j, Z\right) + b^*\right) \quad (13)$$

Most commonly used kernel functions in SVM [27], [28] are defined as follows:

i. Linear kernel function

$$K(x, y) = x^T y \quad (14)$$

ii. Quadratic kernel function

$$K(x, y) = (1 + x^T y)^2 \quad (15)$$

iii. Gaussian RBF kernel function

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (16)$$

where σ is the kernel width

C. Least-Squares SVM

In LS-SVM, QP problem of the standard soft-margin SVM is transformed into linear problem. This transform is performed [29] as

(i) Slack variables ζ_j of the inequality constraints $Y_j\left(w^T X_j + b\right) \geq 1 - \zeta_j \quad for \quad \forall j\ and \quad \zeta_j \geq 0$ are replaced with error variables e_j;

(ii) the term $C\sum_{j=1}^{L}\zeta_j$ in the objective function $\frac{1}{2}w^T w + C\sum_{j=1}^{L}\zeta_j$ is replaced by $\gamma\frac{1}{2}\sum_{j=1}^{L}e_j^2$, where γ is a tuning parameter. Thus, the formulation of the classification problem for LS-SVM is as follows:

Objective function: Minimize $\frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{j=1}^{L}e_j^2$

Subject to the constraints:

$$Y_j\left(w^T X_j + b\right) \geq 1 - e_j \quad for \quad \forall j$$

Lagrangian formulation of the above said problem is expressed by

$$L_{LS}(w,b,e,\alpha) = \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{j=1}^{L}e_j^2 - \sum_{j=1}^{L}\alpha_j\{Y_j(w^T X_j + b) - 1 + e_j\} \quad (17)$$

According to Karush-Kuhn-Tucker condition [30], the following conditions for optimality [31] are obtained:

$$\frac{\partial L_{LS}}{\partial w} = 0 \quad \Rightarrow \quad w = \sum_{j=1}^{L}\alpha_j Y_j X_j \quad (18)$$

$$\frac{\partial L_{LS}}{\partial b} = 0 \quad \Rightarrow \quad \sum_{j=1}^{L}\alpha_j Y_j = 0 \quad (19)$$

$$\frac{\partial L_{LS}}{\partial e_j} = 0 \quad \Rightarrow \quad \alpha_j = \gamma e_j \quad for \quad \forall j \quad (20)$$

$$\frac{\partial L_{LS}}{\partial \alpha_j} = 0 \quad \Rightarrow \quad Y_j(w^T X_j + b) - 1 + e_j = 0 \quad for \quad \forall j \quad (21)$$

The above set of linear equations is used to find the solution of the problem.

III. Classifier Performance Measures

For evaluating the performance of classifier to classify MCCs as benign and malignant, mainly confusion matrix [32] and ROC analysis [33] are used. A confusion matrix is a table that contains information about actual and predicted classifications done by a classifier. Table I shows confusion matrix.

TABLE I
Confusion Matrix

		Actual		
		Positive	Negative	
Predicted	Positive	TPs	FPs	Positive Predictive Value
	Negative	FNs	TNs	Negative Predictive Value
		Sensitivity	Specificity	Accuracy

In confusion matrix, *Sensitivity*, *Specificity*, *Positive Predictive Value*, *Negative Predictive Value* and *Accuracy* are defined as

$$Sensitivity = \frac{TPs}{TPs + FNs} \quad (22)$$

$$Specificity = \frac{TNs}{TNs + FPs} \quad (23)$$

$$Positive\ Predictive\ Value = \frac{TPs}{TPs + FPs} \quad (24)$$

$$Negative\ Predictive\ Value = \frac{TNs}{TNs + FNs} \quad (25)$$

$$Accuracy = \frac{TPs + TNs}{TPs + FPs + TNs + FNs} \quad (26)$$

where, *TPs*, *TNs*, *FNs* and *FPs* are number of true positive decisions, number of true negative decisions, number of false negative decisions and number of false positive decisions taken by a classifier respectively.

ROC analysis is based on statistical decision theory that has been widely used in medical decision making. In ROC analysis, ROC curve is a popular tool to measure classifier performance in CAD system. ROC curve is a plot of classifier's *Sensitivity* versus its *1- Specificity* at all possible threshold values. To draw ROC curve, x-axis is *1-Specificity* and y-axis is *Sensitivity*. The terms *Sensitivity*, *Specificity* and *1- Specificity* are synonymous with *True Positive Rate*, *True Negative Rate* and *False Positive Rate* respectively. ROC curve depicts the tradeoffs between *True Positive Rate* and *False Positive Rate* to describe the inherent discrimination capacity of CAD system. Area under the ROC curve (A_Z) is an important criterion for evaluating diagnostic performance [34]. The ROC curve is in the range between 0.0 and 1.0. So, A_Z lies between 0.0 and 1.0. The value of A_Z is equal to 1.0 when CAD system has perfect performance i.e. *True Positive Rate* is 100% and *False Positive Rate* is 0%. The value of A_Z can be computed by Trapezoidal rule or Simpson's rule.

Hosmer and Lemeshow [35] gave the following general rule to measure the discrimination power of a classifier based on ROC curve:

- If $0.5 \leq Az < 0.6$: This means no discrimination i.e. fail classifier
- If $0.6 \leq Az < 0.7$: This means poor discrimination i.e. poor classifier
- If $0.7 \leq Az < 0.8$: This means fair discrimination i.e. fair classifier
- If $0.8 \leq Az < 0.9$: This means good discrimination i.e. good classifier
- If $0.9 \leq Az \leq 1.0$: This means excellent discrimination i.e. excellent classifier

IV. Experimental Results and Discussion

In order to establish the credibility of LS-SVM classifier for classifying MCCs as benign or malignant, a comparative evaluation of the relative performance of LS-SVM classifier for different kernel functions is made. For this comparative evaluation, experiments are performed on data extracted from mammogram images of DDSM database [36]. All experiments are conducted on MATLAB 7.7 software. Confusion matrix and ROC analysis are used to measure the performance of classifier. In this study, a total of 380 suspicious areas are collected, which contain 235 malignant and 145 benign samples, from mammogram images of DDSM database. It is a standard benchmark database for mammographic image analysis research community that is maintained at the University of South Florida. 50 features are extracted for each suspicious region of mammograms. Such features are shown in Appendix. After this, an optimal subset of 23 features is selected by Particle Swarm Optimization (PSO) method [37]. In this study, linear, quadratic and Gaussian radial basis kernel functions are considered. 191 samples are randomly selected from 380 samples for training purpose and the remaining samples are used for testing. The training samples are not used during the testing stage.

First, linear kernel function is chosen and 10 experiments are run to measure the performance of LS-SVM with it for classifying MCCs. Tabular results of 10 random experimental runs of LS-SVM with linear kernel function are shown in Table II. Due to page constraint, confusion matrices and ROC curves of only 1st and last experimental runs are shown in Figs. 1 and 2, respectively. The average accuracy from confusion matrices is 0.8884 while average accuracy from ROC curves is 0.8867 with a sensitivity of 0.9274 and a specificity of 0.8250. A common ROC curve of 10 random experimental trials is plotted by logarithmic function. Such curve is shown in Fig. 3. Area under this ROC curve is 0.9398. Thus, the overall accuracy of LS-SVM with linear kernel function is 0.9050 that is shown in Table V.

TABLE II
Tabular Results of 10 Random Experimental Trials of LS-SVM with Linear Kernel Function for Classifying MCCs as Benign or Malignant

Random Experimental Trial No.	Confusion Matrix		Accuracy from Confusion Matrix	Accuracy from A_z	Sensitivity	Specificity
1	105	10	0.8836	0.8754	0.8974	0.8611
	12	62				
2	110	15	0.8836	0.8853	0.9402	0.7917
	7	57				
3	110	9	0.9153	0.9122	0.9402	0.8750
	7	63				
4	109	17	0.8677	0.8690	0.9316	0.7639
	8	55				
5	110	17	0.8730	0.8766	0.9402	0.7639
	7	55				
6	111	15	0.8889	0.8929	0.9487	0.7917
	6	57				
7	110	12	0.8995	0.8986	0.9402	0.8333
	7	60				
8	108	10	0.8995	0.8942	0.9231	0.8611
	9	62				
9	102	8	0.8783	0.8687	0.8718	0.8889
	15	64				
10	110	13	0.8942	0.8941	0.9402	0.8194
	7	59				
Mean			0.8884	0.8867	0.9274	0.8250
S. D.			0.0142	0.0142	0.0243	0.0459

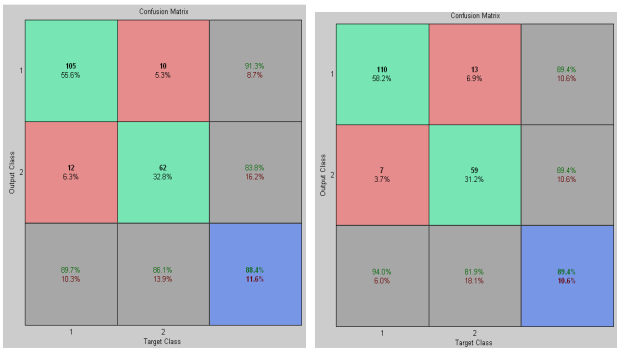

Fig. 1 Confusion matrices for 1st and last experimental trials of LS-SVM with Linear Kernel function to differentiate benign and malignant MCCs

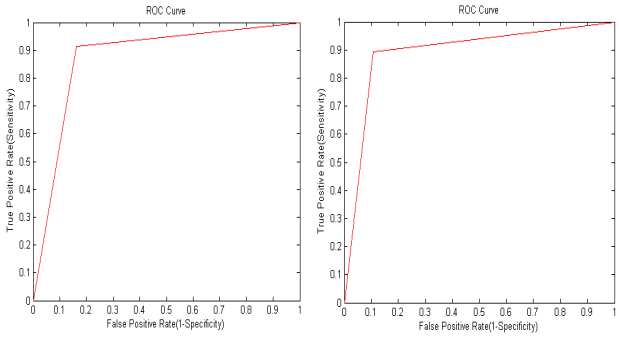

Fig. 2 ROC curves for 1st and last experimental trials illustrate the performance of LS-SVM with Linear Kernel function to differentiate benign and malignant MCCs

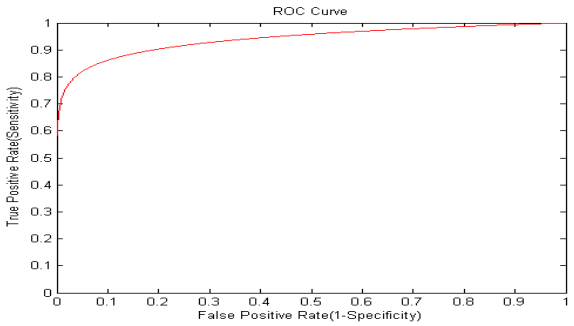

Fig. 3 Common ROC curve for 10 random experimental trials illustrates the performance of LS-SVM with Linear Kernel function to differentiate benign and malignant MCCs

After the performance analysis of linear kernel function LS-SVM, quadratic kernel function is considered. Tabular results of 10 random experimental runs of this classifier are shown in Table III. Confusion matrices and ROC curves of 1st and last experimental runs are shown in Figs. 4 and 5, respectively. The average accuracy from confusion matrices is 0.8371 while average accuracy from ROC curves is 0.8270 with a sensitivity of 0.8521 and a specificity of 0.8125. A common ROC curve obtained from 10 random experimental trials is shown in Fig. 6. Area under this ROC curve is 0.8769. Thus, the overall accuracy of LS-SVM with quadratic kernel function is 0.8470 that is shown in Table V.

TABLE III
TABULAR RESULTS OF 10 RANDOM EXPERIMENTAL TRIALS OF LS-SVM WITH QUADRATIC KERNEL FUNCTION FOR CLASSIFYING MCCS AS BENIGN OR MALIGNANT

Random Experimental Trial No.	Confusion Matrix		Accuracy from Confusion Matrix	Accuracy from A_Z	Sensitivity	Specificity
1	100	14	0.8360	0.8253	0.8547	0.8056
	17	58				
2	98	11	0.8413	0.8308	0.8376	0.8472
	19	61				
3	102	15	0.8413	0.8317	0.8718	0.7917
	15	57				
4	101	16	0.8307	0.8205	0.8632	0.7778
	16	56				
5	100	14	0.8360	0.8253	0.8547	0.8056
	17	58				
6	99	14	0.8307	0.8196	0.8462	0.8056
	18	58				
7	97	11	0.8360	0.8256	0.8291	0.8472
	20	61				
8	95	10	0.8307	0.8214	0.8120	0.8611
	22	62				
9	104	16	0.8466	0.8391	0.8889	0.7778
	13	56				
10	101	14	0.8413	0.8310	0.8632	0.8056
	16	58				
Mean			0.8371	0.8270	0.8521	0.8125
S. D.			0.0055	0.0061	0.0221	0.0294

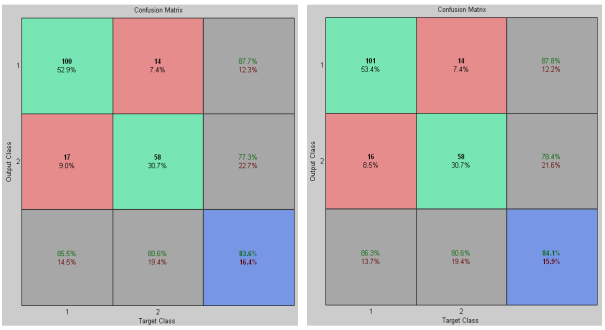

Fig. 4 Confusion matrices for 1st and last experimental trials of SVM with Quadratic Kernel function and LS method to differentiate benign and malignant MCCs

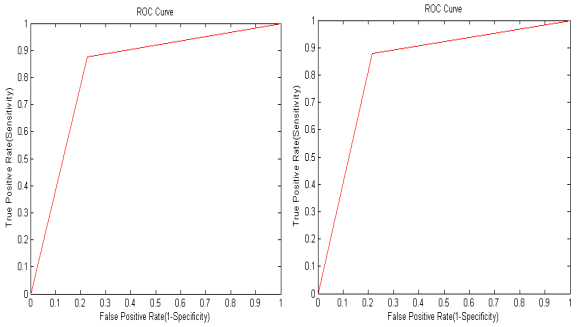

Fig. 5 ROC curves for 1st and last experimental trials illustrate the performance of LS-SVM with Quadratic Kernel function to differentiate benign and malignant MCCs

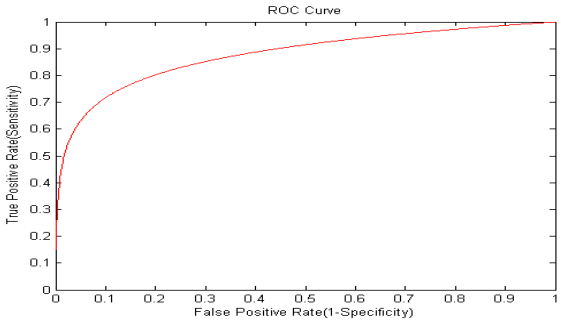

Fig. 6 Common ROC curve for 10 random experimental trials illustrates the performance of LS-SVM with Quadratic Kernel function to differentiate benign and malignant MCCs

At the end, Gaussian radial basis kernel function is considered. In Gaussian radial basis kernel function, the value of sigma (σ) is set to 2. Tabular results of 10 random experimental runs of this classifier are shown in Table IV. Confusion matrices and ROC curves of 1st and last experimental runs are shown in Figs. 7 and 8, respectively. The average accuracy from confusion matrices is 0.8656 while average accuracy from ROC curves is 0.8676 with a sensitivity of 0.9282 and a specificity of 0.7639. A common ROC curve obtained from 10 random experimental trials is shown in Fig. 9. Area under this ROC curve is 0.8717. Thus, the overall accuracy of SVM with Gaussian radial basis kernel function is 0.8683 that is shown in Table V.

TABLE IV
TABULAR RESULTS OF 10 RANDOM EXPERIMENTAL TRIALS OF LS-SVM WITH GAUSSIAN RADIAL BASIS KERNEL FUNCTION FOR CLASSIFYING MCCS AS BENIGN OR MALIGNANT

Random Experimental Trial No.	*Confusion Matrix*		*Accuracy from Confusion Matrix*	*Accuracy from A_Z*	*Sensitivity*	*Specificity*
1	108	15	0.8730	0.8708	0.9231	0.7917
	9	57				
2	108	19	0.8519	0.8526	0.9231	0.7361
	9	53				
3	111	20	0.8624	0.8719	0.9487	0.7222
	6	52				
4	108	15	0.8730	0.8708	0.9231	0.7917
	9	57				
5	110	17	0.8730	0.8766	0.9402	0.7639
	7	55				
6	107	16	0.8624	0.8592	0.9145	0.7778
	10	56				
7	113	21	0.8677	0.8853	0.9658	0.7083
	4	51				
8	111	19	0.8677	0.8761	0.9487	0.7361
	6	53				
9	103	12	0.8624	0.8532	0.8803	0.8333
	14	60				
10	107	16	0.8624	0.8592	0.9145	0.7778
	10	56				
Mean			0.8656	0.8676	0.9282	0.7639
S. D.			0.0067	0.0109	0.0239	0.0382

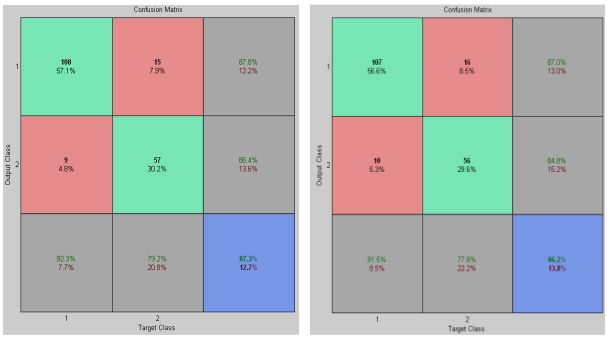

Fig. 7 Confusion matrices for 1st and last experimental trials of LS-SVM with Gaussian Radial Basis Kernel function to differentiate benign and malignant MCCs

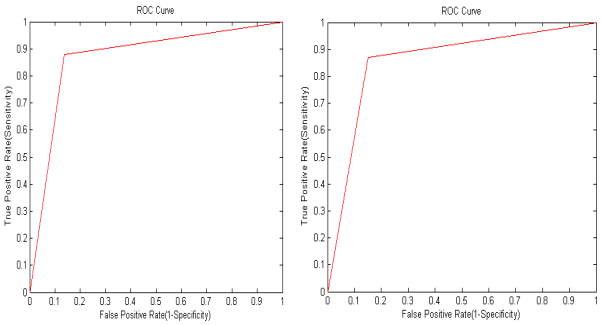

Fig. 8 ROC curves for 1st and last experimental trials illustrate the performance of LS-SVM with Gaussian Radial Basis Kernel function to differentiate benign and malignant MCCs

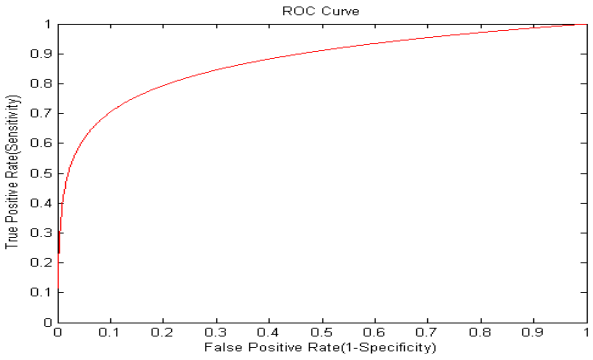

Fig. 9 Common ROC curve for 10 random experimental trials illustrates the performance of LS-SVM with Gaussian Radial Basis Kernel function to differentiate benign and malignant MCCs

TABLE V
PERFORMANCE OF LS-SVM WITH DIFFERENT KERNEL FUNCTIONS FOR DIFFERENTIATING BENIGN AND MALIGNANT MCCS

Kernel Function	Average Accuracy from Confusion Matrices	Average Accuracy from ROC Curves	Accuracy from Common ROC Curve	Overall Accuracy
Linear	0.8884	0.8867	0.9398	0.9050
Quadratic	0.8371	0.8270	0.8769	0.8470
Gaussian Radial Basis	0.8656	0.8676	0.8717	0.8683

V. CONCLUSION AND FUTURE SCOPE

In this paper, an attempt is made to establish the credibility of LS-SVM classifier with different kernel functions for classifying MCCs as benign or malignant. For this purpose, three kernel functions: linear, quadratic and Gaussian radial basis are used. For this study, experiments are conducted on MATLAB 7.7 software. Experiments are performed on 380 suspicious areas collected from mammogram images of DDSM database. A set of 23 most suitable features selected by PSO is used. The performance of the classifier is measured from confusion matrix and ROC analysis. From the experimental results, it is observed that the overall correct classification rate of LS-SVM with linear, quadratic and Gaussian radial basis kernel functions is 90.50%, 84.70% and 86.83%, respectively. From these experimental results, it is observed that LS-SVM with linear kernel function belongs to excellent classifiers category according to Hosmer and Lemeshow's rule, while LS-SVM with quadratic and Gaussian radial basis kernel functions belongs to good classifiers category. Results of this study are quite promising. The proposed work can be a part of developing a CAD system for early detection of breast cancer. Thus, this research work could, in a way, significantly contribute towards eventually detecting DCIS type breast cancer which is the main challenge for radiologists.

Although the achieved performance is satisfactory for mammogram images of DDSM databases, further testing can also be performed on the mammogram images of other standard databases.

APPENDIX

TABLE VI
EXTRACTED FEATURES FROM MAMMOGRAM

Feature No.	Feature
1	Mean from Gray Level Histogram Moments Method
2	Standard Deviation from Gray Level Histogram Moments Method
3	Relative Smoothness from Gray Level Histogram Moments Method
4	3rd Moment from Gray Level Histogram Moments Method
5	4th Moment from Gray Level Histogram Moments Method
6	Uniformity from Gray Level Histogram Moments Method
7	Havrda and Charvat Entropy from Gray Level Histogram Moments Method
8	Maximum Probability from Gray Level Co-occurrence Matrix (GLCM)
9	Contrast from GLCM
10	Energy from GLCM
11	Homogeneity from GLCM
12	Correlation from GLCM
13	Sum Average from GLCM
14	Sum of Squares: Variance from GLCM
15	Sum Variance from GLCM
16	Difference Variance from GLCM
17	Information Measure of Correlation 1st from GLCM
18	Information Measure of Correlation 2nd from GLCM
19	Havrda and Charvat Entropy for GLCM from GLCM
20	Havrda and Charvat Sum Entropy from GLCM

21	Havrda and Charvat Difference Entropy from GLCM
22	Average from Window based Statistical Texture Feature Extraction Method
23	Stand Deviation from Window based Statistical Texture Feature Extraction Method
24	Relative Smoothness from Window based Statistical Texture Feature Extraction Method
25	Skewness from Window based Statistical Texture Feature Extraction Method
26	Kurtosis from Window based Statistical Texture Feature Extraction Method
27	Bussyness from Window based Statistical Texture Feature Extraction Method
28	Potential of a point from Window based Statistical Texture Feature Extraction Method
29	Point Mask from Window based Statistical Texture Feature Extraction Method
30	Average Energy from Window based Statistical Texture Feature Extraction Method
31	Energy Variance from Window based Statistical Texture Feature Extraction Method
32	Volatility from Window based Statistical Texture Feature Extraction Method
33	Mean from Wavelet Domain
34	Standard Deviation from Wavelet Domain
35	Spectral Entropy from Wavelet Domain
36	Mean from Fourier Domain
37	Standard Deviation from Fourier Domain
38	Spectral Entropy from Fourier Domain
39	Mean of Areas
40	Standard Deviation of Areas
41	Maximum Area
42	Minimum Area
43	Mean of Compactness
44	Standard Deviation of Compactness
45	Maximum Compactness
46	Minimum Compactness
47	Mean of 2nd Central Moment
48	Standard Deviation of 2nd Central Moment
49	Maximum 2nd Central Moment
50	Minimum 2nd Central Moment

ACKNOWLEDGMENT

Authors are greatly indebted to the Department of Electronics and Communications Engineering, SLIET, Longowal-148106 (District: Sangrur) Punjab for providing excellent lab facilities that make this work feasible.

REFERENCES

[1] Hassanien, A. E., "Fuzzy rough sets hybrid scheme for breast cancer detection," *Image and Vision Computing*, vol. 25, no. 2, pp. 172-183, 2007.

[2] Xu, J. and Tang, J., "Detection of clustered microcalcifications using an improved texture based approach for computer aided breast cancer diagnosis system," *CSI Communications*, vol. 31, no. 10, pp. 17-20, 2008.

[3] Sickles, E. A., "Breast calcifications: mammographic evaluation," *Radiology*, vol. 160, no. 2, pp. 289-293, 1986.

[4] Wallis, M. G., Walsh, M. T. and Lee, J. R., "A review of false negative mammography in a symptomatic population," *Clinical Radiology*, vol. 144, no. 1, pp. 13-15, 1991.

[5] Varela, C., Tahoces, P. G., Mendez, A. J., Souto, M. and Vidal, J. J., "Computerized detection of breast masses in digitized mammograms," *Computers in Biology and Medicine*, vol. 37, no. 2, pp. 214-226, 2007.

[6] Wei, L., Yang, Y., Nishikawa, R. M. and Jiang, Y., "A study of several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Trans. Medical Imaging*, vol. 24, no. 3, pp. 371-380, 2005.

[7] Kramer, D. and Aghdasi, F., "Classifications of microcalcifications in digitized mammograms using multiscale statistical texture analysis," Proc. South African Symposium on Communications and Signal Processing (COMSIG-98), Rondebosch, South African, 7-8 Sep. 1998, pp. 121-126.

[8] Bruce, L. M. and Adhami, R. R., "Classifying mammographic mass shapes using the wavelet transform modulus-maxima method," *IEEE Trans. Medical Imaging*, vol. 18, no. 12, pp. 1170-1177, 1999.

[9] Bottema, M. J. and Slavotinek, J. P., "Detection and classification of lobular and DCIS (small cell) microcalcifications in digital mammograms," *Pattern Recognition Letters*, vol. 21, no. 13-14, pp. 1209-1214, 2000.

[10] Nicandro, C.-R., Hector, G. A.-M., Humberto, C.-C., Luis, A. N.-F. and Rocio, E. B.-M., "Diagnosis of breast cancer using Bayesian networks: a case study," *Computers in Biology and Medicine*, vol. 37, no. 11, pp. 1553-1564, 2007.

[11] Hassanien, A. E., "Fuzzy rough sets hybrid scheme for breast cancer detection," *Image and Vision Computing*, vol. 25, no. 2, pp. 172-183, 2007.

[12] Christoyianni, I., Koutras, A., Dermatas, E. and Kokkinakis, G., "Computer aided diagnosis of breast cancer in digitized mammograms," *Computerized Medical Imaging and Graphics*, vol. 26, no. 5, pp. 309-319, 2002.

[13] Halkiotis, S., Botsis, T. and Rangoussi, M., "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks," *Signal Processing*, vol. 87, no. 7, pp. 1559-1568, 2007.

[14] Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A. and Tourassi, G. D., "Training neural network classifiers for medical decision making: the effect of imbalanced datasets on classification performance," *Neural Networks*, vol. 21, no. 2-3, pp. 427-436, 2008.

[15] Vapnik, V.N., *The nature of statistical learning theory*, 2nd ed. Springer-Verlag, NY, USA, 2000, p 131.

[16] Arodz, T., Kurdziel, M., Sevre, E. O. D. and Yuen, D. A., "Pattern recognition techniques for automatic detection of suspicious-looking anomalies in mammograms," *Computer Methods and Programs in Biomedicine*, vol. 79, no. 2, pp. 135-149, 2005.

[17] El-Naqa, I., Yang, Y., Wernick, M. N., Galatsanos, N. P. and Nishikawa, R. M., "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Medical Imaging*, vol. 21, no. 12, pp. 1552-1563, 2002.

[18] Fu, J.C., Lee, S. K., Wong, S. T., Yeh, J. Y., Wang, A. H. and Wu, H. K., "Image segmentation features selection and pattern classification for mammographic microcalcifications," *Computerized Medical Imaging and Graphics*, vol.29, no. 6, pp. 419-429, 2005.

[19] Mavroforakis, M. E., Georgiou, H. E., Dimitropoulos, N., Cavouras, D. and Theodoridis, S., "Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers," *Artificial Intelligence in Medicine*, vol. 37, no. 2, pp. 145-162, 2006.

[20] Wei, L., Yang, Y. and Nishikawa, R.M., "Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis," *Pattern Recognition*. vol. 42, no. 6, pp. 1126–1132, 2009.

[21] Suykens, J. A. K. and Vandewalle, J., " Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293-300, 1999.

[22] Vanschoenwinkel, B. and Manderick, B., "Appropriate kernel functions for support vector machine learning with sequences of symbolic data," *Deterministic and Statistical Methods in Machine Learning* (Eds. Joab Winkler, Mahesan Niranjan and Neil Lawrence), Lecture Notes in Computer Science, 3635, Springer-Verlag, Berlin, Germany, pp. 256-280, 2005.

[23] Rardin, R. L., *Optimization in Operation Research*, 2nd ed. Pearson Education, Inc., Delhi, India, 2003, p 810.

[24] Cortes, C. and Vapnik, V., "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[25] Burges, C. J. C., "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.

[26] Ayat, N. E., Cheriet, M. and Suen, C.Y., "Automatic model selection for the optimization of SVM kernels," *Pattern Recognition*, vol. 38, no. 10, pp. 1733-1745, 2005.
[27] Karatzoglou, A., Meyer, D. and Hornik, K., "Support vector machines in R," *Journal of Statistical Software*, vol. 15, no. 9, pp. 1-28, 2006.
[28] Malon, C., Uchida, S. and Suzuki, M., "Mathematical symbol recognition with support vector machines," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1326-1332, 2008.
[29] Adankon, M. M. and Cheriet, M., "Model selection for the LS-SVM: application to handwriting recognition," *Pattern Recognition*, vol. 42, no. 12, pp. 3264-3270, 2009.
[30] Taha, H. A., *Operation Research: an introduction*, 7th ed. Pearson Education, Inc., New Delhi, India, 2006, p 765.
[31] Suykens, J. A. K., Gestel, T. V., Brabanter, J. D., Moor, B. D. and Vandewalle, J., *Least Squares Support Vector Machines*, World Scientific, Co., Singapore, 2002, p 71.
[32] Kohavi, R. and Provost, F., "Glossary of Terms,", *Journal Machine Learning-Special Issue on Applications of Machine Learning and the Knowledge Discovery Process*, vol. 30, no. 2-3, pp. 271-274, 1998.
[33] Fawcett, T., "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006.
[34] Karnan, M. and Thangavel, K., "Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications," *Computer Methods and Programs in Biomedicine*, vol. 87, no. 1, pp. 12-20, 2007.
[35] Hosmer, D. W. and Lemeshow, S., *Applied Logistic Regression*, 2nd ed. John Wiley & Sons, Inc., NJ, USA, 2000, p 164.
[36] http://marathon.csee.usf.edu/Mammography/Database.html.
[37] Kennedy, J. and Eberhart, R. 1995. Particle swarm optimization. Proc. IEEE International Conference on Neural Networks, Perth, Australia, 27 Nov.- 01 Dec. 1995, vol. 4, pp. 1942-1948.

Baljit Singh Khehra was born in 1974 at Ferozepur, Punjab, India. He received the Bachelor of Engineering degree in Computer Engineering from Punjab Technical University, Jalandhar, Punjab, India in 1998 and the Master of Technology in Computer Science and Engineering from Punjabi university, Patiala, Punjab, India in 2005. Currently, he is pursuing the Ph.D. degree in the area of medical image processing from Sant Longowal Institute of Engineering and Technology, Longowal, Sangrur, Punjab, India.

He is an associate professor in the Department of Computer Science & Engineering of Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, Punjab, India. He has published more than 50 research papers in various international journals and refereed conferences. His main research interests are in image processing, medical imaging, artificial neural networks and genetic algorithms.

He is a member of the Institution of Engineers (India), Computer Society of India (CSI), Punjab Academy of Sciences, International Association of Computer Science and Information Technology and the International Association of Engineers (IAENG).

Prof. (Dr.) Amar Partap Singh Pharwaha was born in 1967 at Sangrur, Punjab, India. He received the Bachelor of Technology degree in Electronics Engineering from Guru Nanak Dev University, Amritsar, Punjab, India in 1990 and the Master of Technology degree in Instrumentation from Regional Engineering College, Kurukshetra, Haryana, India in 1994. He also got the Ph. D. degree in Electronics and Communication Engineering from Punjab technical university, Jalandhar, Punjab, India in 2005.

He is working as Professor in the Department of Electronics and Communications Engineering at Sant Longowal Institute of Engineering and Technology, Longowal, Sangrur, Punjab, India. He has published more than 100 research papers at various international and national level symposia/ conferences and journals. His research interests in virtual instrumentation, artificial neural networks and medical electronics.

He is a fellow of the Institution of Engineers (India), and life member of Instrument Society of India, Indian Society of Technical Education (India), Punjab Academy of Sciences and the International Association of Engineers (IAENG).