

MCs Detection with Combined Image Features and Twin Support Vector Machines

Xinsheng Zhang

VIPSL, School of Electronic Engineering, Xidian University, Xi'an China
School of Management, Xi'an University of Arch. & Tech., Xi'an, China
Email: xinsheng.zh@hotmail.com

Xinbo Gao and Ying Wang

VIPSL, School of Electronic Engineering, Xidian University, Xi'an China
Email: {xbgao, wyang}@lab202.xidian.edu.cn

Abstract—Breast cancer is a common form of cancer diagnosed in women. Clustered microcalcifications(MCs) in mammograms is one of the important early sign. Their accurate detection is a key problem in computer aided detection (CDAe). In this paper, a novel approach based on the recently developed machine learning technique - twin support vector machines (TWSVM) to detect MCs in mammograms. The ground truth of MCs in mammograms is assumed to be known as a priori. First each MCs is preprocessed by using a simple artifact removal filter and a high-pass filter. Then the combined image feature extractors are employed to extract 164 image features. In the combined feature domain, the MCs detection procedure is formulated as a supervised learning and classification problem, and the trained TWSVM is used as a classifier to make decision for the presence of MCs or not. A large number of experiments were carried out to evaluate and compare the performance of the proposed MCs detection algorithm. Experimental results show that the proposed TWSVM classifier is more advantageous for real-time processing of MCs in mammograms.

Index Terms—mammogram, support vector machine, twin support vector machine, microcalcification, ROC curves

I. INTRODUCTION

Breast cancer continues to be a significant public health problem in the world. Mammography is, at present, one of the most effective methods for early detection of breast cancer. One of the important early signs of breast cancer is the appearance of microcalcification clusters (MCs) in mammogram, which appear in 30%-50% of mammographically diagnosed cases.

Because of the importance in early breast cancer diagnosis, accurate detection of MCs has become an important problem. Up till now, a number of different approaches have been applied to the detection of MCs. Concerning image segmentation and specification of regions of interest(ROI), several methods have been reviewed [1]. In the past decade, the most work reported in the literature employed neural networks for MCs

characterization [2-6]. With the rise and development of SVM, various SVMs are employed to classify the ROIs [7]. SVM emerged from research in statistical learning theory on how to regulate the trade-off between statistical structural complexity and empirical risk. One of the most popular SVM classifiers is the maximum margin one that attempts to reduce generalization error by maximizing the margin between two disjoint half planes. The resulting optimization task involves the minimization of a convex quadratic objective function subject to linear inequality constraints. While the SVM approach achieves a good performance, the computational complexity of the SVM classifier may prove to be burdensome in real-time or near real-time applications because the solution of the convex optimization problem. Recently Jayadeva et al. [8] proposed a nonparallel plane classifiers for binary data classification, termed as twin support vector machine(TWSVM). It aims at generating two nonparallel planes such that each plane is closer to one of the two classes and is as far as possible from the other. TWSVM solves a pair of quadratic programming problems (QPPs), whereas, in SVM, a single QPP. The two QPPs have a smaller size than the one larger QPP, which makes TWSVM work faster than the standard SVM. It can reduce the optimize process of the solution.

In this work, to improve the performance of the available MCs detection algorithms, we propose to improve the computational efficiency of the pervious approaches by using an alternative, recently developed machine learning technique -- twin support vector machine (TWSVM) for MCs detection to distinguish the MCs from the other ROIs. The advantage of this approach is that the TWSVM classifier can yield a decision function that is much faster than the SVM while maintaining its detection accuracy. The classifier is first trained by the real mammograms from DDSM. Then it is used to detect other ROIs.

In our experiments, a database of 112 case clinical mammograms from DDSM is selected as the test bed, which containing 2231 MCs. These mammograms were divided into two subsets, the first one is used for training

and validation and another for test. Compared to several other existing methods, the proposed approach yields superior performance when evaluated using receiver operating characteristic (ROC) curves. It could work 4 times faster than SVM in training stage and achieved the best average sensitivity as high as 93.85% with respect to 5.31% false positive rate and Az=0.9677 in each testing phase.

The rest of the paper is organized as follows: Feature extraction methods are introduced in Section I. A brief review of the theory of TWSVM learning for classification is provided in Section II. The proposed MCs detection algorithm based on TWSVM is formulated in Section III. An evaluation study of the proposed approach is described in Section IV, and the experimental results are given in Section V. Finally, conclusions are drawn in Section VI.

II. COMBINED FEATURES EXTRACTION

The research methodology uses a novel ensemble algorithm with relevance feedback proposed in this paper, for combining a set of features from a number of existing features and detection MCs. As we know, feature extraction is a very important part for the classification. To get the best feature or combination of features and get the high classification rate for microcalcification clusters classification is one of main aims of the proposed research. In our task, a set of 164 features, shown in Table I, is calculated for each suspicious area (ROI) from the textural, spatial and transform domains in our research.

TABLE I COMBINED DIFFERENT KIND OF IMAGE FEATURES USED IN OUR EXPERIMENTS.

Feature groups	Type	#No.
Histogram based texture features	Mean	1
	Standard deviation	2
	Smoothness	3
	Third moment	4
	Uniformity	5
	Entropy	6
Multi-scale histogram features	Histograms with different number of bins 3,5,7,9	7~30(total 24)
Zernike Moments Features[9]	Zernike Moments	31~66 (total 36, D=10)
Combined first 4 moments features	Combined moments feature	67~114 (total 48)
Tamura texture signatures	Tamura features	115~120 (total 6)
Chebyshev transform histogram feature [10]	Chebyshev histogram	121~152 (total 32)
Radon Transform Features	Radon features	153~164 (total 12)

Before training the classifier, we use the feature extractor discussed in Table I to extract MCs features in the feature domain. The 164 features of each block are extracted as follows:

(1) Intensity histogram based texture feature

A frequently used approach for texture analysis is based on statistical properties of the intensity histogram. One class of such measures is based on statistical

moments. The expression for the *n*th moment about the mean is given by

$$\mu_n = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i) \tag{1}$$

where z_i is a random variable indicating intensity, $p(z)$ is the histogram of the intensity levels in a region, L is the number of possible intensity levels, and

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \tag{2}$$

is the mean(average)intensity. Table 2 lists the descriptors used in our experiments based on statistical moments and also on uniformity and entropy.

TABLE II DESCRIPTORS OF THE TEXTURE BASED ON THE INTENSITY HISTOGRAM OF A REGION.

Moment	Formula	#No.
Mean	$m = \sum_{i=0}^{L-1} z_i p(z_i)$	1
Standard deviation	$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2}$	2
Smoothness	$R = 1 - 1/(1 + \sigma^2)$	3
Third moment	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$	4
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$	5
Entropy	$e = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$	6

(2) Multiscale intensity histograms

We compute signatures based on "multi-scale histograms" idea. Idea of multi-scale histogram comes from the belief of a unique representation of an image through infinite series of histograms with sequentially increasing number of bins. Here we used 4 histograms with number of bins being 3,5,7,9 to calculate the image feature, so we get a 1*24 row vectors.

(3) Zernike moments

We use derived moments based on alternative complex polynomial functions, know as Zernike polynomials[9]. They form a complete orthogonal set over the interior of the unit circle $x^2 + y^2 = 1$ and are defined as

$$Z_{pq} = (p+1)/\pi \int_{x^2+y^2 \leq 1} f(x,y) V(\rho,\theta) dx dy, \tag{3}$$

$$V_{pq}(x,y) = V_{pq}(\rho,\theta) = R_{pq}(\rho) \exp(jq\theta), \tag{4}$$

$$R_{pq}(\rho) = \sum_{s=0}^{(p-|q|)/2} \frac{(-1)^s [(p-s)!] \rho^{p-2s}}{s! \left(\frac{p+|q|}{2} - s\right)! \left(\frac{p-|q|}{2} - s\right)!}, \tag{5}$$

where p is a nonnegative integer, q is an integer subject to the constraint $p - |q| = \text{even}$ and $|q| \leq p$, $\rho = \sqrt{x^2 + y^2}$ is the radius from (x,y) to the image centroid, $\theta = \tan^{-1}(y/x)$ is the angle between ρ and x-axis. The Zernike moment Z_{pq} is order p with repetition q . For a digital image, the respective Zernike moments are computed as

$$Z_{pq} = (p+1)/\pi \sum_i f(x_i, y_i) V(\rho, \theta) dx dy, x^2 + y^2 \leq 1, \tag{6}$$

where i runs over all the image pixels. Zernike moments are used as the feature extractor where by the order is varied to achieve the optimal classification performance.

(4) Combined first four moment features

Signatures on the basis of first four moments (also known as mean, std, skewness, kurtosis) for data generated by vertical, horizontal, diagonal and alternative diagonal 'combs'. Each column of the comb results in 4 scalars [mean, std, skewness, kurtosis], we have as many of those [...] as 20. So, 20 go to a 3-bin histogram, producing 1x48 vectors.

(5) Tamura texture signatures

Tamura et al. took the approach of devising texture features that correspond to human visual perception [10]. Six textural features: coarseness, contrast, directionality, line-likeness, regularity and roughness, are defined for the image feature for object recognition.

(6) Transform domains feature

We computes signatures (32 bins histogram) from coefficients of 2D Chebyshev transform (10th order is used in our experiments). Also we used signatures based on the Radon transform as a kind of image features. Radon transform is the projection of the image intensity along a radial line (at a specified orientation angle), total 4 orientations are taken. Transformation $n/2$ vectors (for each rotation), go through 3-bin histogram and convolve into 1x12 vectors.

III. TWIN SUPPORT VECTOR MACHINES

In this paper, MCs detection is formulated as a binary (+1, -1) classification problem. At each location in a mammogram, the trained TWSVM classifier is applied to determine where a MCs object is present or not. We defined $\mathbf{x} \in R^n$ as a pattern to be classified, and y as its class label (i.e., $y \in \{\pm 1\}$). In addition, let $\{(\mathbf{x}_i, y_i), i = 1, 2, \dots, m\}$ denote a given set of m training examples. And meanwhile we define m row vectors $A_i (i = 1, 2, \dots, m)$ in the n dimensional real space R^n , where $\mathbf{A}_i = (A_{i1}, A_{i2}, \dots, A_{in})$ and then $A_i = \mathbf{x}_i^T$. Here data points belonging to classes 1 and -1 are represented by matrices \mathbf{A} and \mathbf{B} . Let the number of patterns in class 1 and -1 be given by m_1 and m_2 . Therefore, the sizes of matrices \mathbf{A} and \mathbf{B} are $(m_1 \times n)$ and $(m_2 \times n)$, respectively.

A. Linear TWSVM Classifier

Twin support vector machine (TWSVM) [8] obtains nonparallel planes around which the data points of the corresponding class get clustered. Each of the two quadratic programming problems in the TWSVM pair has the formulation of a typical SVM, except that not all patterns appear with constrains of either problem at the same time.

The TWSVM classifier is obtained by solving the following pair of quadratic programming problems

$$\begin{aligned} \text{(TWSVM1)} \quad & \min_{\mathbf{w}^{(1)}, \mathbf{b}^{(1)}, \mathbf{q}} \frac{1}{2} (\mathbf{A}\mathbf{w}^{(1)} + \mathbf{e}_1 \mathbf{b}^{(1)})^T (\mathbf{A}\mathbf{w}^{(1)} + \mathbf{e}_1 \mathbf{b}^{(1)}) + c_1 \mathbf{e}_1^T \mathbf{q} \quad (7) \\ \text{s.t.} \quad & -(\mathbf{B}\mathbf{w}^{(1)} + \mathbf{e}_2 \mathbf{b}^{(1)}) + \mathbf{q} \geq \mathbf{e}_2, \quad \mathbf{q} \geq 0, \quad \text{and} \end{aligned}$$

$$\begin{aligned} \text{(TWSVM2)} \quad & \min_{\mathbf{w}^{(2)}, \mathbf{b}^{(2)}, \mathbf{q}} \frac{1}{2} (\mathbf{B}\mathbf{w}^{(2)} + \mathbf{e}_2 \mathbf{b}^{(2)})^T (\mathbf{B}\mathbf{w}^{(2)} + \mathbf{e}_2 \mathbf{b}^{(2)}) + c_2 \mathbf{e}_2^T \mathbf{q} \quad (8) \\ \text{s.t.} \quad & (\mathbf{A}\mathbf{w}^{(2)} + \mathbf{e}_1 \mathbf{b}^{(2)}) + \mathbf{q} \geq \mathbf{e}_1, \quad \mathbf{q} \geq 0, \end{aligned}$$

where $c_1, c_2 > 0$ are parameters and \mathbf{e}_1 and \mathbf{e}_2 are vectors of ones of appropriate dimensions.

This algorithm aims to find two optimal hyperplanes, one for each class, and classifiers points according to which hyperplane a given point is closest to. The first term in the objective function of (7) or (8) is the sum of squared distances from the hyperplane to points of one class. Therefore, minimizing it tends to keep the hyperplane close to points of one class (say class 1). The constraints require the hyperplane to be at a distance of at least 1 from points of the other class (say class -1); a set of error variables is used to measure the error wherever the hyperplane is closer than this minimum distance of 1. The second term of the objective function minimizes the sum of error variables, thus attempting to minimize misclassification due to points belonging to class -1.

B. Kernel TWSVM Classifier

The linear TWSVM can be easily extended to a nonlinear classifier by first using a nonlinear operator $\Phi(\cdot)$ to map the input pattern \mathbf{x} into a higher dimensional space \mathbf{H} , where the function $K(\cdot, \cdot)$ is defined as

$$K(\mathbf{x}, \mathbf{y}) \equiv \Phi^T(\mathbf{x})\Phi(\mathbf{y}). \quad (9)$$

And the kernel-generated surfaces instead of planes are defined as follows:

$$\begin{aligned} K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}^{(1)} + b^{(1)} = 0, \quad \text{and} \quad K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}^{(2)} + b^{(2)} = 0, \end{aligned} \quad (10)$$

where $\mathbf{C}^T = [\mathbf{A} \ \mathbf{B}]^T$, and K is an appropriately chosen kernel, such as polynomial kernel, RBF kernel etc.. Compared with the above arguments, the optimization problem of Kernel TWSVM (KTWSVM1, KTWSVM2) is defined as:

$$\begin{aligned} \text{(KTWSVM1)} \quad & \min_{\mathbf{u}^{(1)}, \mathbf{b}^{(1)}, \mathbf{q}} \frac{1}{2} \|K(\mathbf{A}, \mathbf{C}^T)\mathbf{u}^{(1)} + \mathbf{e}_1 \mathbf{b}^{(1)}\|^2 + c_1 \mathbf{e}_1^T \mathbf{q} \quad (11) \\ \text{s.t.} \quad & -K(\mathbf{B}, \mathbf{C}^T)\mathbf{u}^{(1)} + \mathbf{e}_2 \mathbf{b}^{(1)} + \mathbf{q} \geq \mathbf{e}_2, \quad \mathbf{q} \geq 0, \end{aligned}$$

$$\begin{aligned} \text{(KTWSVM2)} \quad & \min_{\mathbf{u}^{(2)}, \mathbf{b}^{(2)}, \mathbf{q}} \frac{1}{2} \|K(\mathbf{B}, \mathbf{C}^T)\mathbf{u}^{(2)} + \mathbf{e}_2 \mathbf{b}^{(2)}\|^2 + c_2 \mathbf{e}_2^T \mathbf{q} \\ \text{s.t.} \quad & -K(\mathbf{A}, \mathbf{C}^T)\mathbf{u}^{(2)} + \mathbf{e}_1 \mathbf{b}^{(2)} + \mathbf{q} \geq \mathbf{e}_1, \quad \mathbf{q} \geq 0, \end{aligned} \quad (12)$$

where $c_1 > 0$, and $c_2 > 0$ are tuned parameters.

In the paper, we use the popular polynomial kernels and the RBF kernels, which are known to satisfy the Mercer's condition. They are defined as follows,

1) Polynomial kernel:

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^p, \quad (13)$$

where $p > 0$ is a constant that defines as the kernel order.

2) RBF kernel:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right), \quad (14)$$

where $\sigma > 0$ is a constant that defined as the kernel width.

IV. MCs DETECTION WITH TWSVM

For a given digital mammography image, we consider the MCs detection process as the following steps:

- Step 1. Preprocess the mammography image by removing the artifacts, suppressing the inhomogeneity of the background and enhancing the microcalcifications.
- Step 2. At each pixel location in the image, extract a $A_{m \times m}$ small window to describe its surrounding image feature.
- Step 3. Apply feature extraction methods to get the combined feature vector \mathbf{x} .
- Step 4. Use the trained TWSVM classifier to make decision whether \mathbf{x} belongs to MCs class (+1) or not (-1).

A. Mammogram Preprocessing

First, a simple film-artifact removal filter is applied to remove the film-artifacts from mammography image. There are small emulsion continuity faults on the X-ray mammogram films, which look like microcalcifications. These artifacts are usually sharply defined and brighter than the microcalcifications, and the size of the artifacts is within 5×5 pixels in our experiments.

Let us consider two windows centered on a current pixel (x, y) . R_1 is the inner region and R_2 is the surrounding region. If the difference between the pixel value at the current position (x, y) , $I(x, y)$, and the mean value of the surrounding region $A(x, y)$ is larger than a threshold value, T , then the pixel value of current position is substituted by the mean value of the surrounding region R_2 . The output of the filmartifacts removal filtering, $I_1(x, y)$, is defined as

$$I_1(x, y) = \begin{cases} A(x, y), & \text{if } [I(x, y) - A(x, y)] > T \\ I(x, y), & \text{otherwise.} \end{cases} \quad (15)$$

In (15), the threshold value, T , is empirically selected from many digitized mammograms so that the film-artifacts can be removed and the microcalcifications can be preserved.

After the artifacts were removed, our task is to suppress the mammogram background. A high-pass filter is designed to preprocess each mammogram before extracting samples. With the Gaussian filter, $f(x, y) = \exp(-(x^2 + y^2)/(2\sigma^2))$, we use a $m \times m$ window size Gaussian filter where $m = 4\sigma^2 + 1$, experimentally in the study $\sigma = 2$. The output of high-pass filter is denoted by $I_2(x, y) = I_1(x, y) - f(x, y) * I_1(x, y)$, where $*$ is linear convolution.

B. Input Patterns for MCs detection

After the mammogram preprocessing stage, we extract combined image features from $A_{m \times m}$ as a feature vector

\mathbf{x} . $A_{m \times m}$ is a small window of $m \times m$ pixels centered at a location that we concerned in a mammogram image. The choice of m should be large enough to include the MCs (in our experiment, we take $m=115$). The task of the TWSVM classifier is to decide whether the input window $A_{m \times m}$ at each location is a MCs pattern ($y = +1$) or not ($y = -1$).

C. Training Data Set Preparation

The procedure for extracting training data from the training mammograms is given as follows. For each MCs location in a training mammogram set, a window of $m \times m$ image pixels centered at its center of mass is extracted; the area is denoted by $A_{m \times m}^i$, with respect to \mathbf{x}_i after subspace feature extraction, and then \mathbf{x}_i is treated as an input pattern for the positive sample ($y_i = +1$). The negative samples are collected ($y_i = -1$) similarly, except that their locations are randomly selected from the non MCs locations in the training mammograms. In the procedure, no window in the training set is allowed to overlap with any other training window.

D. TWSVM training and Parameter Selection

After the positive and negative training examples are gathered, some parameters should be initialized first, such as the type of kernel function, its associated parameter, and the regularization parameter c_1 and c_2 in the structural risk function. To optimize these parameters, we applied m -fold cross validation method to get the optimal parameters.

Once the best parametric setting (i.e., the type of the kernel function and its associated parameter) is determined, the TWSVM classifier is retrained using all the available samples in the training set to obtain the final form of the decision function. The resulting classifier will be used to detect the MCs in mammogram.

V DATABASE AND PERFORMANCE EVALUATION

A. Mammographic Database

In this research Digital Database for Screening Mammography (DDSM) from University of South Florida is used for experiments. The optical density range of the scanner for the database was 0-3.6. The 12bits digitizer was calibrated so that the gray values were linearly and inversely proportional to the optical density. In DDSM database, the boundaries for the suspicious regions are derived from markings made on the film by at least two experienced radiologists.

B. Performance Evaluation Methods

The use of the overall classification accuracy as an evaluation metric is adequate provided the class distribution among examples is constant and relatively balanced. Furthermore, this evaluation approach also assumes equal error costs, i.e. that a false positive error is equally significant as a false negative error. Since the used data set in our experiments is unbalanced, only

correction of classification accuracy may be misleading. Unfortunately, in real-life problems, these assumptions are not always true. Consequently, the performance of such systems is best described in terms of their sensitivity and specificity quantifying their performance related to false positive and false negative instances. These metrics are based on the consideration that a test point always falls into one of the following four categories:

- False Positive (FP): if the system labels a negative point as positive;
- False Negative (FN): if the system labels a positive point as negative;
- True Positive (TP): if the system correctly predicts the label;
- True Negative (TN): if the system correctly predicts the label.

The sensitivity or true positive of a learning machine, which is also called recall, is defined as the ration between the number of true positive predictions TP and the number of positive instances in the test set. It is defined as follows:

$$Sensitivity = TP / (TP + FN). \tag{10}$$

While the specificity or true negative rate is defined as the ratio between the number of true negative predictions TN and the number of negative instance in the test set. It is defined as

$$Specificity = TN / (TN + FP) \tag{11}$$

The overall accuracy is the ratio between the total number of correctly classified instances and the test set size. It is defined as

$$OverallAccuracy = N_r / N \tag{12}$$

where $N_r = TP + TN$ is the number of correctly classified samples during the test run and $N = TP + TN + FP + FN$ is the complete number of test samples.

Another important measure is the area under curve (Az) wherein the curve is the receiver operating characteristic (ROC) curve [11]. ROC curve is a plotting of true positive as a function of false positive. Higher ROC, approaching the perfection at the upper left hand corner, would indicate greater discrimination capacity.

VI EXPERIMENTAL RESULTS

A. Combined Feature Extraction

The data in the training (and validation), test, and validation sets were randomly selected from the training set. Each selected sample was covered by a 115x115 window whose center coincided with the center of mass of the suspected MCs. The blocks included 2231 with true MCs and 8364 with normal tissue.

Before training the classifier, we first use the feature extraction algorithm to extract MCs feature in combined feature domain. The 164-dimension feature vector will be calculated for each image block. When we get the image feature vector, feature normalization program should be user for normalizing the features to be real numbers in the range of [0, 1]. The normalization is accomplished by the following step: (a) change all the features to be positive

by adding the magnitude of the largest minus value of this feature times 1.01^2 , (b) divided by the maximum value of the same feature. The normalized features are used as the inputs of the proposed ensemble learning algorithm for training and classification.

B. TWSVM Training and Model Selection

For model selection, the TWSVM classifier is first trained by using the 10-fold cross-validation procedure with different model and parametric settings. In our training stage, we used generalization error, which was defined as the total number of incorrectly classified examples divided by the total number of samples classified, as a metric to measure the trained classifier. Generalization error was computed using only the samples used during training. For the sake of convenient, the parametric values of c_1 and c_2 in our experiment are set to be equal (i.e. $c_1 = c_2$). In Fig. 1(a), we summarize the results for trained TWSVM classifier with a polynomial kernel. The estimated generalization error is plotted versus the regularization parameters c_1 and c_2 for kernel order $p = 2$, $p = 3$ and $p = 4$. Similarly, Fig. 4(b) summarizes the results with RBF kernel. Here, the generalization error is plotted for different values of the width σ (2, 7, 10, 15, 17, 20).

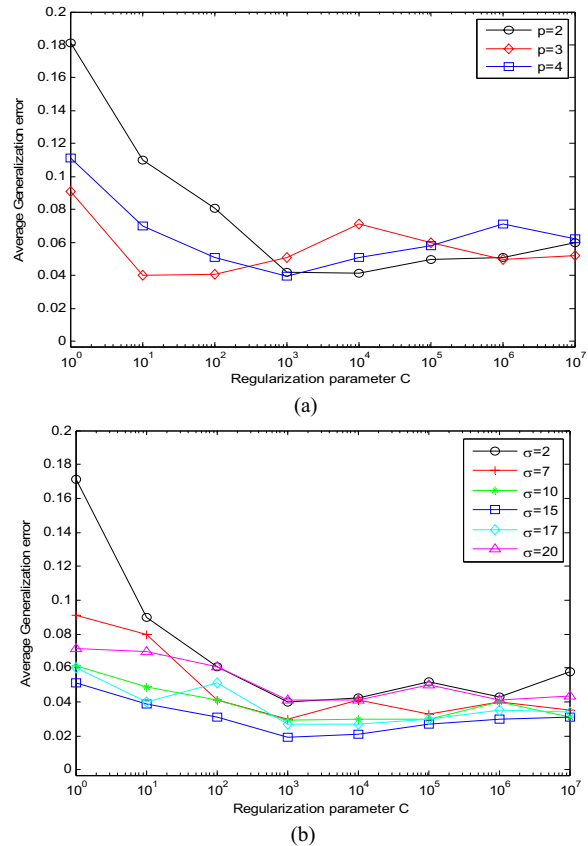


Fig. 1 Plot of average generalization error rate versus regularization parameter C ($c_1 = c_2$) achieved by training TWSVM using (a) a polynomial kernel with orders 2, 3, and 4, and (b) a Gaussian RBF kernel width $\sigma = 2, 7, 10, 15, 17, 20$ by 10-fold validation method.

For the polynomial kernel, it is found that the smallest generalization error is get when $p=3$ and c_1 and c_2 are between 100 and 1000. A similar error level was also achieved by the RBF kernel over a wide range of parameter settings (e.g., $\sigma = 15$ and $c_1, c_2 \in [100, 1000]$). These results indicate that the performance of the TWSVM classifier is not very sensitive to the model parameters. Indeed, essentially similar performance was achieved when σ was varied from 10 to 20.

Having determined that the TWSVM results do not vary significantly over a wide range of parameter settings, next we will focus for the remainder of the paper on a particular, representative configuration of the TWSVM classifier, having a RBF kernel with $\sigma = 15$ and $c_1 = c_2 = 1000$.

For comparison, we also trained the SVM classifier using the same training procedure as for the TWSVM. The best error level (3.1%) was achieved when an RBF kernel with $\sigma = 15$ and the regularization parameter c was set to 10. Similar to TWSVM, the SVM classifier was retrained using these tuned parameters with all the samples in the training set.

For the TWSVM classifier, the number of SVs (on the two nonparallel planes) was 83 (500 positive and 500 negative samples); for SVM the number of SVs (on the two parallel planes) was 285 (500 positive and 500 negative samples). We can see that the TWSVM classifier is much effective than the traditional SVM method.

C. Performance Evaluation Results

The trained classifiers are evaluated using all the mammograms in the test subset. The test results are summarized by using ROC curves in Fig.2 for TWSVM. For comparison, ROC curve is also show for the traditional SVM classifier with the same inputs. Average experimental results of TWSVM and SVM classifier are shown in Table I.

TABLE III. EXPERIMENTAL RESULTS OF TWSVM AND SVM CLASSIFIER.

Method	Sensitivity	Specificity	Az
SVM	0.9201	0.9237	0.9586
TWSVM	0.9385	0.9469	0.9706

From the Fig. 2, it can be shown that, the TWSVM classifier has a higher detection accuracy rate compared to SVM classifier. By using the same extracted feature, compared with SVM, TWSVM has a better detection rate and run about 4 times faster than the SVM when we train the classifier. In particular, the TWSVM classifier achieved a sensitivity of approximately 93.85% with respect to 5.31% false positive rate and Az=0.9706. With the same training data set and test data set, the SVM classifier achieved a sensitivity of 92.01% , 7.63% false positive rate and Az=0.9586.

D. Experimental Execution Times Analysis

Finally, we show in Table IV the computation times for the different methods (implemented with MATLAB on

an Intel Core 2 CPU 1.86 GHz, DDR2 1.99GB PC). The training time is the average amount of time taken during each round of training in the 10-fold cross validation procedure during the training phase; the test time is the average amount of time taken by the trained classifier for classifying each image during the testing phase. Compared to SVM, the TWSVM classifier has reduced the detection time to about 1/4. Note that such a dramatic saving in computation time is achieved without sacrificing the best detection. This is because TWSVM solves a pair of quadratic programming problems (QPPs), whereas, in SVM, a single QPP. The two QPPs have a smaller size than the one larger QPP, which makes TWSVM work faster than the standard SVM.

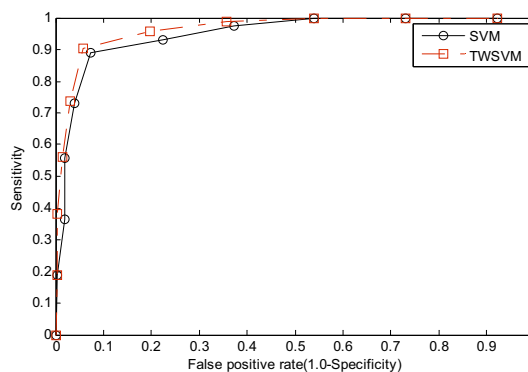


Fig. 2. Roc curves MCS detection by using SVM and TWSVM classifier, where Az (SVM)=0.9586 and Az (TWSVM)=0.9706.

TABLE IV.

THE COMPUTATION TIME FOR TWSVM AND SVM CLASSIFIERS. THE TRAINING TIME IS AVERAGE AMOUNT OF TIME TAKEN DURING EACH ROUND OF TRAINING IN THE 10-FOLD CROSS VALIDATION PROCEDURE, AND THE TESTING TIME IS THE AVERAGE AMOUNT OF TIME TAKEN BY THE TRAINED CLASSIFIER FOR CLASSIFYING EACH IMAGE DURING THE TESTING PHASE

	SVM	TWSVM
Training time(s)	4.9726	1.3127
Testing time(s)	0.3515	0.2801

VII. CONCLUSION

In this paper, we proposed a new algorithm based on TWSVM technique for detection microcalcification clusters in digital mammograms. In this method, combined image features are extracted from each image block of positive and negative samples, and TWSVM is trained through supervised learning with the 164 dimensional feature of each samples to test at every location in a mammogram whether the MCs is present or not. The decision function of the trained TWSVM classifier is determined in terms of twin support vectors that are identified from the examples during training stage. Compared to SVM classifier, TWSVM can achieve about 4 times faster while maintaining the detection accuracy. In our experiments, ROC curves indicated that the TWSVM learning approach yielded the best performance compared with the traditional SVM methods.

ACKNOWLEDGMENT

This work presented in our paper is supported by the National Natural Science Foundation of China (No. 60771068), the Natural Science Basic Research Plan in Shaanxi Province of China (No. 2007F248) and the Youth Scientific Research Foundation in Xi'an Univ. of Arc. & Tech. (No. QN0738). The authors are grateful for the anonymous reviewers who made constructive comments.

REFERENCES

- [1] H. D. Cheng, X. Cai, X. Chen *et al.*, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern Recognition*, vol. 36, no. 12, pp. 2967-2991, 2003.
- [2] S. Halkiotis, T. Botsis, and M. Rangoussi, "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks," *Signal Processing*, vol. 87, no. 7, pp. 1559-1568, 2007.
- [3] R. R. Hernandez-Cisneros, and H. Terashima-Marin, "Evolutionary Neural Networks Applied To The Classification Of Microcalcification Clusters In Digital Mammograms." pp. 2459-2466.
- [4] A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 141-150, 2005.
- [5] L. Bocchi, G. Coppini, J. Nori *et al.*, "Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks," *Medical Engineering & Physics*, vol. 26, no. 4, pp. 303-312, 2004.
- [6] Y. Songyang, and G. Ling, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 115-26, 2000.
- [7] I. El-Naqa, Y. Yongyi, M. N. Wernick *et al.*, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1552-1563, 2002.
- [8] Jayadeva, R. Khemchandani, and S. Chandra, "Twin Support Vector Machines for Pattern Classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 905-910, 2007.
- [9] L. Tsong-Wuu, and C. Yun-Feng, "A comparative study of Zernike moments." pp. 516-519.
- [10] M. S. Tamura H., Yamawaki T., "Texture features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Transactions. on*, vol. 1, no. 8, pp. 460-472, 1978
- [11] J. B. Tilbury, W. J. Van Eetvelt, J. M. Garibaldi *et al.*, "Receiver operating characteristic analysis for intelligent medical systems-a new approach for finding confidence intervals," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 952-963, 2000.

Xinsheng Zhang is presently pursuing his PhD degree in school of Electronic Engineering, Xidian University. He received his M.S. Degree and the B.S. degree in system engineering from Xi'an University of Architecture & Technology, China, in 2004 and 2001, respectively.

His research interests include image processing and machine learning.

Xinbo Gao received the B.S. degree in electronic engineering, and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, in 1994, 1997, and 1999, respectively. From 1997 to 1998, he was with the computer Games Research Institute, Shizuoka University, as a Research Fellow. From 2000 to 2001, he also worked at the multimedia Laboratory, Chinese University of Hong Kong, Hong Kong, as a Research Associate. Since 2003, he has been a full Professor in the School of Electronic Engineering at Xidian University, Xi'an, China, and the Director of Video/Image Processing System Laboratory (VIPSL).

His research interests include image/video processing and analysis, pattern recognition, machine learning and computational Intelligence.