Improving the accuracy in detection of clustered microcalcifications with a context-sensitive classification model

Juan Wang

Department of Electrical and Computer Engineering, Medical Imaging Research Center, Illinois Institute of Technology, Chicago, Illinois 60616

Robert M. Nishikawa

Department of Radiology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

Yongyi Yang^{a)}

Department of Electrical and Computer Engineering, Medical Imaging Research Center, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 13 March 2015; revised 28 September 2015; accepted for publication 18 November 2015; published 22 December 2015)

Purpose: In computer-aided detection of microcalcifications (MCs), the detection accuracy is often compromised by frequent occurrence of false positives (FPs), which can be attributed to a number of factors, including imaging noise, inhomogeneity in tissue background, linear structures, and artifacts in mammograms. In this study, the authors investigated a unified classification approach for combating the adverse effects of these heterogeneous factors for accurate MC detection.

Methods: To accommodate FPs caused by different factors in a mammogram image, the authors developed a classification model to which the input features were adapted according to the image context at a detection location. For this purpose, the input features were defined in two groups, of which one group was derived from the image intensity pattern in a local neighborhood of a detection location, and the other group was used to characterize how a MC is different from its structural background. Owing to the distinctive effect of linear structures in the detector response, the authors introduced a dummy variable into the unified classifier model, which allowed the input features to be adapted according to the image context at a detection location (i.e., presence or absence of linear structures). To suppress the effect of inhomogeneity in tissue background, the input features were extracted from different domains aimed for enhancing MCs in a mammogram image. To demonstrate the flexibility of the proposed approach, the authors implemented the unified classifier model by two widely used machine learning algorithms, namely, a support vector machine (SVM) classifier and an Adaboost classifier. In the experiment, the proposed approach was tested for two representative MC detectors in the literature [difference-of-Gaussians (DoG) detector and SVM detector]. The detection performance was assessed using free-response receiver operating characteristic (FROC) analysis on a set of 141 screen-film mammogram (SFM) images (66 cases) and a set of 188 full-field digital mammogram (FFDM) images (95 cases).

Results: The FROC analysis results show that the proposed unified classification approach can significantly improve the detection accuracy of two MC detectors on both SFM and FFDM images. Despite the difference in performance between the two detectors, the unified classifiers can reduce their FP rate to a similar level in the output of the two detectors. In particular, with true-positive rate at 85%, the FP rate on SFM images for the DoG detector was reduced from 1.16 to 0.33 clusters/image (unified SVM) and 0.36 clusters/image (unified Adaboost), respectively; similarly, for the SVM detector, the FP rate was reduced from 0.45 clusters/image to 0.30 clusters/image (unified SVM) and 0.25 clusters/image (unified Adaboost), respectively. Similar FP reduction results were also achieved on FFDM images for the two MC detectors.

Conclusions: The proposed unified classification approach can be effective for discriminating MCs from FPs caused by different factors (such as MC-like noise patterns and linear structures) in MC detection. The framework is general and can be applicable for further improving the detection accuracy of existing MC detectors. © 2016 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4938059>]

Key words: computer-aided detection (CADE), clustered microcalcifications (MCs), false positives, mammography

1. INTRODUCTION

Breast cancer is the most frequently diagnosed nonskin cancer in women in the US. It is estimated that about 231 840 new

breast cancer cases and 40 290 breast cancer deaths will occur among women in the US in 2015.¹ Mammography is an effective screening tool for diagnosis of breast cancer, which can detect about 80%–90% of breast cancers in

women without symptoms.¹ One important early sign of breast cancer in mammograms is the appearance of clustered microcalcifications (MCs), which are found in 30%–50% of mammographically diagnosed cases.² MCs are tiny calcium deposits which appear as bright spots in mammograms. While often seen, individual MCs can be difficult to detect, because of their subtlety in appearance, variation in shape and size, and inhomogeneity in surrounding tissue.² Accurate detection of MCs is also a critical task in computer-aided diagnosis (CADx), where detected MCs are subject to subsequent analysis as being benign or malignant.

Due to their importance in early diagnosis, there have been great efforts in development of computerized methods for automatic detection of clustered MCs in mammogram images.³ Based on the signal/image processing techniques employed, these MC detection methods can range broadly from image enhancement (e.g., Refs. 4–6), stochastic modeling (e.g., Refs. 7–11), to modern machine learning (e.g., Refs. 12–17). The basic idea in image enhancement methods is to enhance the MC signals while suppressing the tissue background in a mammogram, for example, by difference-of-Gaussians (DoG) filter⁴ or wavelet analysis.^{5,6} In contrast, stochastic modeling methods are designed to exploit the statistical difference between MCs and their surroundings, which include, for example, higher-order statistics,⁷ Markov random field,^{8,9} Gaussian mixture models,¹⁰ and spatial point process models.¹¹ Different from these approaches, machine learning methods treat the MC detection as a two-class classification problem, wherein a decision function is determined with supervised learning from data examples in mammogram images.^{12–17} Such methods include, for example, neural network,^{12,13} boosted learning,¹⁴ relevant vector machine (RVM),¹⁵ and support vector machine (SVM).^{16,17}

Despite these efforts, however, the performance of computerized methods for MC detection is still far from being perfect. One often cited degrading factor in performance is the frequent occurrence of false positives (FPs) in detection when the sensitivity is set at a reasonable accuracy level. In MC detection, there are several known factors for causing FPs, including MC-like noise patterns, linear structures, inhomogeneity in tissue background, imaging artifacts, etc.¹⁸ Consequently, there exist several studies on how to suppress these factors in contributing to FPs. For example, noise equalization techniques were developed for reducing the noise variability in a mammogram,^{19,20} background removal methods were used to suppress the inhomogeneity in tissue background,²¹ and detection algorithms were studied for linear and curve-like structures in a mammogram (attributed to vessels, ducts, fibrous tissue, skinfolds, edges, and other anatomical features).^{22–24}

To deal with the effect of linear structures in MC detection, in our preliminary work,²⁵ we developed a bithresholding scheme, wherein the MC detector output was processed differently at linear structures from the rest of a mammogram according to their respective statistics. This scheme was based on the observation that the MC detector response gets elevated in linear structures and that MCs and linear

structures can overlap with each other in a mammogram. While conceptually simple, this bithresholding scheme was demonstrated to reduce significantly the FPs associated with linear structures for a SVM detector.²⁵

Built on the initial success in Ref. 25, in this work, we develop a unified classification approach to deal with FPs caused both by linear structures and by other noise sources (e.g., MC-like noise patterns, tissue inhomogeneity). For this purpose, we consider a classifier model to which the input features consist of two groups for characterizing the MC signal at a detection location. The first group consists of features defined by the image context at the detection location, which is used for characterizing how a MC is different from its structural background (i.e., presence or absence of linear structures). The second group consists of features derived from the image intensity pattern in the local neighborhood of the detection location. The latter is out of consideration that individual MCs are small in extent and well localized in a mammogram image. As to be explained in detail (Sec. 2.B), the features in the second group are extracted in order to both suppress the inhomogeneity in tissue background and enhance the different aspects of a MC signal at a detection location. To maximize differentiation of true MCs from FPs caused by both linear structures and MC-like noise patterns, we apply a feature selection procedure to determine the most discriminative features for use in the classifier model.

To accommodate the effect of linear structures on the detector response, we introduce a dummy variable in the unified classifier model to characterize the presence (or absence) of linear structures at a detection location. Through this variable, we can adapt the input features to the classifier according to the context of the detection location. Conceptually, the resulting classifier functions in two modes: one with linear structures present and one without. This is similar in spirit to the dual-thresholding scheme,²⁵ where the decision threshold was varied depending on the presence of linear structures; here, the decision function is varied according to the context of the detection location in the unified classifier model.

The proposed approach is general and can be applicable for FP reduction with any MC detectors. In this study, we demonstrated the proposed approach on two existing MC detectors, both of which have been well cited in the literature. The first is the DoG detector,⁴ which is of low computational complexity and representative of image enhancement methods for MC detection; the second is a SVM detector developed in Ref. 16, which is representative of machine learning methods. As to be seen from the experiments, these two detectors differed in their detection performance, thus serving as a good test bed for the proposed approach on MC detectors with varied detection accuracy levels.

Moreover, to demonstrate the flexibility of the proposed unified classifier approach, in the experiments, we implemented the classifier model with two commonly used machine learning algorithms. Among them, one is a SVM classifier with Gaussian kernel,²⁶ and the other is an Adaboost classifier with decision stumps.²⁷

The rest of the paper is organized as follows: in Sec. 2, we present the unified classifier framework, context-dependent feature extraction, and the two classifier models used in the implementation; in Sec. 3, we describe the procedure for demonstration and evaluation of the proposed unified classification approach; the performance results on reduction of FPs in MC detection are analyzed in Sec. 4 and discussions are given in Sec. 5. Finally, conclusions are given in Sec. 6.

2. METHODS

In this study, we formulate a two-class classification framework for reducing FPs in MC detection, wherein a pattern classifier is used to determine whether a detection from a MC detector in a mammogram image is a true positive (TP) or not. The classifier model is trained with supervised learning, for which a set of known detection examples (both TPs and FPs) from the MC detector is used.

2.A. Motivation and overview of the unified classifier model

As noted in the Introduction, there are a number of sources that can contribute to the occurrence of FPs in MC detection. To illustrate this, in Fig. 1(a), we show two examples of mammogram ROIs, both of which contain clustered MCs; for better visualization, the contrast in these images was adjusted so that the individual MCs become more visible. In Fig. 1(b), we show the output of the DoG detector in Ref. 4 when applied to these ROIs, wherein the intensity value indicates the likelihood of the presence of a MC signal at a given location; similar results are shown in Fig. 1(c) for the SVM

detector in Ref. 16. As can be seen, while the output by both detectors is notably higher at most of the MCs (i.e., correct detection), it also gets higher at many other locations in the two ROIs. Depending on the operating threshold, the latter would lead to potential FPs.

As noted from the examples in Fig. 1, the potential FPs in the detector output occur at two dominant sources, one is at locations along linear structures, and the other is at locations exhibiting MC-like noise patterns. In these examples, the effect of linear structures is noted to be particularly more pronounced in the detector response. The reason is that when examined locally, a small segment of a linear structure resembles the image features of a MC (i.e., higher contrast and image gradient), thereby triggering a higher response in the MC detector.

In order to accommodate FPs caused by different sources, in this work, we develop a unified classification approach in which we incorporate two separate groups of image features relevant to a MC signal in a mammogram. The first group of features is used to characterize the image intensity pattern in the local neighborhood of the detection location; for convenience, let this group of features be denoted collectively by vector $\mathbf{x}_m$. The second group of features is related to the image context of the detection location and is used to characterize how a MC object is different from its structural background. These context features will be defined differently depending on the presence or absence of a linear structure at the detection location. Let this group of features be denoted, respectively, by vector $\mathbf{x}_l$ when a linear structure is present, and by vector $\mathbf{x}_b$ when a linear structure is absent. The specific definitions of these features are given later in detail.

FIG. 1. (a) Two sample mammogram ROIs containing both MCs and linear structures; (b) corresponding output of DoG detector in Ref. 4, wherein MCs are indicated by “+,” and detections are indicated by green diamond symbols; and (c) corresponding output of SVM detector in Ref. 16. For both detectors, the output is noted to be higher at linear structures.

Next, to deal with the distinctive effect of linear structures on the detector response, we introduce a dummy variable D to indicate the presence ($D = 1$) or absence ($D = 0$) of a linear structure at a detection location. In traditional regression analysis,²⁸ dummy variables are used routinely to characterize the presence (or absence) of some categorical effects on the outcome. Here, we borrow this concept for classifying TPs from FPs in MC detection. By incorporating this variable into the classifier model, we aim to adapt the group of context features, i.e., $\mathbf{x}_l$ or $\mathbf{x}_b$, according to the presence or absence of a linear structure.

Specifically, let vector $\mathbf{x}$ denote the collection of the two groups of features introduced above, i.e., $\mathbf{x} = [\mathbf{x}_m, D\mathbf{x}_l, (1-D)\mathbf{x}_b, D]$. Then, the unified classifier model can be written in general as

$$f(\mathbf{x}) = F(\mathbf{x}_m, D\mathbf{x}_l, (1-D)\mathbf{x}_b, D), \quad (1)$$

where $F(\cdot)$ denotes the classifier function.

In this study, we consider a nonlinear model for the decision function $F(\cdot)$ based on supervised learning, as to be described below in detail. For the purpose of illustration, consider the special case of a linear model. The corresponding classifier model for Eq. (1) can be written in the following form:

$$f(\mathbf{x}) = \beta_1^T \mathbf{x}_m + \beta_2^T D\mathbf{x}_l + \beta_3^T (1-D)\mathbf{x}_b + \beta_d D + \beta_0, \quad (2)$$

where $\beta_0, \beta_1, \beta_2, \beta_3$, and β_d denote the corresponding model coefficients of the different feature components. Specifically, β_2 and β_d together control the effect when a linear structure is present, while β_3 controls the effect when it is absent. In contrast, β_0 and β_1 control the common effect both when a linear structure is present and when it is absent.

Conceptually, the unified classifier model in Eq. (1) operates in two different modes: one for $D = 1$, and the other for $D = 0$. Of course, one may accomplish this alternatively by using two separate classifier models, one for locations with linear structures and the other for locations without. However, with such an alternative approach, the two classifiers would need to be trained separately, which is suboptimal in utilizing the training data. In contrast, the unified model in Eq. (1) will benefit from training samples both with and without linear structures. This is illustrated by the common effect coefficients β_0 and β_1 in the special case in Eq. (2). In addition, when two separate classifiers are used in operation, it is not clear how to optimally balance their individual decision levels in order to achieve a given performance level.

2.B. Classifier input features

As described above, we consider for the classifier model the two broad groups of image features that are relevant to MC signals. Below, we describe these features in detail.

2.B.1. Features on local image intensity pattern

To describe the MC signal at a detection location, we use the image values extracted from those pixels within a local neighborhood of the location. This is out of consideration that

MCs are small objects (typically 0.1–1.0 mm in diameter²⁹) and well localized in a mammogram image. Specifically, we define a circular window of diameter 0.7 mm centered at a detection location and extract the image pixels within this window. Such a window is chosen to be large enough to cover a MC object and its immediate surrounding background, while avoiding potential interference from any nearby structures (e.g., neighboring MCs) at a detection location. With a spatial resolution of 50 $\mu\text{m}/\text{pixel}$, this circular window consists of 121 pixels (as illustrated Fig. 2). The image values at these pixels are arranged into a vector for input to the classifier.

2.B.2. Features on context of MCs

To characterize the background at a detection location, we extract a set of quantitative features from a local structural neighborhood of the location. Depending on the presence or absence of linear structures, these features will be computed differently at a detection location.

2.B.2.a. Structural contrast of MC. When a MC is present, the image intensity is noted to be higher than its surroundings. To quantify the signal strength relative to its background, we extract the image contrast at detection location (u, v) as

$$C_0 = \frac{I(u, v) - \mu_0}{\sigma_0}, \quad (3)$$

where $I(u, v)$ is the image intensity at (u, v) , and μ_0 and σ_0 denote the mean and standard deviation, respectively, of the image intensity values of the structural background at (u, v) . Specifically, when a linear structure is present at (u, v) , μ_0 and σ_0 are estimated from only those image pixels along the associated linear structure; otherwise, they are estimated from the pixels within a circular window of radius 0.7 cm centered at (u, v) (as defined in Sec. 2.B.1). Thus, when a linear structure is present at a detected location, the associated upward bias in the image signal is automatically suppressed in the resulting contrast measure C_0 .

Fig. 2. Illustration of local circular window used for extracting image features.

2.B.2.b. Statistics of structural background. To quantify the image characteristics at a detection location (u,v) , we use the mean and standard deviation of the image intensity in a structural neighborhood of (u,v) , denoted by μ_1 and σ_1 , respectively. Specifically, when no linear structure is present at (u,v) , the image pixels within a 5×5 window centered at (u,v) are used [which is approximately the average size of a MC with resolution at $50 \mu\text{m}/\text{pixel}$ (Ref. 29)]; otherwise, the image pixels within the associated linear structure are used. Thus, both μ_1 and σ_1 are expected to be higher in value when a MC is present at (u,v) ; when a linear structure is present, μ_1 and σ_1 serve as a measure of the strength and variability of the linear structure, respectively.

Altogether there are a total of three context features, namely, C_0 , μ_1 , and σ_1 , extracted at a given detection location (u,v) , which are computed according to the structural background at (u,v) . When a linear structure is present, these features are used to form $\mathbf{x}_l$ ($D=1$) for the classifier; otherwise, they are used to form $\mathbf{x}_b$ ($D=0$). As to be demonstrated later in the experiments, these structure related features are useful for discriminating true MCs from FPs both when linear structures are present and when they are absent.

2.B.3. Multidomain feature extraction for MC enhancement

In a mammogram image, the detectability of MCs is greatly hampered by the inhomogeneity in their surrounding background (e.g., variation in tissue density). To deal with this difficulty, we preprocess the mammogram image in the following three ways prior to feature extraction: (1) remove the local mean at a detection location so that the image intensity pattern (Sec. 2.B.1) has zero mean (denoted as *Org*); (2) prefilter the image with a narrow-band high-pass filter¹⁶ (denoted as *HP*), and (3) prefilter the image with the DoG filter⁴ (denoted as *DoG*).

The above processing steps are intended to suppress the image background and enhance different aspects of the MC signals. Specifically, *Org* is applied to subtract the local background without altering the frequency content of a MC signal at a detection location; *HP* is for enhancing the edges of a MC while suppressing the image background; and finally, *DoG* is to enhance the contrast of a MC relative to its background (the DoG filter is a band-pass filter). Thus, these different processed images, namely, *Org*, *HP*, *DoG*, can complement each other for characterizing the MCs in a mammogram. For convenience, these three images are referred to as three separate domains hereafter.

Subsequently, feature extraction is applied separately for each of the three domains. At a detection location, this will result in a total of 363 features (121 per domain) on the local image intensity pattern and nine features (three per domain) on the image context. These features are then used to form the vector $\mathbf{x} = [\mathbf{x}_m, D\mathbf{x}_l, (1-D)\mathbf{x}_b, D]$. Of course, these features will be highly redundant for a detection location, which is especially true for the 363 features in $\mathbf{x}_m$. To deal with this issue, we employ a feature selection

procedure to determine among them the most discriminative features for our classification task in Sec. 3.D.2. As to be demonstrated in the experiments, these multidomain features are more powerful than those from individual domains for discriminating MCs from FPs.

2.C. Classifier implementation

The proposed unified classifier model in Eq. (1) can be implemented with any traditional classifiers. In this study, we demonstrate this approach on two types of classifiers: SVM, and Adaboost. Both classifiers are found to have good generalization performance and are widely used in various machine-learning problems. For example, SVM was applied in handwritten digit recognition,³⁰ object recognition,³¹ face detection,³² and text categorization;³³ and Adaboost was also applied in face detection,³⁴ text categorization,³⁵ as well as music recommendation.³⁶

2.C.1. Classifier implementation with SVM

SVM (Ref. 26) is a supervised learning algorithm based on the principle of structural risk minimization. For notational simplicity, consider the case of a linear classifier,

$$f(\mathbf{x}) = \beta^T \mathbf{x} + \beta_0, \quad (4)$$

where β and β_0 are unknown parameters to be determined from the training dataset $\{(\mathbf{x}_i, y_i), i = 1, 2, \dots, N\}$. In a SVM formulation, these parameters are solved from the following structural risk minimization:²⁶

$$\begin{aligned} \min_{\beta, \beta_0} \quad & \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n \varepsilon_i \\ \text{subject to} \quad & y_i f(\mathbf{x}_i) \geq 1 - \varepsilon_i, \text{ and } \varepsilon_i \geq 0, i = 1, 2, \dots, n. \end{aligned} \quad (5)$$

In Eq. (5), the parameter C is used to control the trade-off between the model complexity and empirical risk.

The above formulation can be extended to the case of a nonlinear classifier by using the “kernel trick,” where the input vector $\mathbf{x}$ is first mapped to a higher-dimensional space via a nonlinear mapping, which is implicitly defined by a kernel function $K(\cdot, \cdot)$, and then classified by a linear classifier in this mapped space. The resulting classifier function can be rewritten as

$$f(\mathbf{x}) = \sum_{k=1}^{N_s} y_k \alpha_k K(\mathbf{x}, \mathbf{s}_k) + b, \quad (6)$$

where $\mathbf{s}_k$, $k = 1, 2, \dots, N_s$ are the so-called support vectors, which are a subset of the training samples. Both the support vectors and the coefficients α_k in Eq. (6) are determined from the training data via structural risk minimization.

In this study, we use the Gaussian RBF kernel, which has the following form:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \quad (7)$$

where $\sigma > 0$ is a parameter that defines the kernel width.

2.C.2. Classifier implementation with Adaboost

Adaboost²⁷ is a boosting learning algorithm to form a committee decision function from a sequence of weak learners (whose individual performance is only slightly better than random guessing). With input $\mathbf{x} = [\mathbf{x}_m, D\mathbf{x}_l, (1-D)\mathbf{x}_b, D]$, the Adaboost classifier function can be written as

$$f(\mathbf{x}) = \sum_{i=1}^M \alpha_i f_i(\mathbf{x}), \quad (8)$$

where $f_i(\mathbf{x})$, $i = 1, 2, \dots, M$, are the weaker classifiers, and their corresponding weighting factors α_i are determined based on their classification accuracy during training.²⁷

In this study, we use decision stumps for the weak classifiers $f_i(\mathbf{x})$ owing to their computational simplicity. A decision stump is a decision tree with only two terminal nodes.³⁴ Accordingly, the classifier $f_i(\mathbf{x})$ in the i th iteration of Adaboost is computed as following:

$$f_i(\mathbf{x}) = \begin{cases} 1, & x^{(i)} \geq T_i \\ -1, & x^{(i)} < T_i \end{cases} \quad (9)$$

where $x^{(i)}$ denotes the optimal feature component determined in step i , and T_i denotes the corresponding decision threshold. Both $x^{(i)}$ and T_i are optimized accordingly to the weighted error criterion during the training step.²⁷

3. PERFORMANCE EVALUATION

3.A. Mammogram datasets

To demonstrate the proposed approach, we first make use of a dataset of screen-film mammogram (SFM) images collected by the Department of Radiology at the University of Chicago (U of C). The dataset consists of a total of 141 mammogram images from 66 cases (32 cancer/34 benign). These cases were collected consecutively over time and were all sent for biopsy due to the subtlety of their MC lesions. Each mammogram has at least one cluster of MCs which were histologically proven. These mammograms were digitized with a spatial resolution of 50 $\mu\text{m}/\text{pixel}$ and 12-bit grayscale with a dimension of 3000 $\times$ 5000 pixels using a Lumiscan film digitizer (Lumisys; Sunnyvale, CA). The MCs in each mammogram were manually identified by a group of experienced radiologists. Among the 66 cases in the dataset, the clustered MCs were rated as visible in 35 cases and as subtle in the other 31 cases by the readers. There were a total of 6654 MCs in these mammograms, which are used as the ground truth in our evaluation studies. In Fig. 3, we show a box-whisker plot of the number of MCs in individual cases separately for benign and malignant cases in the dataset. The median number of MCs was 18 for benign cases and 24.5 for malignant cases.

In our experiment implementation, the dataset was randomly partitioned into two subsets, each containing 33 cases, one used for training and the other for testing. It is noted that most of the cases have more than one view (mediolateral oblique view, craniocaudal view, or views from both breasts). To avoid any potential bias, the different views

from one case were always assigned together to either the training or testing subset, but never both. The training subset was used for optimizing the associated parameters of the classifiers, and the testing subset was used exclusively for performance evaluation.

In addition, we also demonstrated the proposed approach on a set of full-field digital mammogram (FFDM) images collected at U of C. It consists of a total of 188 mammogram images from 95 cases (43 cancer/52 benign). Each mammogram has at least one cluster of MCs. They were acquired using a Senographe 2000D FFDM system (General Electric Medical Systems; Milwaukee, WI) with a spatial resolution 100 $\mu\text{m}/\text{pixels}$. As in the SFM dataset above, the MCs in each mammogram were manually identified, and there were a total of 8928 MCs identified. The dataset was randomly divided into two subsets: one with 47 cases for training and the other with 48 cases for testing.

3.B. Preprocessing and feature extraction

For each SFM image, a noise equalization procedure¹⁹ was applied for equalizing the intensity-dependent noise in the image. To suppress the variability among different cases, each mammogram image was further normalized to have zero mean and unit standard-deviation over the tissue area upon preprocessing, respectively, by *Org*, *HP*, or *DoG* (Sec. 2.B.3).

For each mammogram image in the dataset, a linear structure detection procedure²⁵ was applied to detect the location map of linear structures. This location map was used subsequently to set the dummy variable D in the classifier model. The detection procedure in Ref. 25 was based on the line operator in Ref. 37, which was demonstrated to be effective for detecting linear structures in mammograms.²³ The basic idea of this line operator is to apply directional template-matching with different orientation angles; the directional template detector achieves its maximum output when its orientation is in perfect alignment with a linear structure at a detection location. The line detector output was subsequently subject to a discriminant analysis against false detections; the discriminant parameters were determined from a set of training mammograms.²⁵

Fig. 3. The box-whisker plot of the number of MCs over individual cases in the SFM dataset with 66 cases.

3.C. MC detectors for demonstration

As mentioned earlier in the Introduction, in this study, we demonstrate the proposed approach on two existing MC detectors: the DoG detector in Ref. 4 and the SVM detector developed in Ref. 16. In our implementation, the operating point for each of the two detectors was set at a level such that 90% of the marked MCs are detected on the set of training images. This is to insure a high sensitivity level in the detection. Of course, a high sensitivity level (i.e., TP rate) will also inevitably lead to many FPs in detection. Our unified classifier was applied subsequently to discriminate FPs from true MCs in the detection.

3.D. Classifier model training and optimization

3.D.1. Preparation of training dataset

For optimizing the unified classifier model, we first extracted a set of TP samples and FP samples from the set of training mammogram images. Specifically, from each mammogram image, all the marked MCs were extracted as TP samples (i.e., $y_i = 1$); afterward, twice as many FP samples (i.e., $y_i = -1$) were selected randomly from the detection results of the MC detector on the image (either DoG or SVM). These FP samples were equally distributed among sites both with (i.e., $D = 1$) and without linear structures ($D = 0$).

3.D.2. Feature selection

To determine the most useful features for discriminating MCs from FPs among the set of extracted features (Sec. 2.B), we apply a sequential forward feature selection (FSS) procedure.³⁸ The procedure starts with an initial empty set and successively adds to this set those features that can improve the classification accuracy the most in an iterative fashion. In the experiments, we used the logistic regression classifier for feature selection with the FSS procedure owing to its low computational complexity. For this purpose, the training dataset was randomly divided into two equal-sized subsets, of which one was used for training the classifier with the selected features, and the other for assessing its classification accuracy at each step. During feature selection, the number of selected features was varied from 50 with an increment of 10; in the end, the set of features with the best classification accuracy was selected for subsequent development.

3.D.3. Classifier training and optimization

For the unified SVM classifier, we need to determine the parameter C in Eq. (5) and the kernel width σ in Eq. (7). For this purpose, we apply a four-fold cross-validation procedure³⁸ on the training dataset, for which the training samples are randomly divided into four equal-sized subsets. Afterward, each of the four subsets is held out in turn for testing while the rest three subsets are used together for training the classifier. In the end, the parametric setting with the smallest test error is chosen. The classifier is then retrained with all the training

samples at this selected setting. Subsequently, the classifier is applied on the set of test mammograms for performance evaluation. In the experiments, a grid search was used for the parameters in which the following values were used: $\{0.1, 1, 10, 100\}$ for C and $\{5, 10, 15, 20, 25\}$ for kernel width σ .

Similarly, the above procedure is also applied for the unified Adaboost classifier in which the parameter M in Eq. (8) is determined.

3.D.4. Performance evaluation using free-response receiver operating characteristic (FROC)

To evaluate the detection performance, we conduct a FROC analysis, which is routinely used for CAD performance evaluation. A FROC curve is a plot of the TP rate of MC clusters versus the average number of FPs per image with the decision threshold varied continuously over its operating range.

In the FROC analysis, the detected MCs in a mammogram are first grouped into clusters. In our experiments, we used the criterion described in Ref. 39, which was reported to yield more realistic performance than several other alternatives. Specifically, a group of objects is considered to be a cluster if the objects are connected with nearest-neighbor distances less than 0.5 cm and there are at least three objects within a square area of 1 cm². A detected cluster is considered as a TP cluster if (1) it includes at least two true detected MCs and (2) its center of gravity is within 1 cm of that of a known true MC cluster. Likewise, a detected cluster is considered as a FP cluster provided that (1) it contains no true MCs, or (2) the distance between its center of gravity and that of any known cluster is larger than 1 cm.

It is noted that FROC curves can be sensitive to the detection criteria used.³⁹ However, the relative ordering of different detectors tends to be the same with the same criteria.¹⁶ In this study, we mainly focus on the operating range when the FP rate is not more than 2 FP clusters per image, which is of main interest in clinical practice.

To reduce the effect of case variation, we apply a bootstrapping procedure on the set of test mammograms for obtaining the FROC.⁴⁰ A total of 20 000 bootstrap samples are used, based on which the partial area under the FROC curve ($pAUC$) is obtained. This bootstrapping procedure is also used to perform statistical comparison of the performances between two methods.⁴⁰

To speed up the FROC analysis, in the experiments, we first applied a prescouting step as in Ref. 11, during which only the most suspicious regions in a mammogram image were identified for further consideration. This was based on the fact that the spatial extent of clustered MCs is typically limited (<1 cm² in area), while a whole mammogram image can be rather large in size. For each image, a set of MC candidates were first detected by the MC detector under test. The detection threshold was set to be four standard-deviations above the mean value of the detector output for the whole image, and only those candidates sufficiently large in size (over 0.03 mm²) were kept. Afterward, the detected candidates were grouped into clusters, and a square region of

5×5 cm was extracted for each cluster at its center of gravity. Note that the region was chosen to be much larger in area than a typical MC cluster. In the end, up to four such regions (with most detected candidates) in each image were selected for further processing by the proposed approach. The number of detections was counted as the actual number of cluster candidates in all the regions according to the detection criteria.

For the purpose of comparison, we also tested a bithresholding scheme for reducing FPs caused by linear structures in MC detection²⁵ and a detection algorithm based on spatial point processing (SPP) modeling.¹¹

4. RESULTS

For clarity, we present the evaluation results achieved by the proposed unified approach on FP reduction for the two MC detectors in separate subsections. For each MC detector, we give the results for both of the unified classifiers implemented in this study (i.e., unified SVM and unified Adaboost). To avoid potential confusion, we first give the results on the SFM dataset and then give the results on the FFDM dataset in Sec. 4.D.

4.A. FP reduction for DoG detector

4.A.1. Unified SVM classifier

In Fig. 4, we show the FROC curve obtained by the baseline DoG detector, together with the resulting FROC curve upon FP reduction with the proposed unified SVM classifier. As can be seen, the FROC curve gets notably higher (hence better detection performance) for the unified SVM. A statistical comparison between the unified SVM and the baseline DoG yields a difference of 0.3572 in $pAUC$ (p -value $< 1.0 \times 10^{-4}$) for FP rate over the range of [0,1] clusters/image. For consistency, this same FP range is also used for $pAUC$ values given subsequently. In particular, with TPF at 85%, the FP rate is reduced from 1.16 (DoG) to 0.33 clusters/image (unified SVM), nearly a 71.6% reduction. Moreover, with FP rate at 0.5 clusters/image, the sensitivity is improved from 53.0% (DoG) to 89.5% (unified SVM).

For comparison, the FROC curve obtained with the bithresholding scheme applied to the DoG detector response is also shown in Fig. 4. As can be seen, the unified SVM classifier is much higher than the bithresholding scheme (difference in $pAUC = 0.3165$, p -value $< 1.0 \times 10^{-4}$).

For the unified SVM, the parameter values determined from training were as follows: $\sigma = 20$ and $C = 10$. Also, the number of features used was 100, which were determined from the feature selection procedure (Sec. 3.D.2). These features are indicated in Fig. 6(a) and further discussed later in Sec. 4.C.

4.A.2. Unified Adaboost classifier

For ease of comparison, we show also in Fig. 4 the resulting FROC curve upon FP reduction with the unified Adaboost

Fig. 4. Comparison of FROC curves obtained with (1) DoG detector (baseline), (2) FP reduction with unified SVM classifier (unified SVM), and (3) FP reduction with unified Adaboost classifier (unified Adaboost). For comparison, FP reduction with a bithresholding scheme is also shown (Bithresholding).

classifier. As can be seen, the unified Adaboost is similar in performance to the unified SVM (difference in $pAUC = 0.0245$; p -value = 0.8324). The FROC curve of the unified Adaboost is also notably higher than that of the baseline DoG (difference in $pAUC = 0.3328$, p -value $< 1.0 \times 10^{-4}$). In particular, with TPF at 85%, the FP rate is reduced from 1.16 (DoG) to 0.36 clusters/image (unified Adaboost), nearly a 69.0% reduction; likewise, with FP rate at 0.5 clusters/image, the sensitivity is improved from 53.0% (DoG) to 88.0% (unified Adaboost).

For the unified Adaboost, the parameter M was determined to be 100 from the cross-validation procedure (Sec. 3.D.3). Note that the same input features were used for both the unified Adaboost and the unified SVM.

4.B. FP reduction for SVM detector

4.B.1. Unified SVM classifier

In Fig. 5, we show the FROC curve obtained by the baseline SVM detector, together with the resulting FROC curve upon FP reduction with the proposed unified SVM classifier. As can be seen, the FROC curve gets notably higher for the unified SVM (difference in $pAUC = 0.1309$, p -value = 0.0004). In particular, with TPF at 85%, the FP rate is reduced from 0.54 (baseline SVM) to 0.30 clusters/image (unified SVM), nearly a 44.4% reduction. Moreover, with FP rate at 0.5 clusters/image, the sensitivity is improved from 83.4% (baseline SVM) to 90.8% (unified SVM).

For comparison, the FROC curves obtained with the bithresholding scheme to the baseline SVM detector response and with the SPP method are also shown in Fig. 5. The unified SVM classifier is noted to be higher than the bithresholding scheme (difference in $pAUC = 0.1047$, p -value = 0.0042). Similarly, the unified SVM is also higher than the SPP method (difference in $pAUC = 0.0611$, p -value = 0.0407).

FIG. 5. Comparison of FROC curves obtained with (1) SVM detector (baseline), (2) FP reduction with unified SVM classifier (unified SVM), and (3) FP reduction with unified Adaboost classifier (unified Adaboost). For comparison, FP reduction with the bithresholding scheme (bithresholding) and SPP method is also shown.

For the unified SVM classifier, the parameter values determined from training were as follows: $\sigma = 10$ and $C = 10$. Also, the number of features used was 60, which were determined from the feature selection procedure (Sec. 3.D.2). These features are indicated in Fig. 6(b) and further discussed later in Sec. 4.C.

4.B.2. Unified Adaboost classifier

For ease of comparison, we show also in Fig. 5 the resulting FROC curve upon FP reduction with the unified Adaboost classifier. As can be seen, the unified Adaboost is similar in performance to the unified SVM above (difference in $pAUC = 0.0138$; p -value = 0.6905). The FROC curve of the unified Adaboost is also notably higher than that of the baseline detector (difference in $pAUC = 0.1171$, p -value = 0.0077). In particular, with TPF at 85%, the FP rate is reduced from 0.54 (baseline SVM) to 0.25 (unified

Adaboost) clusters/image, nearly a 53.7% reduction; likewise, with FP rate at 0.5 clusters/image, the sensitivity is improved from 83.4% (baseline SVM) to 89.4% (unified Adaboost).

For the unified Adaboost classifier, the parameter M was determined to be 180 from the cross-validation procedure (Sec. 3.D.2). The same input features were used for both the unified Adaboost and the unified SVM.

4.C. Input features

As mentioned above (Sec. 4.A), a total of 100 features were used in the unified classifiers for FP reduction in the detection output of the DoG detector. A further examination of these features revealed that they consisted of the following components: feature D , 10 context features from $D\mathbf{x}_l$ and $(1-D)\mathbf{x}_b$ (Sec. 2.B.2), and 89 features from the image intensity pattern (Sec. 2.B.1). These context features were (1) for $D = 0$, C_0 and μ_1 from *Org*, σ_1 from *DoG*; and (2) for $D = 1$, C_0 from both *HP* and *Org*, σ_1 from *HP*, *DoG*, and *Org*, μ_1 from *Org* and *DoG*. Similarly, among the 89 features on image intensity pattern, 35 were from *HP*, 18 from *DoG*, and 36 from *Org*; the corresponding pixel locations of these features are indicated in Fig. 6(a).

Similarly, there were 60 features used in the unified classifiers for the output of the SVM detector. They consisted of the following: feature D , 7 context features for $D\mathbf{x}_l$ and $(1-D)\mathbf{x}_b$, and 52 features from the image intensity pattern. These context features were (1) for $D = 0$, σ_1 from *Org*; and (2) for $D = 1$, C_0 from both *HP* and *Org*, σ_1 from *HP*, *DoG*, and *Org*, μ_1 from *Org*. Similarly, among the 52 features on the image intensity pattern, 22 were from *HP*, 12 from *DoG*, and 18 from *Org*; the corresponding pixel locations of these features are indicated in Fig. 6(b).

These results indicate that features from the three domains, namely, *Org*, *HP*, and *DoG*, all contributed to the discrimination between MCs and FPs in the unified classifiers. That is, the features extracted from the three domains complement each other and can be more powerful than that from individual domains. This issue is further examined subsequently in Sec. 5.B.

FIG. 6. (a) Pixel locations of the 100 selected features in three domains for the DoG detector (red circle: *HP*, green plus sign: *DoG*, blue triangle: *Org*); (b) pixel locations of the 60 selected features in three domains for the SVM detector.

FIG. 7. Comparison of FROC curves obtained with (1) DoG detector (baseline), (2) FP reduction with unified SVM classifier (unified SVM), and (3) FP reduction with unified Adaboost classifier (unified Adaboost). For comparison, FP reduction with a bithresholding scheme is also shown (Bithresholding).

4.D. MC detection in FFDM images

4.D.1. FP reduction for DoG detector

In Fig. 7, we show the FROC curves obtained on the FFDM dataset by (1) the baseline DoG detector, (2) the bithresholding FP reduction scheme, (3) FP reduction with the unified SVM classifier, and (4) FP reduction with the unified Adaboost classifier. As can be seen, the FROC curves get notably higher for both the unified SVM and unified Adaboost, which are similar in performance (difference in $pAUC = 0.0043$, $p\text{-value} = 0.5661$). The difference in $pAUC$ between the unified SVM and the baseline DoG is 0.1574 ($p\text{-value} = 0.0008$). In particular, with TPF at 85%, the FP rate is reduced from 1.59 (DoG) to 0.28 clusters/image (unified SVM), nearly a 82.4% reduction. With FP rate at 0.5 clusters/image, the sensitivity is improved from 74.6% (DoG) to 88.9% (unified SVM). Moreover, the FROC curve of the unified SVM classifier is notably higher than that of the bithresholding scheme (difference in $pAUC = 0.1100$, $p\text{-value} = 0.0071$).

4.D.2. FP reduction for SVM detector

We also tested the unified classifiers for FP reduction with the SVM detector on the FFDM dataset. In the interest of brevity, the obtained FROC curves by the different methods are not shown. As above, the unified SVM and unified Adaboost are similar in performance (difference in $pAUC = 0.0032$, $p\text{-value} = 0.4118$). The FROC curve of the unified SVM is higher than that of the baseline SVM (difference in $pAUC = 0.1211$, $p\text{-value} = 0.0001$). With TPF at 85%, the FP rate is reduced from 0.98 (baseline SVM) to 0.37 clusters/image (unified SVM), nearly a 61.9% reduction. Moreover, with FP rate at 0.5 clusters/image, the sensitivity is improved from 76.3% (baseline SVM) to 86.6% (unified SVM). Furthermore, the unified SVM classifier is higher than the bithresholding scheme (difference

in $pAUC = 0.0747$, $p\text{-value} = 0.0040$) and the SPP method (difference in $pAUC = 0.0759$, $p\text{-value} = 0.0049$).

5. DISCUSSIONS

5.A. FP reduction performance for different MC detectors

By comparing the results in Figs. 4 and 5, it is interesting to note that the FROC curves achieved are similar for the two MC detectors (i.e., DoG and SVM) upon FP reduction by the proposed unified classifiers. Consider the unified Adaboost, for example, with TPF = 85%, the FP rate is reduced to 0.36 clusters/image for the DoG detector and 0.25 clusters/image for the SVM detector. A statistical comparison between the two yields a difference of 0.0361 in $pAUC$ ($p\text{-value} = 0.1200$). Similarly, for the unified SVM, with TPF = 85%, the FP rate is reduced to 0.33 clusters/image for the DoG detector and 0.30 clusters/image for the SVM detector. A statistical comparison between the two yields a difference of 0.0077 in $pAUC$ ($p\text{-value} = 0.4067$). These results indicate that the proposed unified classifier approach can be equally effective for MC detectors with different accuracy levels. While the FP rate is higher in the DoG detector than the SVM detector, the unified classifiers could reduce the FP rate to a similar level for the two detectors.

Furthermore, from Figs. 4 and 5, it is noted that the improvement achieved by the unified classifiers over the baseline is much higher in the DoG detector (Fig. 4) than that in the SVM detector (Fig. 5). However, it is also noted that the bithresholding scheme achieved similar improvement over the baseline for the two detectors. This is due to the fact that the bithresholding scheme is aimed for reducing FPs mainly caused by linear structures. Compared to the SVM detector, the DoG detector yields more FPs (associated with MC-like noise patterns) for a given TP level; this was also illustrated in the examples earlier in Fig. 1. These results demonstrate that the unified classifiers are indeed effective for removing FPs caused by both linear structures and MC-like noise patterns.

5.B. Saliency of multidomain features

The results in Sec. 4.C show that the most discriminative input features to the unified classifiers consist of features from all three domains, i.e., *Org*, *HP*, and *DoG*. To illustrate how these multidomain features complement each other, we further investigated and compared their discriminating power against their single-domain counterparts. For this purpose, we replaced in the unified classifiers the multidomain features by their respective counterparts in each single domain, i.e., *Org*, *HP*, or *DoG*, and compared their FP reduction performance.

In Fig. 8, we show the FROC curves obtained on the SFM dataset by the unified Adaboost classifier with features from individual *Org*, *HP*, and *DoG* domains for the DoG detector. For comparison, the FROC curve obtained with the multidomain features is also shown. As can be seen, the FROC curves from the individual domains are notably lower than that of multidomain. A statistical comparison between the

FIG. 8. Comparison of FROC curves obtained with (1) multidomain features (multidomain), (2) *Org* domain features (*Org* domain), (3) *HP* domain features (*HP* domain), and (4) *DoG* domain features (*DoG* domain).

multidomain and the individual domains yields the following differences in $pAUC$: 0.0928 (p -values = 0.0006), 0.0365 (p -values = 0.0972), and 0.1008 (p -values = 0.0081), for *Org*, *HP*, and *DoG*, respectively.

Similar comparison results were also obtained for the SVM detector with the Adaboost classifier using multidomain features. A statistical comparison between the multidomain and the individual domains yields the following differences in $pAUC$: 0.0531 (p -values = 0.0427), 0.0561 (p -values = 0.0207), and 0.1306 (p -values = 0.0001), for *Org*, *HP*, and *DoG*, respectively. The FROC curves are omitted here for brevity.

5.C. Limitations

One note is that the cases used in this study all contain clustered MCs in evaluating the detection performance. It might be desirable to also include a number of normal cases (i.e., without any MCs). However, we note that the spatial extent of clustered MCs in a mammogram is typically well localized within a small area ($<1 \text{ cm}^2$); the overwhelming majority of area in a mammogram image does not have any MCs and thus can be viewed as a substitute for normal mammograms in terms of FP detections. Therefore, the reported FP rate per image in the FROC analysis will likely remain little changed when normal cases are included.

6. CONCLUSION

In this study, we investigated a unified classification approach for suppressing FPs caused by different sources in MC detection, which include MC-like noise patterns, linear structures, inhomogeneity in tissue background, etc. To accommodate the effect of these heterogeneous sources, we developed a classifier model to which the input features were adapted according to the image context at a detection location through the use of a dummy variable; for characterizing

the MC signals, the input features were defined from both the local image intensity pattern and the image context at a detection location; furthermore, to suppress the effect of inhomogeneity background, the input features were extracted from three different domains aimed for enhancing the MCs in a mammogram image. We implemented the proposed approach with two commonly used machine learning algorithms, namely, SVM and Adaboost. We tested these unified classifiers on two different MC detectors and evaluated their performance using FROC analysis on a set of 141 SFM images from 66 cases and 188 FFDM images from 95 cases. The results show that the proposed unified classifiers could be equally effective for improving the detection accuracy of the two MC detectors; the unified classification approach could significantly reduce the FPs caused by both MC-like noise patterns and linear structures.

ACKNOWLEDGMENT

This work was supported by NIH/NIBIB under Grant No. R01EB009905.

^{a)}Author to whom correspondence should be addressed. Electronic mail: yy@ece.iit.edu

¹R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *Ca-Cancer J. Clin.* **65**(1), 5–29 (2015).

²M. Lanyi, *Diagnosis and Differential Diagnosis of Breast Calcifications* (Springer-Verlag, Berlin, 1988).

³K. Thangavel, M. Karnan, R. Sivakumar, and A. K. Mohideen, "Automatic detection of microcalcification in mammograms—A review," *Int. J. Graphics Vision Image Process.* **5**(5), 31–61 (2005).

⁴J. Dengler, S. Behrens, and J. F. Desaga, "Segmentation of microcalcifications in mammograms," *IEEE Trans. Med. Imaging* **12**(4), 634–642 (1993).

⁵R. N. Strickland and H. I. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms," *IEEE Trans. Med. Imaging* **15**(2), 218–229 (1996).

⁶C. H. Chen and G. G. Lee, "On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis," *Graph. Models Image Process.* **59**(5), 349–364 (1997).

⁷M. N. Gürcan, Y. Yardimci, A. E. Çetin, and R. Ansari, "Detection of microcalcifications in mammograms using higher order statistics," *IEEE Signal Process. Lett.* **4**(8), 213–216 (1997).

⁸N. Karssemeijer, "Stochastic model for automated detection of calcifications in digital mammograms," *Image Vision Comput.* **10**(6), 369–375 (1992).

⁹B. Caputo, E. Torre, S. Bouattour, and G. E. Gigante, "A new kernel method for microcalcification detection: Spin glass-Markov random fields," in *Studies in Health Technology and Informatics* (IOS Press, Amsterdam, 2002), pp. 30–34.

¹⁰P. C. de-la Higuera, J. I. Arribas, E. Muñoz-Moreno, and C. Alberola-López, "A comparative study on microcalcification detection methods with posterior probability estimation based on Gaussian mixture models," in *27th Annual International Conference of the Engineering in Medicine and Biology Society* (IEEE, Shanghai, China, 2006), pp. 49–54.

¹¹H. Jing, Y. Yang, and R. M. Nishikawa, "Detection of clustered microcalcifications using spatial point process modeling," *Phys. Med. Biol.* **56**(1), 1–17 (2011).

¹²H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided diagnosis of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Med. Phys.* **22**(10), 1555–1567 (1995).

¹³S. Yu and L. Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Trans. Med. Imaging* **19**(2), 115–126 (2000).

¹⁴A. Oliver, A. Torrent, X. Lladó, M. Tortajada, L. Tortajada, M. Sentís, J. Freixenet, and R. Zwigelaar, "Automatic microcalcification and cluster detection for digital and digitized mammograms," *Knowl.-Based Syst.* **28**, 68–75 (2012).

- ¹⁵L. Wei, Y. Yang, R. M. Nishikawa, M. N. Wernick, and A. Edwards, "Relevance vector machine for automatic detection of clustered microcalcifications," *IEEE Trans. Med. Imaging* **24**(10), 1278–1285 (2005).
- ¹⁶I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galasanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Med. Imaging* **21**(12), 1552–1563 (2002).
- ¹⁷A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani, "An SVM classifier to separate false signals from microcalcifications in digital mammograms," *Phys. Med. Biol.* **46**(6), 1651–1663 (2001).
- ¹⁸T. Ema, K. Doi, R. M. Nishikawa, Y. Jiang, and J. Papaioannou, "Image feature analysis and computer-aided diagnosis in mammography: Reduction of false-positive clustered microcalcifications using local edge-gradient analysis," *Med. Phys.* **22**(2), 161–169 (1995).
- ¹⁹W. J. H. Veldkamp and N. Karssemeijer, "Normalization of local contrast in mammograms," *IEEE Trans. Med. Imaging* **19**(7), 731–738 (2000).
- ²⁰K. J. McLoughlin, P. J. Bones, and N. Karssemeijer, "Noise equalization for detection of microcalcification clusters in direct digital mammogram images," *IEEE Trans. Med. Imaging* **23**(3), 313–320 (2004).
- ²¹H. P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. MacMahon, and P. M. Jokich, "Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography," *Med. Phys.* **14**(4), 538–548 (1987).
- ²²R. Zwiggelaar and C. R. M. Boggis, "The benefit of knowing your linear structures in mammographic images," in *Proceedings of Medical Image Understanding and Analysis*, 2002.
- ²³R. Zwiggelaar, S. M. Astley, C. R. M. Boggis, and C. J. Taylor, "Linear structures in mammographic images: Detection and classification," *IEEE Trans. Med. Imaging* **23**(9), 1077–1086 (2004).
- ²⁴S. Chen and H. Zhao, "False-positive reduction using RANSAC in mammography microcalcification detection," *Proc. SPIE* **7963**, 79631V (2011).
- ²⁵J. Wang, Y. Yang, and R. M. Nishikawa, "Reduction of false positive detection in clustered microcalcifications," in *IEEE International Conference on Image Processing* (IEEE, Brussels, Belgium, 2011), pp. 1433–1437.
- ²⁶C. Corinna and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
- ²⁷Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.* **55**(1), 119–139 (1997).
- ²⁸B. L. Bowerman, R. O'Connell, and A. Koehler, *Forecasting, Time Series, and Regression: An Application Approach* (South-Western, Cincinnati, OH, 2005).
- ²⁹D. B. Kopans, *Breast Imaging* (Lippincott Williams, New York, NY, 1998).
- ³⁰C. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: Benchmarking of state-of-art techniques," *Pattern Recognit.* **36**, 2271–2285 (2003).
- ³¹S. Lyu, "Mercer kernels for object recognition with local features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, San Diego, CA, 2005), Vol. 2, pp. 223–229.
- ³²Y. Li, S. Gong, J. Sherrah, and H. Liddell, "Support vector machine based multi-view face detection and recognition," *Image Vision Comput.* **22**, 413–427 (2004).
- ³³J. Joachims, *Text Categorization with Support Vector Machines: Learning with Many Relevant Features* (Springer, Berlin, 1998).
- ³⁴P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, Kauai, HI, 2001), Vol. 1, pp. 1–511.
- ³⁵P. Nardiello, F. Sebastiani, and A. Sperduti, "Discretizing continuous attributes in AdaBoost for text categorization," in *Advances in Information Retrieval* (Springer-Verlag, Berlin, 2003), pp. 320–334.
- ³⁶D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green, "Automatic generation of social tags for music recommendation," in *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA, 2007), pp. 385–392.
- ³⁷R. N. Dixon and C. J. Taylor, "Automated asbestos fiber counting," in *Institute of Physics Conference Series* (Institute of Physics, Bristol, 1979), Vol. 44.
- ³⁸T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, NY, 2009).
- ³⁹R. M. Nishikawa, "Current status and future directions of computer-aided diagnosis in mammography," *Comput. Med. Imaging Graphics* **31**(4), 224–235 (2007).
- ⁴⁰F. W. Samuelson and N. Petrick, "Comparing image detection algorithms using resampling," in *International Symposium on Biomedical Imaging: From Nano to Macro* (IEEE, Arlington, VA, 2006), pp. 1312–1315.