

## Research Article

# A New Approach for Clustered MCs Classification with Sparse Features Learning and TWSVM

**Xin-Sheng Zhang**

*School of Management, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China*

Correspondence should be addressed to Xin-Sheng Zhang; [xinsheng.zh@gmail.com](mailto:xinsheng.zh@gmail.com)

Received 10 August 2013; Accepted 14 November 2013; Published 9 February 2014

Academic Editors: Y. Lu, J. Shu, and F. Yu

Copyright © 2014 Xin-Sheng Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In digital mammograms, an early sign of breast cancer is the existence of microcalcification clusters (MCs), which is very important to the early breast cancer detection. In this paper, a new approach is proposed to classify and detect MCs. We formulate this classification problem as sparse feature learning based classification on behalf of the test samples with a set of training samples, which are also known as a “vocabulary” of visual parts. A visual information-rich vocabulary of training samples is manually built up from a set of samples, which include MCs parts and no-MCs parts. With the prior ground truth of MCs in mammograms, the sparse feature learning is acquired by the  $l_p$ -regularized least square approach with the interior-point method. Then we designed the sparse feature learning based MCs classification algorithm using twin support vector machines (TWSVMs). To investigate its performance, the proposed method is applied to DDSM datasets and compared with support vector machines (SVMs) with the same dataset. Experiments have shown that performance of the proposed method is more efficient or better than the state-of-art methods.

## 1. Introduction

Breast cancer is the most common tumor disease in women, with the increasing incidences in recent years. And also, it is one of the major death causes among middle-aged women in the world. Currently, digital mammograms are one of the most reliable methods to perform the early diagnosis, which is very important for the effectiveness of treatment methods.

In digital mammograms, an important sign of the early breast cancer is the existence of MCs. They always exist in 30%–50% of mammographically diagnosed cases, which are present with tiny bright spots of different morphology. Microcalcifications are small calcifications with different shapes and densities, approximately 0.1–1 mm in diameter. Isolated microcalcifications are not dangerous, but a microcalcification cluster might be an early sign of breast cancer [1], which is a region including more than three microcalcifications per  $5 \text{ mm} \times 5 \text{ mm}$ .

However, there is only about 3% of useful information in mammograms, which can be seen by doctors with the naked eye. Due to the fact that most details in mammograms cannot be perceived by human eyes, it is even very difficult for a skillful radiologist to find the sign of early breast cancer, that

is, MCs, as a result missing the best time for treatment. So, one of the key techniques for early diagnosis of the breast cancer is to detect MCs and to judge whether they are malignant or not in mammograms.

According to recent researches, there are several existing criteria to characterize the MCs shape properties. Among them, one of the well known is the category criterion proposed by Le Gal et al. [2], which illustrates five groups (shown in Figure 1). Different groups identify different kinds of MCs in the ascending order of the degree of malignancy. As shown in Figure 1, the first group describes the O-shaped calcifications and partially calcified ones, known as teacup calcifications. The second one includes regular and round calcifications with uniform density. The third is composed of calcifications with the same shape and smaller size than the second one. Class IV is also called salt shaped, which is irregular MCs related to the high degree of malignancy. Type V is also closely related to a very high degree of malignancy, which is called vermicular shaped.

Up till now, computer aided diagnosis (CAD) is still a useful tool in breast cancer detection to improve the accuracy of radiologists and to help radiologists to read mammogram films. It may provide good help to radiologists in interpreting

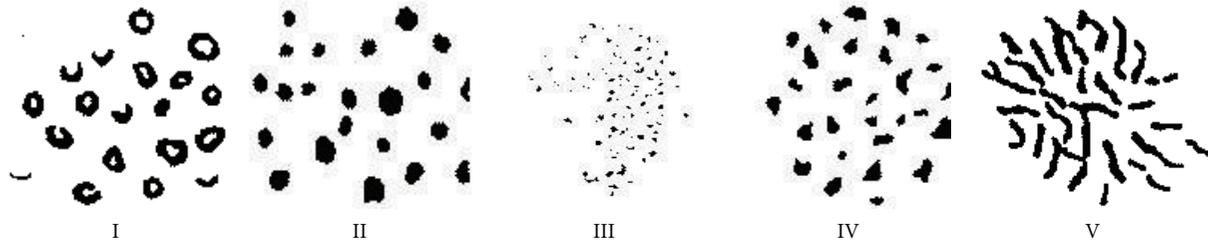


FIGURE 1: Le Gal's MCs classification standards. Type I: annular; Type II: regularly punctiform; Type III: dusty; Type IV: irregularly punctiform; Type V: vermicular calcification.

mammograms to detect MCs and classify them into malignancy or not. A large number of researchers in this field have been trying to find effective methods to automatically detect MCs and categorize them as normal, benign, or malignant.

Because it is very important in breast cancer diagnosis, the detection accuracy of MCs has become a crucial application task and research. Recently a lot of methods have been developed. These approaches have been also greatly assisting radiologists and doctors in diagnosing the disease [3–5]. Among them, several methods focus on image enhancement and segmentation of regions of interests (ROIs), such as local threshold and classical image filter [6, 7], optimal filters [7, 8], fractal models [9], wavelet and multiscale analysis [6, 10, 11], and mathematical morphology [12, 13]. Various classification approaches based on machine learning have also been presented to detect and classify MCs, such as rule-based systems [14], fuzzy logic systems [15–19], statistical methods based on Markov random fields [20, 21], support vector machines (SVMs) [20, 22], twin support vector machines (TWSVMs) [23, 24], and twin support tensor machines (TWSTMs) [24]. In the last ten years, a lot of researches reported in the literatures have used neural networks for MCs characterization [9, 10, 12, 25–28]. With the development of SVMs, various SVMs have been designed to categorize ROIs [29].

However, how to successfully apply the mammography technology to detect breast cancer and design a breast cancer detection system greatly depends on the careful designing of the two important modules: feature selection and sample classification. A lot of well-established methods have been proposed to address this challenge problem. According to [30], these methods can be categorized into the following groups: (a) traditional methods, such as linear discriminant analysis (LDA), K-nearest neighbor (KNN), logistic regression (LR), and generalized partial least square (GPLS); (b) classification trees and aggregation methods, such as classification and regression tree (CART), bagging and boosting (BB), random forest (RF), and ensemble learning (EL); (c) machine learning based methods, such as neural network (NN), support vector machines (SVMs), and twin support vector machines (TWSVMs); and (d) generalized methods, such as flexible discriminant analysis (FDA), bias discriminant analysis (BDA), mixture discriminant analysis (MDA), and shrunken centroid method.

To detect the early sign of this disease and to aid doctors to diagnose breast cancer in early stage, a novel approach for MCs classification is proposed based on sparse feature

learning and representation with TWSVMs, which is inspired by the recent progress in  $l_1$ -norm minimization-based approaches [31, 32]. These approaches, such as compressive sensing for sparse signal reconstruction, basis pursuit denoising, and the Lasso algorithm for features selection, have been well developed. Inspired from the above well-established approaches, our approach presented in this paper is based on the belief that the key problem to finding a solution to the problem depends on learning the suitable representation.

Especially, to extract high-level and conceptual information, for example, the existence of an MC in a mammogram block, it is very important for us to convert the low-level input, such as the pixel value, to high-level and more meaningful representations. Through this transformation, the feature learning and detection process will be well constructed.

Ideally, a test example can be represented just from the training samples of the same category. Therefore, when the test sample is expressed as a linear combination of the entire training sample, the coefficient vector will be sparse. That is, there will be relatively few nonzero coefficients in the vector. Test samples from the same category will have a similar sparse representation, while test samples from different categories will lead to different sparse representations. So the sparse representation coefficients can be treated as the more meaningful and discriminant information for the samples classification. In order to get the sparse coefficient vector, we use  $l_1$ -regularized least square [33] to solve the problem.

To achieve a good performance for MCs detection, we designed two methods to achieve the goal of the detection system. The first one is the sparse discriminant analysis algorithm, which is achieved by computing the residuals of sparse coefficients of the test sample between the centroid sparse coefficients of training samples. As we have known traditional supervised learning methods always use a training procedure to create a classification model for testing. But the proposed sparse representation based approach does not contain the separate training and testing sections. We directly achieved the classification goal out of the testing samples' sparse representation according to the training samples. Another unique feature of the new method is that no model selection is needed. The second one is designed by the combination of sparse representation and the state-of-the-art classifier TWSVMs. We employ the sparse representation approach as a feature learning method in terms of the coefficient vector for samples feature extraction, and then we feed it with the

trained TWSVMs to formulate the detection method as a supervised learning approach.

The paper is organized in five sections. Technology backgrounds of our approach are presented in Section 2; sparse representation based MCs detection algorithm and TWSVM based MCs detection with sparse representation are given in Section 3. Thereafter, the sparse representation based MCs detection algorithms are formulated, and experimental results are illustrated in Section 4 accordingly. Finally, conclusions are drawn in Section 5.

## 2. Technology Backgrounds

**2.1. Image Sparse Representation and Learning.** Given a training dataset  $\{(\mathbf{x}_i, l_i); i = 1, \dots, n\}$ ,  $\mathbf{x}_i \in R^d, l_i \in \{1, 2, \dots, N\}$ , where  $\mathbf{x}_i$  is the  $i$ th sample, a  $d$ -dimension column vector contains MCs features,  $d$  is the number of features, and  $l_i$  represents the label of the  $i$ th sample with  $N$  as the number of categories, and a test sample  $\mathbf{y} \in R^d$ , the problem of sparse representation aims to find a column vector  $\mathbf{c} = [c_1, c_2, \dots, c_n]^T$  such that

$$\mathbf{y} = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \dots + c_n \mathbf{x}_n, \quad (1)$$

and  $\|\mathbf{c}\|_0$  is minimized, where  $\|\mathbf{c}\|_0$  represents the  $l_0$ -norm, which means that it is equal to the number of non-zero components in the vector  $\mathbf{c}$ .

Suppose that we define a matrix by putting  $\mathbf{x}_i$  as the  $i$ th column of  $\mathbf{A} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ ; we can convert the problem of sparse representation into

$$\mathbf{c} = \min_{\mathbf{c}' \in R^n} \|\mathbf{c}'\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{c}. \quad (2)$$

How to get the close solution of the sparse representation problem is NP-hard, because it is a combinatorial optimization. If we replace the  $l_0$ -norm in (2) with  $l_p$ -norm, an approximation solution can be gotten. Thus,

$$\mathbf{c} = \min_{\mathbf{c}' \in R^n} \|\mathbf{c}'\|_p \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{c}, \quad (3)$$

where the  $l_p$ -norm of a vector  $\mathbf{u}$  is defined as  $\|\mathbf{u}\|_p = (\sum_i |u_i|^p)^{1/p}$ . A generalized version of (3), which allows for certain degree of noise, is defined to find a vector  $\mathbf{c}$ , when the following objective function is minimized:

$$J(\mathbf{c}, \lambda) = \min_{\mathbf{c}} \{ \|\mathbf{A}\mathbf{c} - \mathbf{y}\|_2 + \lambda \|\mathbf{c}\|_p \}, \quad (4)$$

where the scalar regularization  $\lambda$  is a positive parameter, which balances the trade-off between sparsity and reconstruction error.

Recently, development in the theory of compressed sensing and sparse representation reveals that, if the solution of (2) is sparse enough, the solution of the  $l_0$ -minimization problem is equal to the solution of the following  $l_1$ -minimization problem [33], which takes  $p = 1$  in (4):

$$\mathbf{c} = \min_{\mathbf{c}' \in R^n} \|\mathbf{c}'\|_1 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{c}, \quad (5)$$

$$J(\mathbf{c}, \lambda) = \min_{\mathbf{c}} \{ \|\mathbf{A}\mathbf{c} - \mathbf{y}\|_2 + \lambda \|\mathbf{c}\|_1 \}.$$

We can solve the problem in polynomial time by quadratic programming or standard linear programming methods. If the solution is very sparse, there will be more efficient methods to solve this problem.

**2.2. Twin Support Vector Machines.** Twin support vector machines (TWSVMs) are a new binary data classifier proposed by Jayadeva et al. [34], which aims at obtaining two nonparallel planes close to two nonparallel planes such that each plane is closer to one of the two classes and is as far as possible from the other. There are two quadratic programming problems (QPPs) to be solved in the TWSVMs. But each QPP is of a smaller size instead of a large one as we have in the traditional SVMs. So, to some extent, TWSVMs work much faster than the stand SVMs facing the same classification problem; that is, it is more efficient

According to (6) we can solve the following two QPPs to obtain the TWSVM classifier:

$$\begin{aligned} & \text{(TWSVM1)} \\ & \min_{\mathbf{w}^{(1)}, b^{(1)}, \mathbf{q}} \frac{1}{2} (\mathbf{A}\mathbf{w}^{(1)} + \mathbf{e}_1 b^{(1)})^T (\mathbf{A}\mathbf{w}^{(1)} + \mathbf{e}_1 b^{(1)}) + c_1 \mathbf{e}_1^T \mathbf{q} \\ & \text{s.t.} \quad -(\mathbf{B}\mathbf{w}^{(1)} + \mathbf{e}_2 b^{(1)}) + \mathbf{q} \geq \mathbf{e}_2, \quad \mathbf{q} \geq 0, \end{aligned} \quad (6)$$

$$\begin{aligned} & \text{(TWSVM2)} \\ & \min_{\mathbf{w}^{(2)}, b^{(2)}, \mathbf{q}} \frac{1}{2} (\mathbf{B}\mathbf{w}^{(2)} + \mathbf{e}_2 b^{(2)})^T (\mathbf{B}\mathbf{w}^{(2)} + \mathbf{e}_2 b^{(2)}) + c_2 \mathbf{e}_1^T \mathbf{q} \\ & \text{s.t.} \quad (\mathbf{A}\mathbf{w}^{(2)} + \mathbf{e}_1 b^{(2)}) + \mathbf{q} \geq \mathbf{e}_1, \quad \mathbf{q} \geq 0, \end{aligned} \quad (7)$$

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

The first term, in the objective function of (6) or (7), is defined by the sum of squared distances from the points of each class to the hyper plane. So, minimizing the objective function tends to keep the hyper plane close to points which belong to one class (say class +1). Meanwhile, the constraints request of the hyper plane to be a distance of no less than 1 to the points belonging to the other class (say class -1). In the model, we use a set of error variables to measure the computing error, whenever the distance to the hyper plane is less than 1. In (7), the objective function aims at minimizing the total error variables, so as to attempt to minimize the misclassification with respect to the points belonging to class -1.

TWSVMs are composed of two QPPs. And the objective function in each QPP, corresponding to a particular class and constrains, is determined by the patterns belonging to the other class. In TWSVM1, patterns of class +1 are clustered near the hyper plane  $\mathbf{x}^T \mathbf{w}^{(1)} + b^{(1)} = 0$ . Similarly, in TWSVM2, patterns of class -1 are clustered around the hyper plane  $\mathbf{x}^T \mathbf{w}^{(2)} + b^{(2)} = 0$ . The Lagrange equation corresponding to TWSVM1 is given by

$$\begin{aligned} & L(\mathbf{w}^{(1)}, b^{(1)}, \mathbf{q}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \\ & = \frac{1}{2} (\mathbf{A}\mathbf{w}^{(1)} + \mathbf{e}_1 b^{(1)})^T (\mathbf{A}\mathbf{w}^{(1)} + \mathbf{e}_1 b^{(1)}) + c_1 \mathbf{e}_1^T \mathbf{q} \\ & \quad - \boldsymbol{\alpha}^T (-(\mathbf{B}\mathbf{w}^{(1)} + \mathbf{e}_2 b^{(1)}) + \mathbf{q} - \mathbf{e}_2) - \boldsymbol{\beta}^T \mathbf{q}, \end{aligned} \quad (8)$$

where  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{m_2})^T$  and  $\beta = (\beta_1, \beta_2, \dots, \beta_{m_2})^T$  are the vectors of Lagrange multipliers. The Karush-Kuhn-Tucker (K.K.T.) conditions for TWSVM1 are given by

$$\begin{aligned} \mathbf{A}^T (\mathbf{A}\mathbf{w}^{(1)} + \mathbf{e}^1 b^{(1)}) + \mathbf{B}^T \alpha &= 0, \\ \mathbf{e}_1^T (\mathbf{A}\mathbf{w}^{(1)} + \mathbf{e}^1 b^{(1)}) + \mathbf{e}_2^T \alpha &= 0, \\ c_1 \mathbf{e}_2 - \alpha - \beta &= 0, \\ -(\mathbf{B}\mathbf{w}^{(1)} + \mathbf{e}_2 b^{(1)}) + \mathbf{q} &\geq \mathbf{e}_2, \quad \mathbf{q} \geq 0, \\ \alpha^T (-\mathbf{B}\mathbf{w}^{(1)} + \mathbf{e}_2 b^{(1)}) + \mathbf{q} - \mathbf{e}_2 &= 0, \quad \beta^T \mathbf{q} = 0, \\ \alpha &\geq 0, \quad \beta \geq 0. \end{aligned} \quad (9)$$

If we define  $\mathbf{H} = [\mathbf{A} \ \mathbf{e}_1]$ ,  $\mathbf{G} = [\mathbf{B} \ \mathbf{e}_2]$ , and  $\mathbf{u} = [\mathbf{w}^{(1)}; b^{(1)}]^T$ , we can get

$$\mathbf{u} = -(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{G}^T \alpha, \quad (10)$$

where  $\mathbf{H}^T \mathbf{H}$  is positive semidefinite. With the K.K.T. conditions and the Lagrange equation of problem TWSVM1, we can get the Wolfe dual of TWSVM1 as follows:

(DTWSVM1)

$$\begin{aligned} \max_{\alpha} \quad & \mathbf{e}_2^T \alpha - \frac{1}{2} \alpha^T \mathbf{G} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{G}^T \alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1. \end{aligned} \quad (11)$$

Similarly, if we consider TWSVM2; then we can also obtain its dual as

(DTWSVM2)

$$\begin{aligned} \max_{\gamma} \quad & \mathbf{e}_2^T \gamma - \frac{1}{2} \gamma^T \mathbf{H} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{H}^T \gamma \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1. \end{aligned} \quad (12)$$

And we define the augmented  $\mathbf{v} = [\mathbf{w}^{(2)}; b^{(2)}]^T$  as

$$\mathbf{v} = -(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{H}^T \gamma. \quad (13)$$

If we get the vector  $\mathbf{u}$  and  $\mathbf{v}$  from (10) and (13), the hyper planes

$$\mathbf{x}^T \mathbf{w}^{(1)} + b^{(1)} = 0, \quad \mathbf{x}^T \mathbf{w}^{(2)} + b^{(2)} = 0 \quad (14)$$

can be obtained. Suppose that we have a new sample  $\mathbf{x} \in \mathbf{R}^n$ , which is assigned to class  $p$  ( $p = 1, 2$ ). The data belongs to which category will be determined by the closer plane to the sample; that is,

$$\mathbf{x}^T \mathbf{w}^{(p)} + b^{(p)} = \min_{l=1,2} |\mathbf{x}^T \mathbf{w}^{(l)} + b^{(l)}|, \quad (15)$$

where  $|\cdot|$  represents a vertical distance of the point  $\mathbf{x}$  to the plane  $\mathbf{x}^T \mathbf{w}^{(l)} + b^{(l)} = 0$ ,  $l = 1, 2$ .

From the K.K.T. conditions, we can observe that patterns lie on the hyper plane given by  $\mathbf{x}^T \mathbf{w}^{(1)} + b^{(1)} = 0$  of class  $-1$

given  $0 < \alpha_i < c_1$  ( $i = 1, 2, \dots, m_2$ ), and such patterns of class  $-1$  are called support vectors of class  $-1$  as they play a key role when we determine the required hyper plane. A similar observation can be gotten from the problem TWSVM2. From the example shown in Figure 2, one can find the difference between TWSVMs and traditional SVMs.

### 3. Materials and Methods

**3.1. Database and Evaluation Metrics.** The digital database for screening mammography (DDSM) [35] was built by the University of South Florida, which is available for research at [36]. In our experiments, all images manually selected from this database are intensity images, digitized at  $43.5 \mu\text{m}/\text{pixel}$  and 12-bit gray scale. To evaluate our methods, we totally selected a set of 267 images from the DDSM, which are all clinical mammograms, to form an evaluation database.

To summarize quantitatively the performance of the proposed method, we used receiver operating characteristic (ROC) curves [37]. Receiver operating characteristic analysis is based on statistical decision theory, which is a commonly used criterion for classification performance measure. We can get a ROC curve by figuring the plotting of the classifier's sensitivity (also known as true positive classification rate) as a function of the classifier's specificity (also known as false positive rate). Sensitivity is a probability of correctly classifying a target object. Specificity is a probability of incorrectly classifying a non-target object. The area under the ROC curve ( $A_z$ ) is defined as an accepted way to compare the classifiers performance. Higher  $A_z$  would characterize the greater discrimination capacity. A good classifier should have a true positive rate of 1.0 (or 100%) and the false positive rate of 0.0 accordingly with respect to an  $A_z$  of 1.0.

**3.2. MCs Classification Based on Sparse Representation.** Ideally, the nonzero entries in the estimated  $\mathbf{c}$  will be associated with all the columns in  $\mathbf{A}$  from a single category. So we can easily designate the test image  $\mathbf{y}$  to that category. However, because of the existence of noise, the nonzero entries sometimes may be related to multiple categories. A lot of the state-of-art classifiers can resolve the problem. For example, we can simply designate  $\mathbf{y}$  to the category with the largest entry of  $\mathbf{c}$ . However, such heuristics cannot model the structure of the subspace associated with MCs blocks.

To better model the structure, we classify  $\mathbf{y}$  based on how much the coefficients relate to the training sample of each category reproducing  $\mathbf{y}$  alternatively. For each category  $i$  we define the corresponding characteristic function  $\delta_i : \mathbf{R}^n \rightarrow \mathbf{R}^n$  to select the coefficients associated with the  $i$ th category. If  $x \in \mathbf{R}^n$ ,  $\delta_i(x) \in \mathbf{R}^n$  is a new vector whose nonzero entries are the entries in  $x$  that belongs to the category  $i$  and whose entries associated with all the other subjects are zero or very close to zero. The classification algorithm can be summarized as follows.

*Algorithm 1* (MCs detection and classification based on sparse representation (MCs-SRC)). Input: We have a matrix of training images  $\mathbf{A} \in \mathbf{R}^{m \times n}$  from two categories (MCs or not), a linear transform  $\mathbf{D} \in \mathbf{R}^{d \times m}$ , and an error tolerance  $\epsilon$ .

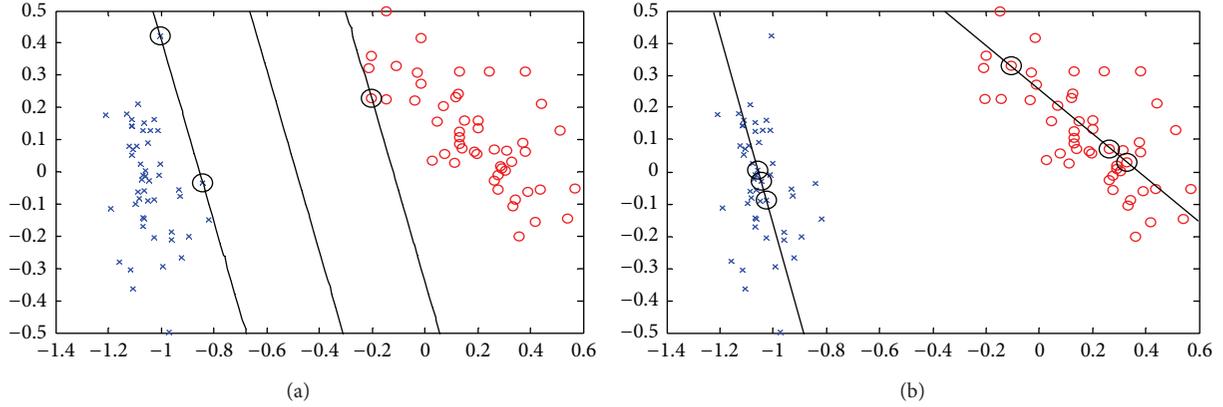


FIGURE 2: An example of SVMs and TWSVMs: (a) SVMs and (b) TWSVMs. All the points of class +1 are represented by a “x” and those of class -1 by “o.”

- (1) Compute features  $\tilde{\mathbf{y}} = \mathbf{D}\mathbf{y}$  and  $\tilde{\mathbf{A}} = \mathbf{D}\mathbf{A}$ , and normalize  $\tilde{\mathbf{y}}$  and columns of  $\tilde{\mathbf{A}}$  to unit length.
- (2) Solve the convex optimization problem:

$$\mathbf{c} = \min_{\mathbf{c}' \in \mathbb{R}^n} \|\mathbf{c}'\|_1 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{c},$$

$$J(\mathbf{c}, \lambda) = \min_{\mathbf{c}} \{ \|\mathbf{A}\mathbf{c} - \mathbf{y}\|_2 + \lambda \|\mathbf{c}\|_1 \}.$$
(16)

- (3) Compute the residuals  $r_i(\mathbf{y}) = \|\tilde{\mathbf{y}} - \tilde{\mathbf{A}}\mathbf{c}\|_2$  for  $i = +1, -1$ .

Output:  $f(\mathbf{y}) = \arg \min_i r_i(\mathbf{y})$ .

**3.3. MCs Classification Approach with Twin Support Machines and Sparse Representation (TWSVMs-SR).** Given a set of mammogram blocks, each block is transformed and represented in terms of the sparse coefficients with respect to the parts from the vocabulary constructed in the image sparse representation and learning stage. Each block is then transformed into a vector and represented as a sparse feature vector based on the vocabulary parts (the negative and positive samples in the sparse learning set). Then we can use the learned sparse coefficients as each block’s sparse feature. So we consider the sparse representation approach as a feature extraction method in terms of the coefficient vector, and then we feed it with the state-of-the-art classifier (TWSVMs) to formulate the detection method as a supervised learning approach.

If we have a set of training blocks labeled as positive (MCs) or negative (non-MCs), each image is represented as a feature vector using the method described above. These feature vectors are then fed as inputs to TWSVMs, which learns to classify an image block as a member or not a member of the object category, under some associated confidence. When we get the learned TWSVMs model, we can use the learned classifier as a reliable detector to perform the detection task.

## 4. Results and Discussion

Up till now, we have illustrated our new methods to detect MCs in mammograms. In this section, we will evaluate the

performance of our methods by using the real mammogram data obtained from DDSM. In our experiments, we use the training, test, and validation sets, which were randomly selected from the preprocessed image blocks. The blocks included 3000 with true MCs and 3000 with normal tissue. We chose 75% of the blocks for training and 25% for test.

For a given digital mammogram, we formulate the MCs detection and classification approach as the following steps.

*Step 1.* We first preprocess the mammogram to remove the artifacts, suppress inhomogeneity of the image background, and enhance microcalcifications in the breast area.

*Step 2.* At each pixel location of the image, we manually select a small window ( $\mathbf{x} = A_{m \times m}$ , where  $m = 115$ ) to describe its surrounding image feature.

*Step 3.* Get the sparse representation of each image block by using the proposed methods.

*Step 4.* Apply the proposed MCs detection methods to decide whether  $\mathbf{x}$  belongs to MCs or not.

In our experiments, they are designed to quantitatively verify the performance of sparse representation based methods for MCs detection and classification by using mammograms. To get an accurate performance measure in this study, a stratified 5-fold cross validation method is employed. We also compared our approaches with the state-of-the-art algorithm, SVMs, which have been successfully applied in MCs detection.

All the experiments are performed on a notebook computer with DUO Intel 2.54 G CPU and 4 G memory under Windows 7. MATLAB 2011 is used to implement sparse representation based MCs detection methods. To get the sparse representation of each image block, we employed the  $l_1$ - $l_s$  MATLAB package to achieve the optimization.  $l_1$ - $l_s$  is a toolbox in MATLAB implementation of the interior-point method for  $l_1$ -regularized least squares solution.

Before we performed the MCs detection methods, we first used the  $l_1$ - $l_s$  package to learn the sparse transform matrix with the training dataset. The trained sparse transform matrix

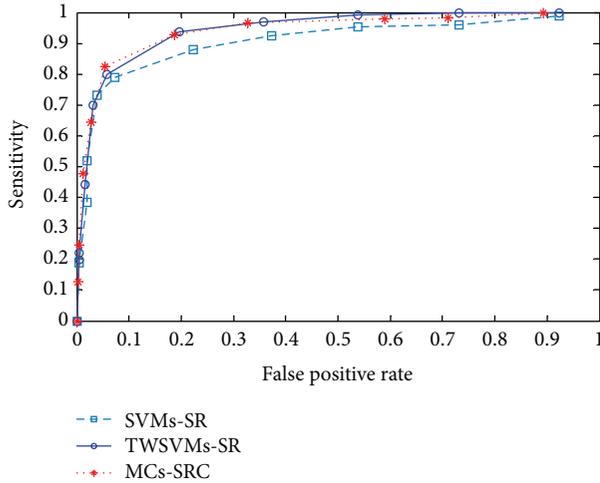


FIGURE 3: Comparisons of ROC curves of MCs detection and classification using the proposed methods.

is evaluated using all the mammograms in the test dataset. Because TWSVMs do not vary significantly over a wide range of parameter settings, we chose the fixed parameters of TWSVMs, having an RBF kernel with  $\sigma = 15$  and  $c_1 = c_2 = 1000$  in our experiments.

First, we did the experiments using MCs-SRC algorithm to perform the MCs detection and classification. The test results are summarized with ROC curves in Figure 3 for the MCs-SRC method. For comparison, ROC curve is also shown for the sparse representation based TWSVMs and SVMs methods with the same inputs.

From Figure 3, it can be shown that the proposed algorithm has a higher detection accuracy rate compared with SVMs and TWSVMs with the same dataset and configuration. By using the same test samples, compared with SVMs and TWSVMs, the proposed method has a better detection performance when we train the classifier. In particular, the proposed method achieved the averaged sensitivity of approximately 92.17% with respect to 7.83% false positive rate and  $Az = 0.9507$ . With the same training data set and test data set, the TWSVMs classifier achieved a sensitivity of 90.01%, 9.63% false positive rate, and  $Az = 0.9459$ , and SVMs achieved a sensitivity of 87.13%, 9.88% false positive rate, and  $Az = 0.9298$ .

To evaluate the stability of our methods, we repeat the sampling 50 times so that we can compute the mean and standard deviation of the detection accuracy, sensitivity, and specificity. We perform the detection and classification task in 20 rounds, and in each round we randomly select training samples from 95% of training samples to 5% to train classifiers. The trained classifiers are evaluated using the other 500 test samples. Average experimental results of the MCs-SRC method, compared with SVMs and TWSVMs, are shown in Table 1.

From Table 1, we can see that the MCs classification with Twin Support Machines and sparse representation (TWSVMs-SR) approach has a bit better performance than the other two. But as we have known, the MCs-SRC method

TABLE 1: Experimental results of the proposed MCs-SRC method for MCs detection, compared with sparse representation based TWSVMs and SVMs methods.

Methods	Sensitivity	Specificity	$Az$
MCs-SRC	$90.84 \pm 1.07\%$	$92.37 \pm 0.78\%$	$0.9407 \pm 0.0564$
TWSVMs-SR	$92.07 \pm 0.89\%$	$89.93 \pm 0.91\%$	$0.9678 \pm 0.0977$
SVMs-SR	$87.53 \pm 0.94\%$	$89.72 \pm 0.88\%$	$0.9304 \pm 0.1001$

does not need to train a classifier, because the features are acquired by learning, so it will get the classification more efficiently.

## 5. Conclusions

In this paper, a novel approach is described to aid breast cancer detection and classification using digital mammograms. The proposed method is based on sparse feature learning and representation, which expresses a testing sample as a linear combination of the built vocabulary (training samples). The sparse coefficient vector is obtained by using  $l_1$ -regularized least square method. MCs detection and classification are achieved by defining discriminating functions from the coefficient vector for each category (called MCs-SRC) and TWSVMs based algorithm using sparse representation as a feature learning method (called TWSVMs-SR). The demonstrated experiments show that TWSVMs-SR method gets the best performance, and MCs-SRC can match the performance achieved by the state-of-art classifier. Furthermore, the MCs-SRC approach does not need to select optimal model parameters of the used classifier.

## Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The author in this paper would like thank Shaanxi Provincial Education Department. All the work presented is supported by the key discipline of Shaanxi Province and the Scientific Research Program funded by Program no. 12JK0741. The author is also very grateful to the editor and anonymous reviewers who made their kind constructive comments and suggestions.

## References

- [1] D. Kopans, *Breast Imaging*, Lippincott Williams & Wilkins, New York, NY, USA, 2nd edition, 2006.
- [2] M. Le Gal, G. Chavanne, and D. Pellier, "Diagnostic value of breast clustered microcalcifications discovered by mammography (on 227 cases with histological verification and without an associated mass in the breast)," *Bulletin du Cancer*, vol. 71, no. 1, pp. 57–64, 1984.
- [3] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in

- mammograms: a survey," *Pattern Recognition*, vol. 36, no. 12, pp. 2967–2991, 2003.
- [4] C. Marrocco, M. Molinara, C. D'Elia, and F. Tortorella, "A computer-aided detection system for clustered microcalcifications," *Artificial Intelligence in Medicine*, vol. 50, no. 1, pp. 23–32, 2010.
  - [5] A. Vartician, L. Gavriluc, N. Godoroja et al., "610 method of differential breast tumour diagnostics," *European Journal of Cancer Supplements*, vol. 8, no. 3, pp. 236–237, 2010.
  - [6] L. C. S. Romualdo, M. A. C. Vieira, H. Schiabel et al., "Mammographic image denoising and enhancement using the Anscombe transformation, adaptive wiener filtering, and the modulation transfer function," *Journal of Digital Imaging*, vol. 26, no. 2, pp. 183–197, 2013.
  - [7] R. Nakayama, Y. Uchiyama, K. Yamamoto, R. Watanabe, and K. Namba, "Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 2, pp. 273–283, 2006.
  - [8] T. O. Gulsrud and J. H. Husøy, "Optimal filter-based detection of microcalcifications," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1272–1281, 2001.
  - [9] L. Bocchi, G. Coppini, J. Nori, and G. Valli, "Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks," *Medical Engineering and Physics*, vol. 26, no. 4, pp. 303–312, 2004.
  - [10] G. Rezaei-Rad and S. Jamarani, "Detecting microcalcification clusters in digital mammograms using combination of wavelet and neural network," in *Proceedings of the 2nd Conference on Computer Graphics, Imaging, and Vision: New Trends*, pp. 197–201, July 2005.
  - [11] L. Song, Q. Wang, and J. Gao, "Microcalcification detection using combination of wavelet transform and morphology," in *Proceedings of the 8th International Conference on Signal Processing (ICSP '06)*, November 2006.
  - [12] 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.
  - [13] J. M. Mossi and A. Albiol, "Improving detection of clustered microcalcifications using morphological connected operators," in *Proceedings of the 1999 7th International Conference on Image Processing and Its Applications*, pp. 498–501, July 1999.
  - [14] N. Riyahi-Alam, A. Ahmadian, J. N. Tehrani, M. Guiti, M. A. Oghabian, and A. Deldari, "Segmentation of suspicious clustered microcalcifications on digital mammograms: using Fuzzy logic and wavelet coefficients," in *Proceedings of the 7th International Conference on Signal Processing Proceedings (ICSP '04)*, pp. 2226–2228, September 2004.
  - [15] H. D. Cheng, J. R. Chen, R. I. Freimanis, and X. H. Jiang, "A novel fuzzy logic approach to microcalcification detection," *Information Sciences*, vol. 111, no. 1–4, pp. 189–205, 1998.
  - [16] H. Cheng, Y. M. Lui, and R. I. Freimanis, "A novel approach to microcalcification detection using fuzzy logic technique," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 442–450, 1998.
  - [17] H. D. Cheng, J. Wang, and X. Shi, "Microcalcification detection using fuzzy logic and scale space approaches," *Pattern Recognition*, vol. 37, no. 2, pp. 363–375, 2004.
  - [18] J. Jiang, B. Yao, and A. M. Wason, "Integration of fuzzy logic and structure tensor towards mammogram contrast enhancement," *Computerized Medical Imaging and Graphics*, vol. 29, no. 1, pp. 83–90, 2005.
  - [19] M. Mohanalin, B. Beenamol, P. K. Kalra, and N. Kumar, "A novel automatic microcalcification detection technique using Tsallis entropy & a type II fuzzy index," *Computers and Mathematics with Applications*, vol. 60, no. 8, pp. 2426–2432, 2010.
  - [20] C. D'Elia, C. Marrocco, M. Molinara, G. Poggi, G. Scarpa, and F. Tortorella, "Detection of microcalcifications clusters in mammograms through TS-MRF segmentation and SVM-based classification," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 743, pp. 742–745, August 2004.
  - [21] G. G. Lee and C. H. Chen, "Multiresolution wavelet analysis and Gaussian Markov random field algorithm for breast cancer screening of digital mammography," in *Proceedings of the 1996 IEEE Nuclear Science Symposium*, pp. 1737–1741, November 1996.
  - [22] S. Singh, V. Kumar, H. K. Verma, and D. Singh, "SVM based system for classification of microcalcifications in digital mammograms," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06)*, pp. 4747–4750, September 2006.
  - [23] X. Zhang and X. Gao, "Twin support vector machines and subspace learning methods for microcalcification clusters detection," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 5, pp. 215–221, 2012.
  - [24] Z. Xinsheng, G. Xinbo, and W. Ying, "Twin support tensor machines for MCs detection," *Journal of Electronics*, vol. 26, no. 3, pp. 318–325, 2009.
  - [25] R. R. Hernandez-Cisneros and H. Terashima-Marin, "Evolutionary neural networks applied to the classification of microcalcification clusters in digital mammograms," in *Proceedings of the IEEE International Conference on Evolutionary Computation (CEC '06)*, pp. 2459–2466, 2006.
  - [26] J. Jiang, P. Trundle, and J. Ren, "Medical image analysis with artificial neural networks," *Computerized Medical Imaging and Graphics*, vol. 34, no. 8, pp. 617–631, 2010.
  - [27] 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.
  - [28] A. K. Mohanty, M. R. Senapati, and S. K. Lenka, "An improved data mining technique for classification and detection of breast cancer from mammograms," *Neural Computing and Applications*, pp. 1–8, 2012.
  - [29] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.
  - [30] L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 24, no. 2, pp. 371–380, 2005.
  - [31] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, 2004.
  - [32] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

- [33] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale  $\ell_1$ -regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.
- [34] J. Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 905–910, 2007.
- [35] R. M. Nishikawa, "Mammographic databases," *Breast Disease*, vol. 10, no. 3-4, pp. 137–150, 1998.
- [36] G. Peters, C. M. Jones, and K. Daniels, "Why is microcalcification missed on mammography?" *Journal of Medical Imaging and Radiation Oncology*, vol. 57, no. 1, pp. 32–37, 2013.
- [37] J. B. Tilbury, P. W. J. van Eetvelt, J. M. Garibaldi, J. S. H. Curnow, and E. C. Ifeakor, "Receiver operating characteristic analysis for intelligent medical systems—a new approach for finding confidence intervals," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 952–963, 2000.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

