Computer-aided detection (CAD) of breast masses in mammography: Combined detection and ensemble classification

Article *in* Physics in Medicine and Biology · June 2014

DOI: 10.1088/0031-9155/59/14/3697

CITATIONS
13

READS
588

4 authors, including:

Konstantinos Plataniotis
University of Toronto
598 PUBLICATIONS **14,469** CITATIONS

SEE PROFILE

Computer-aided detection (CAD) of breast masses in mammography: combined detection and ensemble classification

2014 Phys. Med. Biol. 59 3697

(http://iopscience.iop.org/0031-9155/59/14/3697)

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine and Biology

Phys. Med. Biol. **59** (2014) 3697–3719 doi:10.1088/0031-9155/59/14/3697

Computer-aided detection (CAD) of breast masses in mammography: combined detection and ensemble classification

Jae Young Choi[1], Dae Hoe Kim[1], Konstantinos N Plataniotis[2] and Yong Man Ro[1,3]

[1] Department of Electrical Engineering, Image and Video Systems Lab, Korea Advanced Institute of Science and Technology (KAIST), 291, Daehak-ro (373-1 Guseong-dong), Yuseong-gu, Daejeon 305-701, Korea
[2] The Edward S Rogers Department of Electrical and Computer Engineering, Bell Canada Multimedia Lab, University of Toronto, 10 King's College Road, Room SFB540 Toronto, Ontario M5S 3G4, Canada

E-mail: ymro@ee.kaist.ac.kr

Received 14 January 2014, revised 27 April 2014
Accepted for publication 2 May 2014
Published 13 June 2014

Abstract

We propose a novel computer-aided detection (CAD) framework of breast masses in mammography. To increase detection sensitivity for various types of mammographic masses, we propose the combined use of different detection algorithms. In particular, we develop a region-of-interest combination mechanism that integrates detection information gained from unsupervised and supervised detection algorithms. Also, to significantly reduce the number of false-positive (FP) detections, the new ensemble classification algorithm is developed. Extensive experiments have been conducted on a benchmark mammogram database. Results show that our combined detection approach can considerably improve the detection sensitivity with a small loss of FP rate, compared to representative detection algorithms previously developed for mammographic CAD systems. The proposed ensemble classification solution also has a dramatic impact on the reduction of FP detections; as much as 70% (from 15 to 4.5 per image) at only cost of 4.6% sensitivity loss (from 90.0% to 85.4%). Moreover, our proposed CAD method performs as well or better

[3] Author to whom any correspondence should be addressed.

(70.7% and 80.0% per 1.5 and 3.5 FPs per image respectively) than the results of mammography CAD algorithms previously reported in the literature.

Keywords: computer-aided detection, combined detection, ensemble classification, mammography

(Some figures may appear in colour only in the online journal)

1. Introduction

Breast cancer is the most common form of cancer among women and is the second leading cause of death (Kopans 2007). Screening mammography is considered the most reliable and cost-effective method for the early detection of breast cancer. However, screening mammography is susceptible to a high rate of false positives (FPs) as well as false negatives, causing a high proportion of women without cancer to undergo further clinical evaluation or breast biopsy, or miss the best time interval for the treatment of cancer (Kopans 2007, Bird *et al* 1992). To cope with the aforementioned limitations, automated screening systems—so-called 'computer-aided detection (CAD)'—are being developed (Nishikawa 2007, Suri and Rangayyan 2006) to reduce the workload of radiologists and to improve the accuracy, specificity, and sensitivity in detection of breast cancer.

While many CAD techniques have been proposed so far, the development of new CAD algorithms of breast cancer is still an active research field. In addition, a few recent studies (Tang *et al* 2009, Pisano *et al* 2005, Ciatto *et al* 2007) show that the performance of the current commercial CAD systems still needs to be improved to meet the requirements of clinic and screening applications. Hence, improving the performance of CAD systems remains a challenging and open question, particularly in regard to the detection of subtle masses on mammograms. This paper aims to present a new and novel method for the automatic detection of breast masses in mammography. From the design aspect of the CAD scheme, the proposed method is innovative and quite different from previously developed CAD algorithms in the following aspects.

- Most of the mass detection algorithms in previous studies published to date (Suri and Rangayyan 2006, Tang *et al* 2009, Cheng *et al* 2005, Oliver *et al* 2010) has been limited to employing a single detection solution. However, it may be difficult for a single detection solution to provide acceptable detection because the detection algorithms have to make compromise to accommodate a wide variability in the appearance of masses, such as shapes, margins, and densities. It is reasonable to assume that different detection algorithms make their errors on different parts of types of masses (Oliver *et al* 2010). Hence, proper integration of useful detection results obtained from the separate algorithms may be desired. In this paper, we develop a novel detection scheme that takes advantage of combined use of supervised (i.e., model-based) and unsupervised detection algorithms (Oliver *et al* 2010). Thanks to using unsupervised detection, the proposed method can utilize the detection criteria generally used by radiologists (e.g., masses have usually dense core regions (Mudigonda*et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010)) for the purpose of mass detection. At the same time, by using supervised detection, the proposed method can learn the subtle characteristics of masses (e.g., textural variation of mass core regions and mass boundary regions), which are hard to be identified by radiologists. For this purpose, an effective region-of-interest (ROI) combination solution has been devised

to maximize detection sensitivity (in the prescreening stage) and at the same time to retain the number of FP detections as small as possible. This is achieved by exploiting the degree of overlapping ratio (detailed in section 4.3) in the process of ROI combination. Using our ROI combination solution, detection information gained from both supervised and unsupervised detection could become complementary for improving detection sensitivity. To the best of our knowledge, the potential improvement in performance by combining the detection results of different algorithms has not been systematically investigated in the area of mammography CAD.

- We develop a new ensemble classification framework for reducing FP detections in mammography CAD. Our method has been devised to achieve much better FP reduction performances than those obtained using a single classifier. Key technical contribution is to produce more specialized base classifiers by finding the best feature representation at every boosting round. Each base classifier trained with a best feature representation covers a smaller, specific region of the whole mass instance space. Consequently, a combination of such specialized base classifiers is likely to boost capability of discriminating mammographic masses from normal tissues, leading to considerable FP reduction. In addition, compared to previous ensemble classification approaches (Santo *et al* 2003, Constantinidis *et al* 2001, Jesneck *et al* 2006) developed for breast lesion classification, the proposed method is significant as follows: (1) our method combines data resampling based on AdaBoost learning (Freund and Schapire 1997) with the use of different feature representations, aiming to create more diverse base classifiers; (2) in order to extend our proposed ensemble classification framework to accommodate general (strong/weak) classifier—extensively used in mammographic CAD systems, we devise an effective strategy that controls the degree of weakness of base classifiers.

2. Background

Previously published techniques related to the detection (or segmentation) of ROIs can largely fall into two approaches (Tang *et al* 2009, Oliver *et al* 2010): unsupervised and supervised detection methods. Common idea of unsupervised detection algorithms (Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010, Kobatake and Murakami 1996) is to locate suspicious regions which are generally brighter and more dense than their surrounding regions in mammograms. However, it may be often hard to reliably make use of the aforementioned knowledge as detection criteria, especially when a high proportion of breast tissue is quite dense (Oliver *et al* 2010). As for supervised detection algorithms, Karssemeijer *et al* (Karssemeijer and Brake 1996) introduced the model of representing stellate pattern information to detect malignant masses on mammograms. Lai *et al* (1989) proposed pattern matching method based on a tumor-like template which defines the tumor by the following three characteristics: brightness contrast, uniform density, and circular shape. Freixenet *et al* (2008) proposed a probabilistic template matching approach to detect breast masses in mammography. Deformable templates were generated by applying the so-called eigenanalysis to breast mass instances.

To our knowledge, the methods explicitly and systematically devoted to the development of combining multiple mass detection algorithms are quite a few, although the idea of combining the results of independent CAD algorithms has been previously introduced in Oliver *et al* (2010), Wei *et al* (2006). Wei *et al* (2006) proposed a dual CAD system approach that combines two mass detection systems in parallel, one was trained with masses of average subtlety (from the current mammograms) and the other with subtle masses (from the prior mammograms).

To merge the information from the two CAD systems for classifying the masses from normal tissues, artificial neural network was used to combine the information of discriminant scores (for each detected object) resulting from the two single CAD systems. In their experimental studies, the dual CAD system achieved impressive improvement in the detection of both average and subtle masses, compared to the conventional single system approach. Oliver *et al* (2010) studied the combined use of four different detection algorithms by using simple addition and multiplication techniques of the corresponding probability images. They indicated that combination of multiple algorithms can be a potential way to improve detection, when compared to using the algorithms individually and separately, and also suggested that more sophisticated combination techniques should be explored as a new topic in CAD systems.

Popular classifiers in the area of pattern recognition have been explored for reducing FP detections in mammographic CAD systems (Tang *et al* 2009, Constantinidis *et al* 2001, Wei *et al* 1995, 1997, Sahiner *et al* 1996, Sampat 2005, Kupinski 1997), including decision tree, k-nearest-neighbors (kNN), neural network, linear discriminant analysis (LDA), and support vector machine (SVM). However, most of them have been focused on the study of the single classifier based methods. There are only a few publications on the design of ensemble classifier system on the classification of mammographic lesions (Santo *et al* 2003, Constantinidis *et al* 2001, Jesneck *et al* 2006, Fung *et al* 2006, Yoon and Kim 2008). The authors in (Constantinidis *et al* 2001) proposed the so-called the augmented behavior knowledge space method for the purpose of classification of circumscribed masses in digital mammograms. However, their study was only focused on the classification of circumscribed masses in digital mammograms. In (Fung *et al* 2006), a classifier learning algorithm based on classical bagging was suggested to handle a small fraction of outlier mammogram images which tend to produce a number of FPs. The authors in Yoon and Kim (2008) proposed the ensemble feature selection method which incorporated recently developed MSVM-RFE feature ranking into conventional AdaBoost learning scheme.

3. Overview

Figure 1 provides an overview of the proposed CAD framework (system). In the process of ROI detection and segmentation, ROIs are *independently* and *separately* extracted from both unsupervised and supervised detection algorithms. To implement unsupervised detection, contour-based detection using multi-level thresholding algorithm (Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010) was adopted for detecting and segmenting mammographic masses. We chose this unsupervised detection approach because it has been well-documented in previous publications in that it can provide 'successful' detection (segmentation) results.

Differing from the previous contour-based detection approaches (Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010), key characteristic of our approach is to integrate an effective preprocessing algorithm with multi-level thresholding technique, especially aiming to increase detection sensitivity of subtle masses on mammograms. In addition, to realize supervised detection scheme, 'wavelet model-based detection' proposed in Campanini *et al* (2004) was utilized. In the next step, detection results coming from both unsupervised and supervised detection schemes are combined via our proposed *ROI combination solution* (detailed in section 4.3). The aim of ROI combination is to maximize the detection sensitivity performance as much as possible (in the context of prescreening stage).

To significantly reduce the number of FPs, the segmented ROIs were used as input for feature extraction. Herein, we used four different feature subspaces: texture, shape, intensity, and spiculation features. Six different types of texture features were used: (1) spatial gray-level

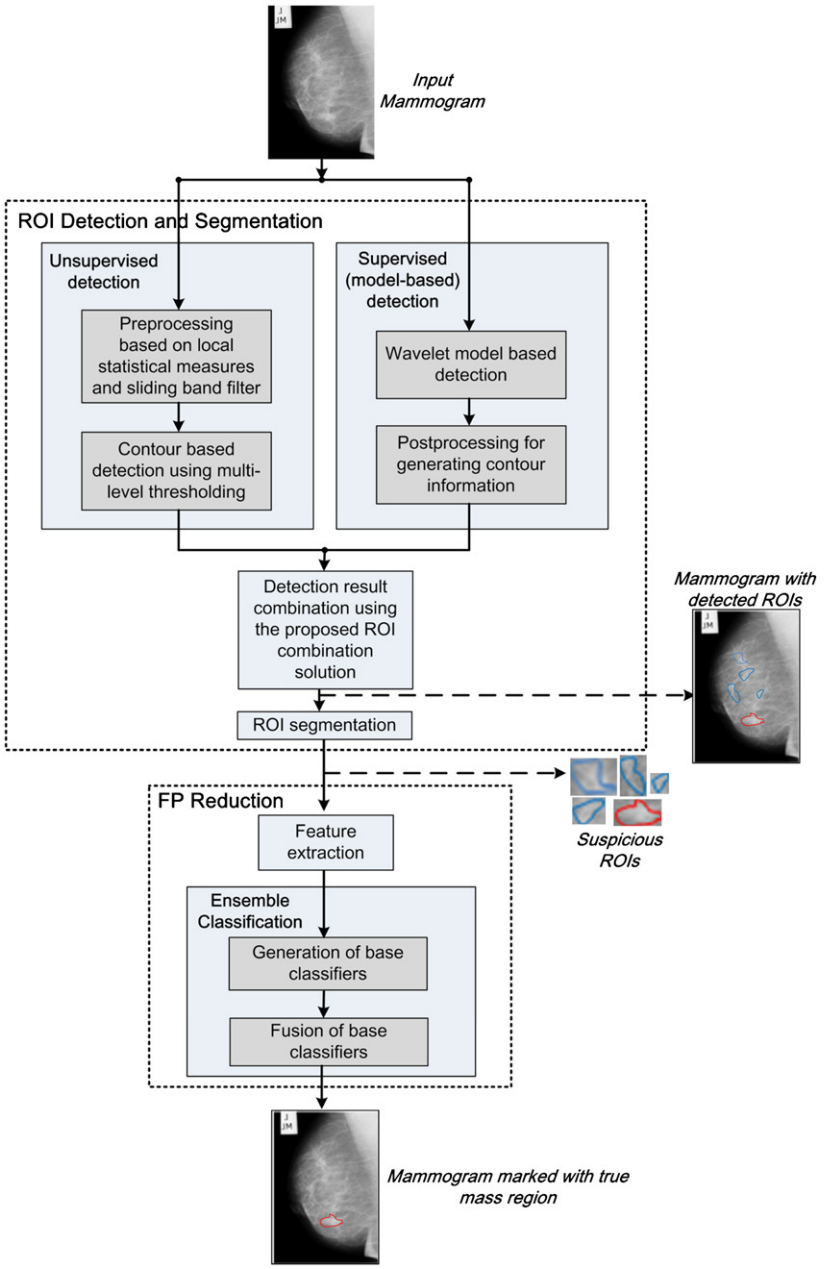


Figure 1. Overview of the proposed CAD framework of breast masses in mammography. The red line in the mammogram with detected ROIs is the successfully segmented contour (or boundary) of a true mass region, while the blue lines represent the contours of FP ROIs.

dependence (SGLD) feature (Wei *et al* 1997), (2) local binary pattern (LBP) feature (Choi and Ro 2012), (3) run length statistics feature (Sahiner *et al* 2001), (4) gray-level difference statistics feature (Sahiner *et al* 1996b), (5) rubber-band straightening transform (RBST) feature

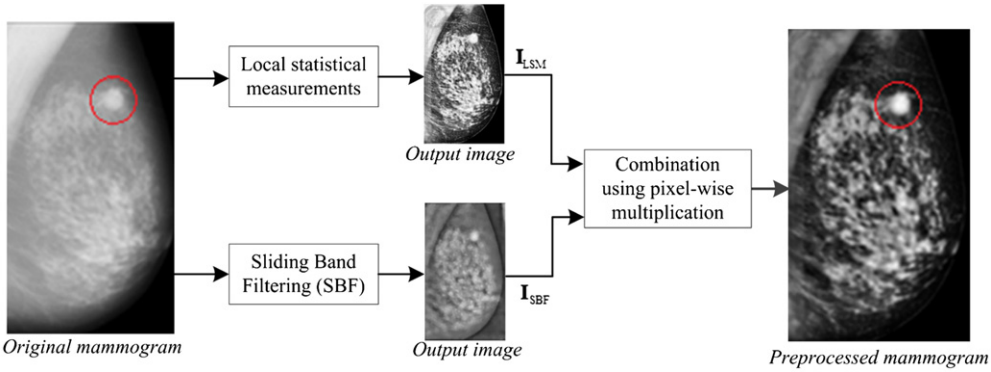


Figure 2. Block diagram of our proposed mammogram enhancement as a preprocessing step for the purpose of improving detection sensitivity of contour-based unsupervised detection. Note that the region enclosed by a red-line circle represents a true mass region.

(Sahiner *et al* 1998b), and (6) texture-flow field feature (Mudigonda *et al* 2001). As for shape features, 'Circularity', 'Extent', 'Convexity', 'Solidity', 'Eccentricity', 'Elongatedness', and 'Compactness', 'Area' were used (Cheng *et al* 2005, Sahiner *et al* 2001). To extract intensity-related features, 'Contrast', 'Average', 'Standard deviation', 'Skewness', and 'Kurtosis' were adopted (Cheng *et al* 2005). In addition, to measure the degree of spiculated patterns, region-based stellated features proposed in (Kim *et al* 2012b) were employed. Thereafter, each of the above-mentioned features was used as a particular feature representation during the generation of base classifiers in the ensemble classification. To classify a given ROI into mass versus normal tissue, the decision outputs of base classifiers included in an ensemble are combined. Each of the steps shown in figure 1 will be described in more detail in the following sections.

4. Combining unsupervised and supervised detection for ROI segmentation

4.1. Unsupervised detection

Note that the contour-based unsupervised detection has been originally designed to detect masses with hyper dense core parts. However, one challenging problem is that mammograms generally have low contrast. Hence, contour-based detection may not work well for mammograms with low contrast. To overcome this, a mammogram enhancement routine (Kim *et al* 2012a) (developed by our group) is incorporated as a preprocessing step into the unsupervised detection. Our proposed enhancement solution had been devised to combine the output images resulting from both local statistical measurements (Dominguez and Nandi 2008) and sliding band filtering (SBF) (Pereira *et al* 2007) (see figure 2 for illustration purpose).

As shown in figure 2, the combination of output images (obtained using both local statistical measurements and SBF) is performed via the pixel-wise multiplication of normalized responses in the following way (Kim *et al* 2012a):

$$\mathbf{I}_{\text{prep}} = (\mathbf{I}_{\text{LSM}})^{\alpha_1} \otimes (\mathbf{I}_{\text{SBF}})^{\alpha_2}, \tag{1}$$

where $\mathbf{I}_{\text{LSM}}$ and $\mathbf{I}_{\text{SBF}}$ are output images (as described in figure 2), each obtained using the responses of local statistical measurements and SBF respectively, and $\otimes$ denotes pixel-wise multiplication operation, and α_1 and α_2 denote weights of each output image. We set $\alpha_1 = \alpha_2 = 0.5$ to make the equal importance of $\mathbf{I}_{\text{LSM}}$ and $\mathbf{I}_{\text{SBF}}$ during the combination. Also

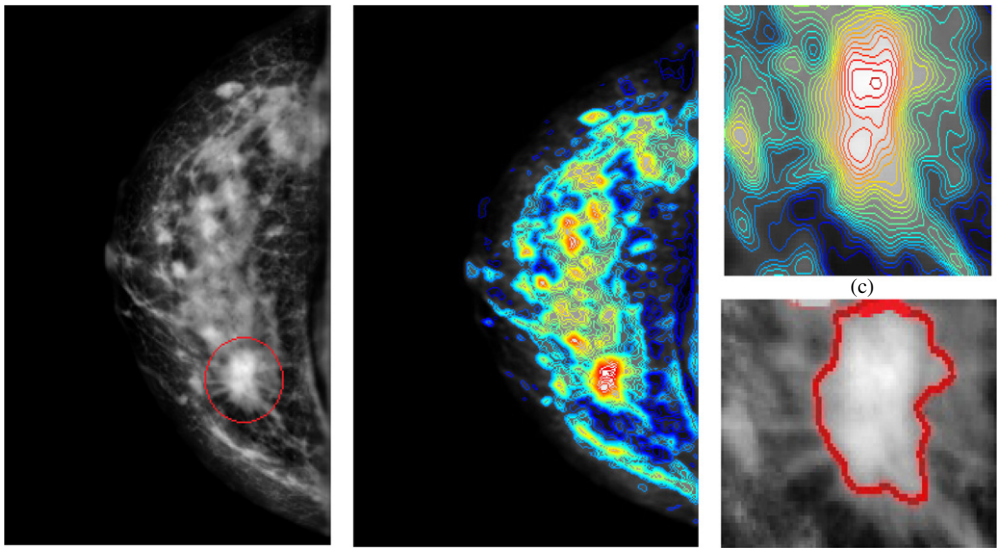


Figure 3. (a) Enhanced mammogram (generated via our preprocessing solution). Note that the region enclosed by a red line circle corresponds to true mass region. (b) Corresponding isocontour map (Hong and Sohn 2010) of the enhanced mammogram shown in (a) generated by performing multi-level thresholding. Note that closed curves marked with the same color represent a set of isocontours generated at the same given intensity level. (c) Magnification of isocontour map of ROI image of a true mass region shown in (a). (d) Successfully segmented contour (denoted by $C^{(i)}_{\text{contour}}$) of mass ROI identified by using the isocontour map shown in (c).

note that in (1), pixel-wise multiplication is used to emphasize either larger local statistical response or SBF response, or both larger local statistical and SBF responses.

In the next step, the contour-based detection is applied to a given preprocessed mammogram (see figure 3). The contour-based detection consists of the following three sequential steps: (1) construction of isocontour map, (2) formation of inclusion tree, and (3) computation of minimum nesting depth. The implementation details of each step have been described in the literature (Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010). Since we make use of an enhanced mammogram for contour-based detection, the mass detection sensitivity is expected to be increased. However, the number of FPs is likely to also increase because the preprocessing could enhance image contrast of FP regions, as well as image contrast of true mass regions. In order to reduce the number of FPs, an effective ROI selection strategy has been devised. The central idea behind our ROI selection strategy is that mass regions have relatively high pixel values of an enhanced mammogram (denoted by $\mathbf{I}_{\text{prep}}$). Based on this idea, ROIs with high pixel values are selected as the final ROIs using a threshold in the following way:

$$\text{Select } \mathbf{C}^{(i)}_{\text{contour}},\ i = 1, \ldots, M \quad \text{if} \quad \max(\mathbf{I}_{\text{prep}}(\mathbf{p}^{(i)})) > \beta \max(\mathbf{I}_{\text{prep}}), \tag{2}$$

where $\mathbf{C}^{(i)}_{\text{contour}}$ is the contour of the corresponding ith ROI (see figure 3(d)), $\mathbf{I}_{\text{prep}}$ is an enhanced mammogram, $\mathbf{p}^{(i)}$ denotes a set of pixel coordinates (x, y) of the region enclosed by the contour $\mathbf{C}^{(i)}_{\text{contour}}$, M is a total number of the detected (or segmented) regions in $\mathbf{I}_{\text{prep}}$, and $\max(\cdot)$ is a function that returns the maximum of its input argument. Note that in (2), the threshold is determined by calculating the maximum of $\mathbf{I}_{\text{prep}}$ (i.e., $\max(\mathbf{I}_{\text{prep}})$). Also, β is a parameter

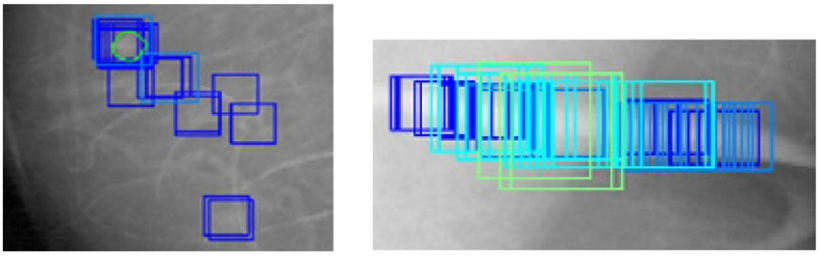


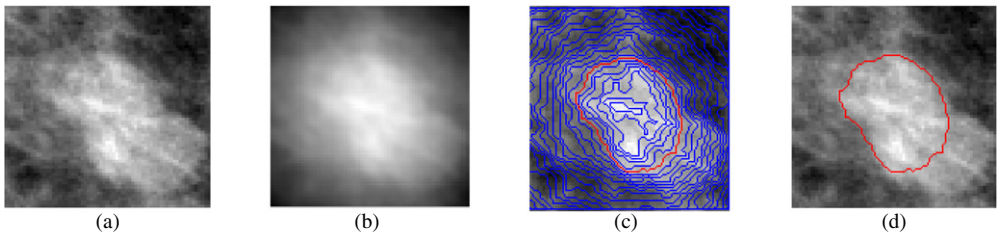
Figure 4. Illustration of overlap problem between continuous crops (scanning masks) occurring during the scanning of a mammogram together with the first SVM classifier used in supervised detection scheme. Note that the different colors represent different sizes (dimensions) of the crop boxes used for scanning.

that adjusts the value of threshold depending on the pixel value of an enhanced mammogram. It should be noted that in our experiments, a good compromise has been found by setting β in the range of (0.6, 0.8). In (2), if the maximum value of a contour $\mathbf{C}^{(i)}_{\text{contour}}$ (denoted by $\max(\mathbf{I}_{\text{prep}}(\mathbf{p}^{(i)}))$ is larger than the threshold $\beta \max(\mathbf{I}_{\text{prep}})$, we select the $\mathbf{C}^{(i)}_{\text{contour}}$ as a final ROI identified from unsupervised detection method.

4.2. Supervised detection

The supervised detection has been employed to learn the characteristics of breast masses. The supervised detection approaches generally require a set of positive training examples (masses) and a set of negative training examples (non-masses). A positive set can include mass examples with the wide diversity of their shape, margin, size, and subtlety. Hence, supervised detection may be beneficial for learning texture, directional, and structural quantities of breast masses, which are hard to be identified by using only unsupervised detection algorithms.

Our supervised detection method has been developed based on wavelet model-based detection algorithm proposed in (Campanini *et al* 2004). For implementing supervised detection method, we scan the mammogram image using a set of scanning masks with different dimension. The aim of scanning is to search all the possible locations of the image to identify areas judged as suspected masses. Scanning masks with five different sizes of 3×3, 6×6, 10×10, 16×16, and 30×30 mm were used to cover mass sizes that are commonly encountered in clinical mammograms (with size ranging from 3 mm to 30 mm, as previously reported in Freixenet *et al* (2008)). Also, as recommended by (Campanini *et al* 2004), scanning masks were resized to a prefixed size of 64×64 pixels using bilinear resizing (interpolation) technique. In next step, overcomplete wavelet analysis was performed on each resized crop, yielding the vector of wavelet coefficients; these vectors are classified as suspect or not by using the first SVM classifier to identify the candidates of breast masses. Note that in the previous wavelet-model-based detection approach (Campanini *et al* 2004), the scanning of the entire mammogram is performed by shifting the mask using a fixed step by the amount of approximately 10% of the linear dimensions of the mask, aiming to maximize detection sensitivity of masses. However, in practice, as illustrated in figure 4, this could cause a certain degree of overlap (superposition) between contiguous crops (scanning masks). To resolve this overlap problem in the previous approach, we first regard detected crops whose overlap ratio is larger than 10% as the same crop images. Then, among the same crop images, a crop with the highest confidence score (obtained from the first SVM classifier) is finally selected as a suspicious candidate ROI to be fed into a second SVM classifier. In the next step, false candidates are eliminated through using a second cascaded SVM classifier.



(a) (b) (c) (d)

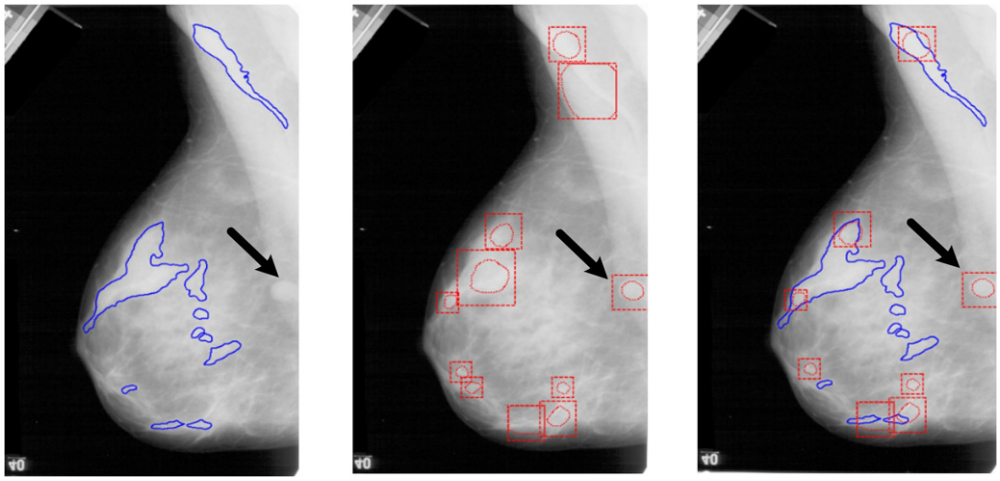
Figure 5. (a) Rectangular ROI (containing mass) detected by using the supervised detection described in section 4.2. (b) Distance pixel suppressed ROI of a given ROI in (a). (c) A series of contours created by using multi-level thresholding technique. Note that the red-colored line represents the final selected contour of mass contained in a given ROI. (d) Rectangular ROI superimposed with the final contour shown in (c).

It should be noted that the previous supervised detection algorithms have been limited to the segmentation of mass candidates only in the form of rectangular ROIs. Hence, contour information of mammographic masses cannot be used for the purpose of FP reduction. Contour information of mammographic masses is important for extracting useful mammographic features (Kim *et al* 2012b, Dominguez and Nandi 2009, Mudigonda *et al* 2000) (such as features related to shape and stellate patterns). To cope with this problem, we devise an effective *postprocessing* step for generating contour of a detected rectangular ROI. Our postprocessing solution is performed based on the radial gradient index (RGI) proposed in (Kupinski and Giger 1998). To create contour information, our method based on RGI consists of the following two sequential steps: (1) an original rectangular ROI (see figure 5(a)) is multiplied by the Gaussian constraint function (Kupinski and Giger 1998), resulting in so-called a 'distance pixel suppressed ROI' (see figure 5(b)); (2) the multi-level thresholding technique (Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010) is applied to a distant pixel suppressed ROI image for generating a series of candidate contours (see figure 5(c)); then contour that maximizes the RGI is selected as the final contour information for a given ROI. The red contour shown in figure 5(d) represents the final selected contour of a mass contained in a given ROI.

4.3. Detection result combination

To combine the detection results from the unsupervised and supervised detection schemes, a novel ROI combination solution has been developed. Note that two ROI combination approaches can be possibly used. The first solution is to make use of a logical 'or' operation, that is, a ROI region remains as suspicious for mass if it was detected by either of the two detection schemes or both. The second solution is to make use of a logical 'and' operation, that is, a region remains suspicious for mass only if it was detected by both detection schemes. The second solution may be helpful for reducing the number of FPs. However, in practice, sensitivity performance could be considerably deteriorated. On the other hand, the first solution is appropriate in terms of increasing sensitivity performance, though it might increase the number of FPs. Referring to literature (Mudigonda *et al* 2001, Wei *et al* 2006), it can be highly desirable that high sensitivity rate should be maintained, prior to performing FP reduction stage. Therefore, we adopt the first solution for combining the detection results.

To determine whether the two detection schemes identify the same region, as well as to reduce the number of FPs as many as possible, an effective ROI combination solution is used during the combination of detection results. We now explain this solution. Let us



(a) (b) (c)

Figure 6. (a) Detection results from unsupervised detection. (b) Detection results from supervised detection. (c) Combined detection results obtained using our proposed ROI combination. Note that true masses are indicated by black colored arrows in all mammograms. Also note that in supervised detection (b), the red-colored dot-line within each ROI box represents a contour (of each detected ROI) obtained using the postprocessing described in section 4.2.

define $\mathrm{ROI}_{\mathrm{unsupervised}}$ and $\mathrm{ROI}_{\mathrm{supervised}}$ as ROI sets; the former consists of M ROI contours identified from the unsupervised detection, while the latter consists of N ROI contours from the supervised detection. Also, let $\mathbf{C}^{(\mathrm{i})}_{\mathrm{unsupervised}}$ and $\mathbf{C}^{(j)}_{\mathrm{supervised}}$ be the ith and jth ROI contours provided that $\mathbf{C}^{(i)}_{\mathrm{unsupervised}} \in \mathrm{ROI}_{\mathrm{unsupervised}}$ and $\mathbf{C}^{(j)}_{\mathrm{supervised}} \in \mathrm{ROI}_{\mathrm{supervised}}$, where $i = 1, \ldots, M$ and $j = 1, \ldots, N$. The value of overlapping ratio between $\mathbf{C}^{(i)}_{\mathrm{unsupervised}}$ and $\mathbf{C}^{(j)}_{\mathrm{supervised}}$ is then calculated as follows:

$$R_{i,j} = \frac{A\left(\mathbf{C}^{(i)}_{\mathrm{unsupervised}}\right) \cap A\left(\mathbf{C}^{(j)}_{\mathrm{supervised}}\right)}{A\left(\mathbf{C}^{(i)}_{\mathrm{unsupervised}}\right) \cup A\left(\mathbf{C}^{(j)}_{\mathrm{supervised}}\right)}, \quad i = 1, ..., M, \quad j = 1, ..., N, \tag{3}$$

where $A(\cdot)$ is a function that returns a set (in $\Re^2$), the elements of which are pixel coordinates (x, y) of the region enclosed by the ROI contour as its input argument. Thus, the numerator in (3) represents the size of a region overlapped between the two ROI contours each generated by unsupervised and supervised detection, while the denominator represents the size of a region obtained by combining the two ROI contours based on the union of two sets. Therefore, the value of $R_{i,j}$ means the degree of overlap between the two detected contours under consideration.

Finally, if the value of $R_{i,j}$ is larger than a prespecified threshold value denoted by Th, the two suspicious regions (i.e., $\mathbf{C}^{(i)}_{\mathrm{unsupervised}}$ and $\mathbf{C}^{(j)}_{\mathrm{supervised}}$) are considered to be the same, otherwise they are assumed to be different regions. The two regions, considered to be the same, are then merged into the single region by selecting the $\mathbf{C}^{(i)}_{\mathrm{unsupervised}}$, while $\mathbf{C}^{(j)}_{\mathrm{supervised}}$ is deleted. Considering a good balance between detection sensitivity and specificity, a good compromise has been found by setting $Th = 0.25$ (i.e., 25% overlap between two different ROI regions) from our experiments.

Figure 6 illustrates the strength of the proposed ROI combination solution. From figure 6, we can see that a mass (indicated by black colored arrow in all mammograms) cannot be identified by the single unsupervised detection, but the missed mass can be identified via the

proposed ROI combination. In addition, as can be seen in figure 6, the number of FPs arising from our ROI combination method is moderately increased, compared to that obtained when using either unsupervised or supervised detection. This demonstrates that the advantage of our method lies in its ability to control the number of FPs on an acceptable level (in the context of prescreening stage) while it is capable of increasing sensitivity. This is accomplished by considering the degree of overlapping ratio during the ROI combination, aiming to find an optimal trade-off between sensitivity and specificity.

5. False-positive reduction using ensemble classification

The proposed ensemble classification is based on the adaptive boosting algorithm (AdaBoost) (Freund and Schapire 1997). We now explain implementation details by referring to figure 7. In Step 0 in figure 7, let $\mathbf{T}$ be a training set composed of N training instances (i.e., ROI images), each denoted by $x_i (i = 1 \ldots, N)$ with a corresponding class label ℓ_i, where $\ell_i \in \{0, 1\}$. Assuming that a total of K different feature representations of a given ROI are yielded from the feature extraction as explained in section 3, we then denote the mth feature representation by f_m (e.g., LBP or SGLD texture features described in section 3) comprising a feature pool denoted by $\mathbf{F}$ for which $f_m \in \mathbf{F}$.

In Step 1 in figure 7, the values of initial distribution $D_{t=0}(i)$ are set equally such that $D_{t=0}(i) = 1/N$. In Step 2. (2) in figure 7, the data resampling is performed to form a resampled set (denoted by $\mathbf{T}_t$)—which is a proportion of the training set such that $\mathbf{T}_t \subset \mathbf{T}$—selectively sampled from the training data set $\mathbf{T}$ according to the distribution $D_t(i)$. It is important to note that in the data resampling process, parameter r is devised to adjust the level of weakness of base classifiers (i.e., ensemble members). The value of r represents the ratio of resampled set size to whole training set size. Hence, the amount of samples in a resampled set is directly proportional to the value of r. Note that increasing the size of the training set generally leads to improved classification performance of classifier learning algorithms (Weiss and Provost 2003, Shavlik *et al* 1991, Raudys and Jain 1991). Based on this fact, it is reasonable to assume that a base classifier—trained on a smaller fraction of the training set (i.e., a smaller size $\mathbf{T}_t$)—should generalize a weaker member of an ensemble. This means that a smaller/larger r value will equivalently lead to a weak/strong (i.e., more/less accurate) base classifier, given the same classifier model. Considering overall classification accuracy, a good compromise has been found by setting r in the range of (0.4, 0.6).

In the proposed algorithm, base classifiers are produced by using different feature representations of the same input, as described in Step 2. (3) in figure 7. Using different feature representations allows finding a set of base classifiers who differ in their decisions to a certain extent complement each other. One key to successful ensemble methods is to construct base classifiers with small error rates (Kuncheva and Whitaker 2003). To account for this, in Step 2. (5) in figure 7, the best base classifier $h_t(\cdot)$ (at each boosting round t) for classifying a weighted version of $\mathbf{T}$ (i.e., weighted training samples) is determined as follows:

$$h_t = \arg\min_{h_{t,m}} \varepsilon_{h_{t,m}}, \tag{4}$$

and

$$\varepsilon_{h_{t,m}} = \sum_{i=1}^{N} D_t(i) |h_{t,m}(x_i) - \ell_i|, \tag{5}$$

where $D_t(i)$ denotes weight distribution for each training sample and $h_{t,m}(\cdot)$ is a base classifier trained with the mth feature representation f_m, where $m = 1, \ldots, K$. Note that $\varepsilon_{h_{t,m}}$ represents the weighted classification error produced by $h_{t,m}(\cdot)$. Using (4) and (5), among K individual base classifiers $h_{t,m}(\cdot)$—each trained with a particular feature representation, we select the

Training stage for generating base (component) classifiers

0. (Input)

(1) Feature pool $\mathbf{F} = \{f_m, m = 1, \ldots, K\}$

(2) Training data set $\mathbf{T}$ consisting of N labeled instances $\{(x_i, \ell_i)\}_{i=1}^{N}$ with class labels $\ell_i \in \{0,1\}$

(3) Total number of ensemble generation rounds T

1. (Initialization)

(1) Initial weight distribution $D_{t=0}(i) = 1/N$, for $i = 1, \ldots, N$ for N training samples included in $\mathbf{T}$

(2) Weight vector $w_{1,i} = D_{t=0}(i)$ for $i = 1, \ldots, N$

(3) $\mathbf{E}_{t=0} = \{\phi\}$ (Ensemble including base classifiers)

2. (Repeat for $t = 1, \ldots, T$)

(1) Compute the distribution for each training sample $D_t(i) = \dfrac{w_{t,i}}{\sum_{i=1}^{N} w_{t,i}}$

(2) Using parameter r $(0 < r < 1)$, select $(r \times 100)$ % hardest training samples per class (as the proportion of whole training samples) according to the distribution to form a resampled subset $\mathbf{T}_t$ $(\mathbf{T}_t \subset \mathbf{T})$

(3) For $m = 1, \ldots, K$

- Build a base classifier $h_{t,m}$ for each feature f_m (along with the m-th feature) using $\mathbf{T}_t$
- Calculate the weighted classification error $\varepsilon_{h_{t,m}}$ each produced by $h_{t,m}$ using $\varepsilon_{h_{t,m}} = \sum_{i=1}^{N} D_t(i) |h_{t,m}(x_i) - \ell_i|$

(4) Construct candidate classifiers denoted by $\mathbf{H}_t = \{h_{t,m}\}_{m=1}^{K}$

(5) Determine the best base classifier h_t with the lowest error ε_{h_t} from $\mathbf{H}_t$, such that $h_t = \underset{h_{t,m}}{\operatorname{argmin}} \varepsilon_{h_{t,m}}$

(6) Define the error ε_t of the best base classifier $\varepsilon_t = \varepsilon_{h_t}$

(7) If $\varepsilon_t = 0$ or $\varepsilon_t > 0.5$, ignore h_t, reinitialize the distribution $D_t(i)$ to $1/N$ and go to step 2.(2)
Else, calculate $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$ and $\mathbf{E}_t = \mathbf{E}_{t-1} \cup \{h_t\}$

(8) Update weight vector $w_{t+1,i} = w_{t,i} \beta_t^{|1 - h_t(x_i) - \ell_i|}$

3. (Output)

Generated ensemble $\mathbf{E} = \{h_t\}_{t=1}^{M}$ and corresponding confidences $\{1/\beta_t\}_{t=1}^{M}$, where $M \leq T$

Testing stage for classification

Classify each testing sample using weighted combination rule

$h_{\text{combined}} = \sum_{t=1}^{M} \alpha_t h_t(x)$ where $\alpha_t = \dfrac{1/\beta_t}{\sum_{t=1}^{P} 1/\beta_t}$

Figure 7. Our proposed ensemble classification algorithm. Note that, as recommended by Kuncheva (2004), if a best base classifier has an error rate greater than $1/2$ in a trial (at each boosting round), then we reinitialize the training set weights to the uniform distribution and continue drawing samples.

best base classifier $h_t(\cdot)$ that yields the most accurate results on a weighted training set. In Step 2. (8), to maintain a set of weights over the T, distribution $D_t(i)$ for each training sample x_i can be determined based on the classification error of a best base classifier; on each round, the values of distribution progress toward increasing the likelihood of difficult samples. This forces a best base classifier generated at next round to focus on the hard-to-classify training samples.

After terminating the generation of base classifiers, to perform classification, the outputs of M generated base classifiers (denoted by $\mathbf{E} = \{h_t\}_{t=1}^{M}$) within an ensemble are combined. For this, the weighted combination rule (Ranawana and Palade 2009) is used in the following way:

$$h_{\text{combined}} = \sum_{t=1}^{M} \alpha_t h_t(x), \tag{6}$$

where $\alpha_t = \frac{1/\beta_t}{\sum_{t=1}^{p} 1/\beta_t}$ and $\beta_t = \varepsilon_t/(1-\varepsilon_t)$, and ε_t is the corresponding classification error associated with the h_t (please refer to (4)). Note that since $1/\beta_t$ is monotonically increasing as ε_t becomes smaller, $1/\beta_t$ would be a reliable indicator of representing the confidence of decision outputted by h_t (Freund and Schapire 1997).

6. Results and discussion

6.1. Evaluation study setup

All mammograms in this study were randomly selected from public Digital Database for Screening Mammography (DDSM) (Heath *et al* 2000). We collected three datasets. The first dataset contained 606 mammograms (each with one mass) from 386 patients. Of the 606 masses, 303 of the masses were found to be malignant and 303 benign. This dataset is referred to as the 'mass dataset 1'. All images in the mass dataset 1 were digitized with a Howtek scanner with 43.5 μm pixel size. The second dataset was composed of 520 mammograms (containing 240 benign and 280 malignant masses) scanned with a Lumisys scanner at a resolution 50 μm per pixel at a bit depth of 12. We refer to these 520 mammograms as the 'mass dataset 2'. The third dataset consisted of 320 normal mammograms (that are free of masses) digitized with a Howtek scanner. The mammograms are referred to as the 'normal dataset'. The normal dataset was only used for assessing the number of FPs generated by CAD algorithms on normal mammograms. Note that although the mammograms in the study datasets were randomly selected, the masses found in clinical practice were well represented in our collected dataset by containing a broad variety of mass shapes, margin characteristics, and breast densities.

The DDSM provides annotations of the masses presented in each image. These annotations were considered as the ground truth in our experiments. Using ground truth information, a generated ROI was considered as a true mass ROI only if it met the following two criteria: (1) the centroid of a segmented region is included in the DDSM annotated area (Kallergi *et al* 1999), and (2) a segmented region intersects with the true mass region more than 40% (Kallergi *et al* 1999). In order to evaluate the performance of the proposed ROI detection (segmentation) method (described in section 4), free receiver operating characteristics (FROC) analysis (Suri and Rangayyan 2006) were used. To assess the FP reduction performance, receiver operating characteristic (ROC) curves were used. In particular, the area under the ROC curve (AUC) (Hanley and McNeil 1983) was used for evaluating the overall FP reduction performance. Besides, the normalized partial area index ($_{\text{p}}$AUC) (Sahiner *et al* 1998a) was used to test FP reduction performance at high sensitivity levels, where p indicates the lowest acceptable sensitivity level. In our experiments, a sensitivity level (threshold) of 90% was used for the computation of partial area index (Sahiner *et al* 1998a), denoted as '$_{0.9}$AUC'. To estimate the statistical significance of the difference between two different FROC curves, the jackknife FROC (i.e., JAFROC) method (Chakraborty and Berbaum 2004) has been adopted. In particular, a well-known Chakraborty's JAFROC software[4] was used. Furthermore, to test the statistical significance of the difference between AUC values, a popular statistical test developed by Hanley *et al* Hanley and McNeil (1983) was used.

[4] www.devchakraborty.com.

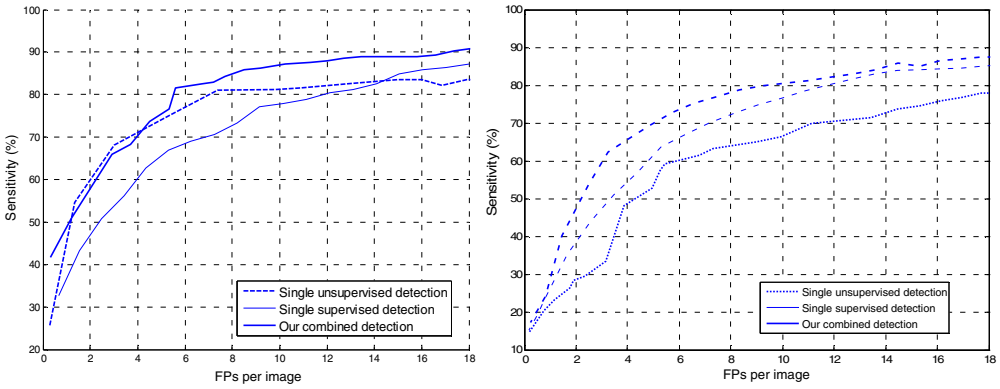


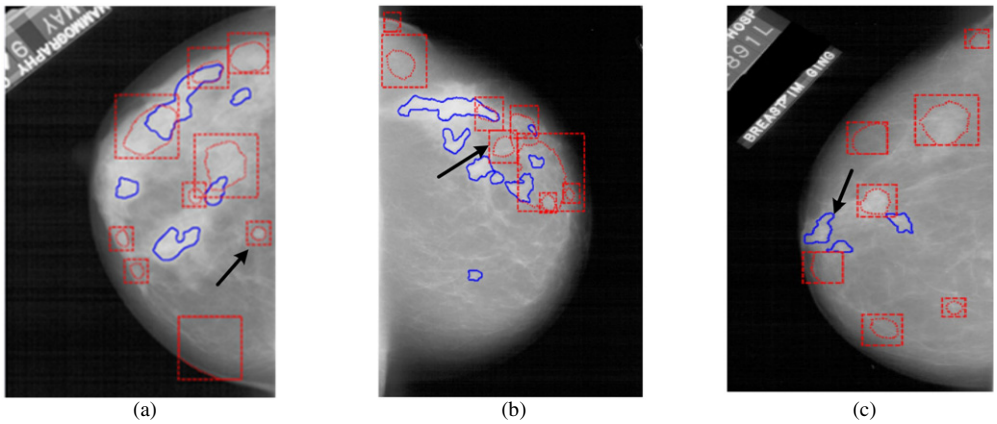
Figure 8. Comparison of the test FROC curves of our combined detection with that of the single unsupervised detection and that of the single supervised detection. (a) Mass dataset 1. (b) Mass dataset 2. Note that the number of FPs per image was estimated using (a) 303 testing mammograms with masses in the mass dataset 1 and (b) 260 testing mammograms with masses in the mass dataset 2.

6.2. Assessing our combined mass detection

To test our combined mass detection method, the mass dataset 1 was randomly divided into two disjoint subsets: 303 training and 303 testing mammogram sets, while 260 training and 260 testing mammogram sets for the mass dataset 2. To guarantee the stability of evaluation results, ten runs of random partitions were executed and all of the FROC results, reported in this section, were averaged over ten runs.

In our proposed ROI detection, each single detection scheme has its own tuning parameter devised for finding the proper balance between the sensitivity and the number of FPs. Specifically, for unsupervised detection, threshold β (shown in (2)) is used as a primarily tuning parameter that controls the sensitivity and the number of FPs. On the other hand, for supervised (model-based) detection, the number of selected ROIs (that are high-ranked as most suspicious regions after the second SVM) is used as a main tuning parameter. The proposed ROI detection has, therefore, two main tuning parameters, each corresponding to the supervised and unsupervised detection. For the sake of stable and fair comparison, the proposed method has been evaluated with different combinations of these two tuning parameters. In the case of supervised detection, we selected 20 operating points (by adjusting the number of ROIs based on their ranking values), while for the case of unsupervised detection, we selected six operating points by setting the value of β to 0.6, 0.7, 0.8, 0.85, 0.9, and 0.95, respectively. For plotting the FROC curves of our method, we first fixed β (as an operating point of unsupervised detection) by choosing one of the aforementioned six different values, and subsequently changing the 20 operating points of supervised detection, yielding a total of six FROC curves (one for a specific β).

In figure 8, the FROC curve of our combined detection method was compared to those of the single detection schemes. Note that to guarantee the stability of comparison against the variation of tuning parameters, reported FROC curve for our method was averaged over all six FROC curves described above. For mass dataset 1, when using only a single supervised detection, we achieved about 67.48%, 77.99%, and 84.82% sensitivity with the FP rate of 5.8, 10.8, and 15.6 per image, respectively, while for a single unsupervised detection, we attained about 76.30%, 81.55%, and 83.56% sensitivity with the FP rate of 5.8, 10.8, and 15.6 per



(a) (b) (c)

Figure 9. Examples of mammograms for validating the benefit of our combined detection method in terms of sensitivity improvement. Note that in each mammogram, the blue-colored solid line is the ROI identified by unsupervised detection, while the red-colored dot line is the ROI marked by supervised detection. True masses are indicated by black-colored arrows in all mammograms. (a) A malignant mass with subtlety rating '2' (ranging from 1 (subtle) to 5 (obvious)). (b) A malignant mass with subtlety rating '2'. (c) A benign mass with subtlety rating '4'.

image, respectively. On the other hand, at the same FP rate, our combined detection reached about 81.82%, 87.23%, and 89.19% sensitivity. Hence, compared to the single unsupervised detection, the detection sensitivity was increased to about 5.52%, 5.68%, and 5.63% with the same FP of 5.8, 10.8, and 15.6 per image, respectively, thanks to using the proposed method. For mass dataset 2, when comparing to the supervised detection, improvement in detection sensitivity, 11.45%, 6.78%, and 4.04%, can be achieved via our combined detection at the FP rate of 3.2, 7.4, and 10.2 per image, respectively, while 28.87%, 13.79%, and 14.42% improvement at the same FP rate in comparison with the unsupervised detection. The difference between the FROC curves (i.e., the proposed method versus unsupervised detection or the proposed method versus supervised detection) is found to be statistically significant by observing $p < 0.001$ through using JAFROC analysis. Note that in ROI detection stage (usually referred to as prescreening stage), emphasis is placed more on the achievement of higher sensitivity than on the achievement of lower FP rate (Oliver *et al* 2010, Mudigonda *et al* 2001). In this context, the ability of our combined detection to increase sensitivity may be beneficial for enhancing overall CAD performance.

Figure 9 shows example mammograms to demonstrate the benefit of our combined detection method in terms of improving detection sensitivity. As shown in figure 9, the masses that were missed by either the single unsupervised or supervised detection can be identified by our combined detection method. Especially, as shown in figures 9(a) and (b), unsupervised detection fails to detect *subtle masses* that do not have typical patterns of masses, such as hyper dense core and spiculated margins. However, our combined detection method is able to identify these subtle masses by taking advantage of detection results of supervised detection. This may reflect a complementary effect by combining multiple detection algorithms, leading to accommodate a wide range of lesion characteristics for improving detection sensitivity.

Referring to literature (Oliver *et al* 2010, Freixenet *et al* 2008), the accuracy of the detection algorithms is reported to be strongly influenced by both mass shape and margin. In this context, it is worth assessing our combined detection method with regard to mass shape and margin. Figure 10 shows the average sensitivity of three different detection algorithms

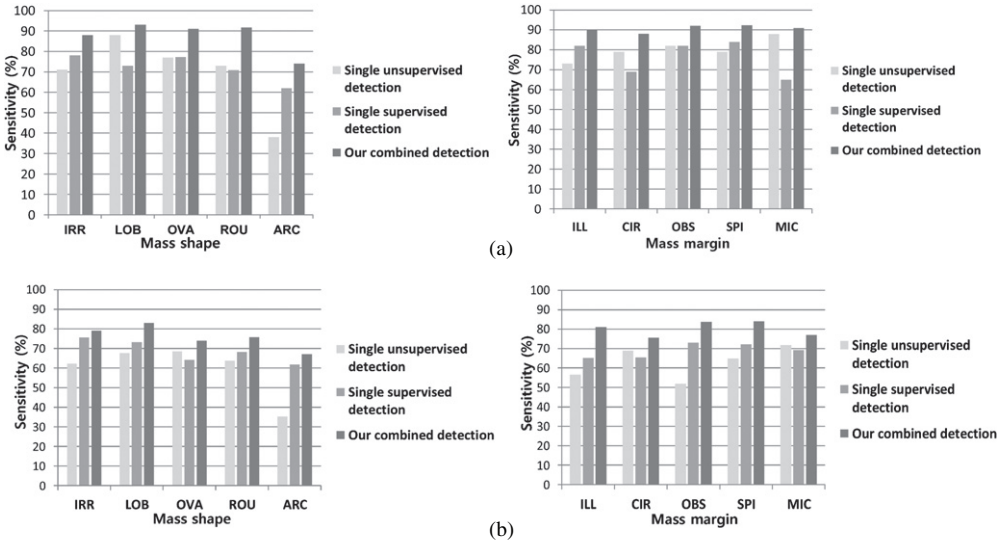


Figure 10. Comparison of the sensitivity of three different detection approaches with respect to the types of mass shape (or margin). Note that for each shape or margin type, the average sensitivity value was calculated based on the three different operating points of 8, 10, and 12 FPs per image. IRR: irregular, LOB: lobulated, OVA: oval, ROU: round, ARC: architectural distortion, ILL: ill defined, CIR: circumscribed, SPI: spiculated, OBS: obscured, MIC: microlobulated. (a) The mass dataset 1. (b) The mass dataset 2.

(note: the evaluation has been based on testing mammograms). For each shape or margin type, the average sensitivity value was calculated based on the three sensibility results (for the sake of stable comparison), corresponding to three different operating points that retained a number of 8, 10, and 12 FPs per image. Looking at the results obtained for mass dataset 1 (figure 10(a)) and mass dataset 2 (figure 10(b)), common observations can be made as follows. Our combined detection performs better than both the single unsupervised and supervised detection for all types of mass shape and margin. In particular, compared to the single unsupervised detection, the improvement in sensitivity (achieved by using our method) is significant when detecting *subtle masses* (not definitely visible) with architecture or irregular shapes, and/or ill-defined margins. This is due to the fact that subtle masses have generally poorly defined shapes and margins (borders). Hence, using only an unsupervised approach (that performs detection using clinical knowledge-based reasoning on the basis of mass shape and margin) may not work well for identifying these subtle masses. However, supervised detection can learn the characteristics (e.g., texture patterns) of subtle masses for a detection purpose. For this reason, the appropriate use of detection results from supervised detection enables our combined method to perform well on the detection of subtle masses.

6.3. Ensemble classification for the reduction of FPs

To demonstrate the effectiveness of our ensemble classification, a ROI dataset was organized by applying our proposed ROI detection to the mass dataset 1 (606 mammograms). Note that during the generation of ROIs, we chose operating threshold which led to an average number of about 15 FPs per image at a detection sensitivity of about 90%. With this fixed operating threshold, FP reduction analysis was performed on all classification approaches used in our

experiments. As a result, a total of 9958 ROIs were automatically generated: 546 mass and 9412 normal tissue ROIs. To achieve the least-biased estimates of classification performance, our evaluation was carried out using the most widely used 5 × 2-Fold cross-validation (cv) (Kuncheva 2004) for testing classifier models. 5 × 2-Fold cv consists of repeating a 2-fold cv procedure five times. In each cv run, we divided the dataset into training and testing halves. The roles are swapped at each fold to generate ten training and testing sets, yielding the final classification performance computed by averaging ten corresponding results. In the proposed ensemble classification (see figure 9), SVM (Vapnik 1998), LDA (Kuncheva 2004), and decision tree (Kuncheva 2004) were used for base classifiers. Specifically, a SVM classifier, which utilizes a radial basis function (as kernel), was constructed. To implement tree base classifiers, 'Gini impurity' (Kuncheva 2004) was used as criterion for choosing a split.

Note that the features described in section 3 were used as different feature representations (i.e., f_m defined in figure 7) for generating base classifiers. We compared the performance of our ensemble classifier against the single SVM, LDA, and decision tree classifier approaches. Note that the single classifier was trained using the same training set, the same as that used to perform our ensemble generation. To guarantee that variation in performance is due only to the difference in the *classifier design approaches* (rather than used classifier models), the parameters of the single classifiers needs to be optimized, especially in order to prevent *over-fitting*. To this end, we adopted a 'grid-search' on the associated parameters of kernel function and the regularization parameter using cross-validation (Chang and Lin 2011) to achieve optimal *generalization* performance. Also, cross-validation based approach (Ji and Ye 2008) was used to find a suitable regularization parameter of the single LDA classifier. Moreover, we applied error-based pruning (Kuncheva 2004) to the constructed full decision tree for finding the optimal sequence of pruned subtree. For comparative purposes, the single classifiers with all the available features were also implemented. To this end, we adopted 'feature-level' fusion strategy (Jain *et al* 2005). In addition, the single classifier with feature selection algorithm was compared with our classifier ensemble. For this, we have implemented a well-established and popular feature selection algorithm for mammographic mass classification, so-called the stepwise feature selection (SFS) proposed in Way *et al* (2010).

The comparative results are given in figure 11. From figure 11, we can see that the results of the proposed ensemble classification are much better than those obtained using the single classifier approach using all available features for all types of base classifiers considered. Especially, a statistical comparison on the values of ${}_{0.9}$AUC between the proposed method and the single classifier approach using all features yields $p < 0.0001$, $p = 0.0019$, and $p < 0.0001$ for SVM, LDA, and decision tree, respectively. Moreover, as seen in figure 11, the proposed method outperforms the single classifier with feature selection in terms of AUC and ${}_{0.9}$AUC for all base classifiers. In particular, considering a statistical comparison of ${}_{0.9}$AUC values, the results of $p < 0.0001$, $p < 0.0001$, and $p < 0.0001$ were observed for SVM, LDA, and decision tree, respectively. The results shown in figure 11 clearly demonstrate that the proposed ensemble classifier can be much more effective than the single classifier based approaches, in terms of designing clinically relevant CAD systems that achieve high specificity performance at high sensitivity.

6.4. Evaluating overall CAD performance

In this section, we present overall CAD performance (detection sensitivity versus number of FPs) of a full CAD system with our combined detection together with ensemble classification. To this end, both mass dataset 1 (606 mammograms) and mass dataset 2 (520 mammograms) were used. Note that used mammogram dataset (i.e., mass dataset 1 or mass dataset 2) was

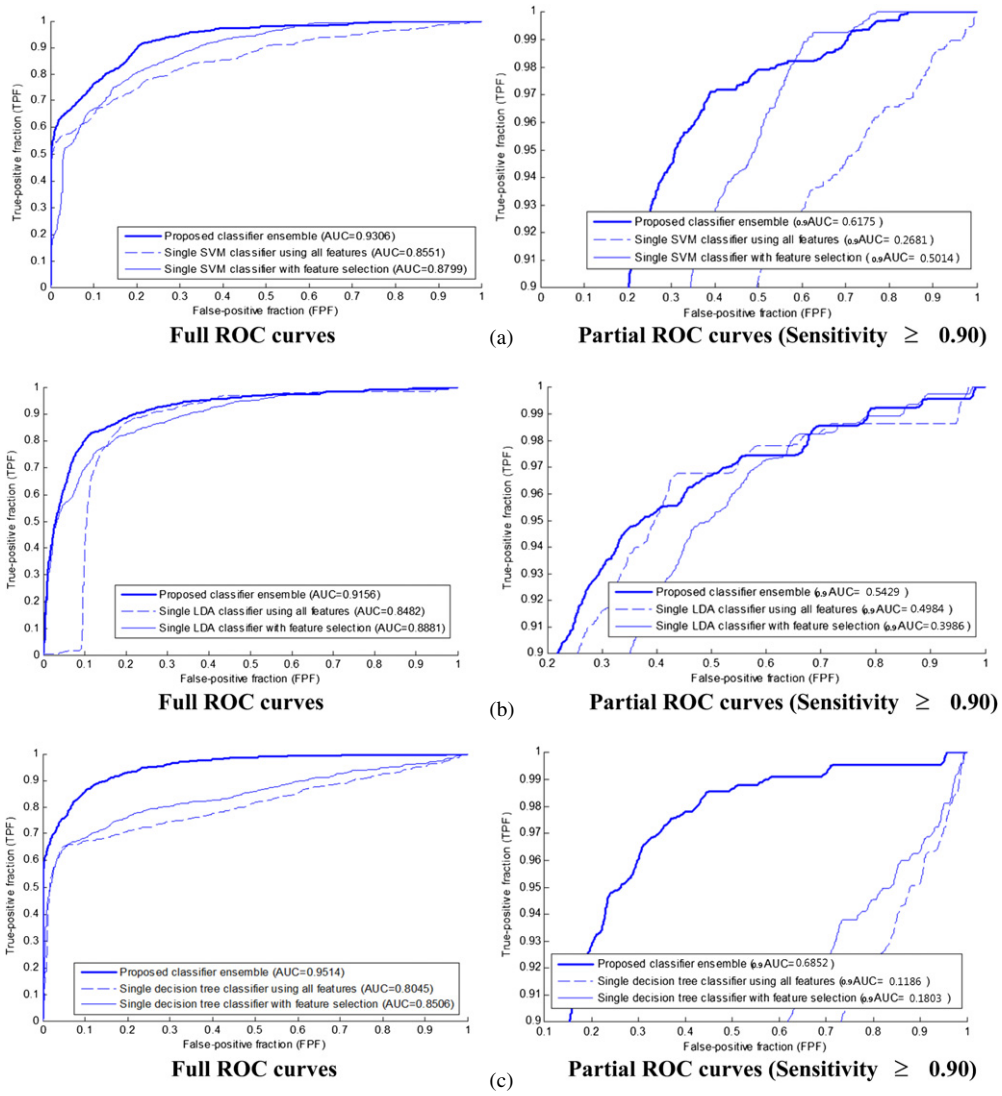


Figure 11. Comparisons of ROC curves to demonstrate the effectiveness of our proposed classification ensemble. The graphs on the left side show the full ROC curves, while the graphs on the right side show only the high-sensitivity region (sensitivity $\geqslant 0.90$). During the implementation of our ensemble classifier, the parameter 'r' was set to 0.5, 0.4, 0.6 for SVM, LDA, and decision tree, respectively, and the number of boosting rounds T was set to 50, 100, and 150 for SVM, LDA, and decision tree. (a) SVM base classifiers. (b) LDA base classifiers. (c) Decision tree base classifiers.

randomly divided into two disjoint subsets: training and testing sets (using the same manner as descried in section 6.1). Note that only training mammograms were used for the purpose of training our combined detection (i.e., supervised detection algorithm) and ensemble classifier, while testing set was used to measure detection sensitivity performance. To evaluate the FP rate, the normal dataset (described in section 6.1) was used. We compared the performance of the proposed CAD system with the following two other CAD systems: (1) CAD system using only the unsupervised detection and a single classifier (referred to as 'single CAD system 1'),

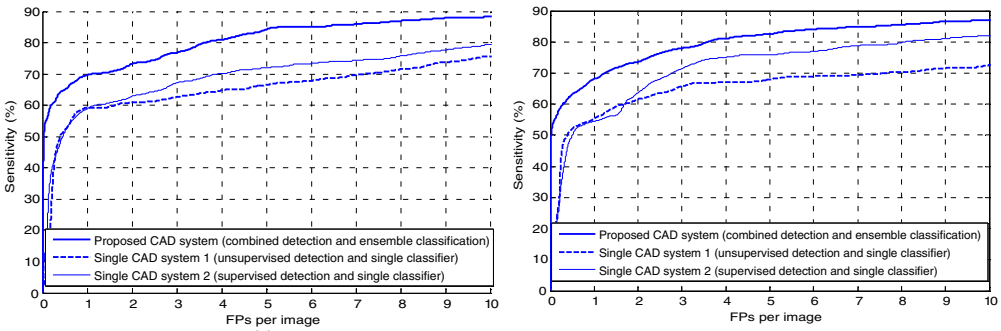


Figure 12. Comparison of the FROC curves to evaluate the overall CAD performance. Decision tree classifier was used to create base classifiers in the proposed CAD system, as well as to construct a single classifier in the other two single CAD systems. (a) The mass dataset 1. (b) The mass dataset 2. Note that a given sensitivity, the average number of FPs per image was computed on the normal data set (520 normal mammograms) in both (a) and (b).

and (2) CAD system using only the supervised detection and a single classifier (referred to as 'single CAD system 2'). Note that for reliable comparison, all features (described in section 3) together with the SFS feature selection (Way *et al* 2010) were used to construct a best-possible single classifier.

The comparison of the FROC curves obtained for both mass dataset 1 and mass dataset 2 is given in figures 12(a) and (b), respectively. For the mass dataset 1 (figure 12(a)), our proposed CAD scheme can achieve a sensitivity of 70.7%, 74.5%, and 82.6% at 1.5, 2.5, and 4.5 FPs per image. On the other hand, the sensitivity of 'single CAD system 1' is 59.4%, 61.3%, and 65.1% at the same FP rate, respectively, while that obtained for 'single CAD system 2' is 60.8%, 64.3%, and 71.4%. For the mass dataset 2 (figure 12(b)), using the proposed CAD system, at 1.5, 2.5, and 4.5 FPs per image (obtained for normal data set), the sensitivity of 71.7%, 76.3%, and 81.9% was achieved, respectively. Note that the improvement in the FROC curves by the proposed CAD system is statistically significant ($p < 0.05$) using the JAFROC method.

Previous reported results obtained using the DDSM DB are explained as follows. The authors in Catarious *et al* (2004) reported a sensitivity of 80% at 4.5 FPs per image for the testing set. In another mass detection study (Catarious *et al* 2006), which used images from the DDSM, the detection accuracy has been reported as 72% at 2 FPs per image for overall mass detection performance (including the malignant and benign masses). The authors in Brake *et al* (2000) reported 81% sensitivity with FPs (per image) ranging from 1.35 to 3.66 FPs for malignant mass detection performance. Note that the aforementioned discussion on the comparison of results reported in previous studies aims at providing context for demonstrating the feasibility of our proposed CAD method, rather than a direct comparison with our results.

It should be stressed that determining overall CAD performance is possible only if comparison of different CAD algorithms is performed using the same set of images, and the same evaluation criteria and protocol (Heath and Bowyer 2000, Nishikawa *et al* 1994). In this context, we tested our CAD method on public DoD BCRP Mammography Datasets (Heath and Bowyer 2000). The central objective for developing these datasets is to provide a common data set of mammograms that can be used to make comparisons of CAD algorithms. We used datasets named 'BCRP_MASS_0' and 'BCRP_MASS_1'. Each case from these two

datasets has at least one, malignant, spiculated mass per case. For details on the aforementioned datasets, we refer the reader to the literature (Heath and Bowyer 2000). As recommended by the guidance in Heath and Bowyer (2000), 'BCRP_MASS_0' (all 156 mammograms and 70 masses) is used for training dataset during the implementation of our CAD algorithm, while 'BCRP_MASS1' for testing set (all 160 mammograms and 81 masses). It is important to note that there is no overlap between the training cases obtained from the 'BCRP_MASS_0' and the testing cases from the 'BCRP_MASS_1'. It should be also noted that we strictly followed the same evaluation criteria (e.g., the way of computing TP and FP rate) recommended in Heath and Bowyer (2000) to allow a direct, fair comparison of the different CAD methods. For comparison purpose, previously reported results (obtained using 'BCRP_MASS_0' and 'BCRP_MASS1') (Heath and Bowyer 2000) were considered to be a baseline performance. According to the experimental results, our CAD method attained about 67%, 79%, and 87% sensitivity of 1.0, 3.0, and 5.0 FPs per image. On the other hand, for the case of baseline performance (obtained using AFUM algorithm developed by Health *et al*. (Heath and Bowyer 2000)), about 60%, 74%, and 85% sensitivity at the same FP rate has been reported. The performance of our CAD algorithm is quite comparable to, or better than, previously reported performance, which further confirms the effectiveness of our proposed CAD algorithm for correctly detecting masses.

7. Conclusions

We have developed a novel CAD algorithm for detecting mammographic masses. Differing from the existing CAD design approaches, we proposed the combined use of two representative detection schemes (i.e., unsupervised and supervised detection) to improve detection sensitivity. In addition, new ensemble classification algorithm has been devised for significantly reducing the number of FP detections. Comparative results validate the usefulness of our combined detection method for improving detection sensitivity compared to the single detection approaches recently proposed in the literature. Improvement in detection sensitivity thanks to using our combined detection is also found to be fairly robust across the mass shape and margin types. Our ensemble classification allows for dramatic reduction of FP detections (e.g., from 4545 down to 697 FPs, or 84.6% reduction) while maintain the high detection sensitivity (from 272 down to 243 benign or malignant masses detected, around 10% sensitivity loss). In conclusion, the proposed CAD algorithm has high potential for improved detection and identification of mammographic lesions in clinical screening process. Our proposed CAD algorithm would be limited by using a single mammographic image. Hence, future work includes extending our work by applying our combined detection and ensemble classification to CAD design framework using two mammographic view images (CC and MLO of the same breast) described in Hupse *et al* (2012), Samulski and Karssemeijer (2011). This extension will enable our current CAD algorithm to move forward a more complete CAD system to fully meet the requirements for routine clinical examinations.

References

Bird R E, Wallace T W and Yankaskas B C 1992 Analysis of cancers missed at screening mammography *Radiology* **184** 613–7

Brake G M, Karssemeijer N and Hendriks J H 2000 An automatic method to discriminate malignant masses from normal tissue in digital mammograms *Phys. Med. Biol.* **45** 2843–57

Campanini R, Dongiovanni D, Iampieri E, Lanconelli N, Masotti M, Palermo G, Riccardi A and Roffilli M 2004 A novel featureless approach to mass detection in digital mammograms based on support vector machines *Phys. Med. Biol.* **49** 961–75

Catarious D M, Baydush A H and Floyd C E 2004 Incorporation of an iterative, linear segmentation routine into a mammographic mass CAD system *Med. Phys.* **31** 1512–20

Catarious D M Jr, Baydush A H and Floyd C E 2006 Characterization of difference of Gaussian filters in the detection of mammographic regions *Med. Phys.* **33** 4104–14

Chakraborty D P and Berbaum K S 2004 Observer studies involving detection and localization: modeling, analysis, and validation *Med. Phys.* **31** 2313–30

Chang C-C and Lin C-J 2011 LIBSVM: a library for support vector machines *ACM Trans. Intell. Syst. Technol.* **2** 27

Cheng H D, Shi X J, Min R, Hu L M, Chi X P and Du H N 2005 Approaches for automated detection and classification of masses in mammograms *Pattern Recognit.* **39** 646–68

Choi J Y and Ro Y M 2012 Multiresolution local binary pattern texture analysis combined with variable selection for application to false positive reduction in computer-aided detection of breast masses on mammograms *Phys. Med. Biol.* **57** 7029–52

Ciatto S *et al* 2007 Computer-aided screening mammography *New Engl. J. Med.* **357** 83–85

Constantinidis A S, Fairhurst M C and Rahman F R 2001 A new multi-expert decision combination algorithms and its application to the detection of circumscribed masses in digital mammograms *Pattern Recognit.* **34** 1527–37

Dominguez A R and Nandi A K 2008 Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection *Comput. Med. Imaging Graph.* **32** 304–15

Dominguez A R and Nandi A K 2009 Toward breast cancer diagnosis based on automated segmentation of masses in mammograms *Pattern Recognit.* **42** 1138–48

Freixenet J, Oliver A, Martií R, Llado X, Pont J, Perez E, Denton E R E and Zwiggelaar R 2008 Eigendetection of masses considering false positive reduction and breast density information *Med. Phys.* **35** 1840–53

Freund Y and Schapire R E 1997 A decision-theoretic generalization of on-line learning and an application to boosting *J. Comput. Syst. Sci.* **55** 119–39

Fung G, Krishnapuram B, Merlet N, Ratner E, Bamberger P, Stoeckel J and Rao R B 2006 Addressing image variability while learning classifiers for detecting clusters of micro-calcifications *Int. Conf. on Digital Mammography* pp 84–91

Hanley J A and McNeil B J 1983 A method of comparing the areas under receiver operating characteristic curves derived from the same cases *Radiology* **148** 839–43

Heath M, Bowyer K, Kopans D, Moore R and Kegelmeyer P J 2000 The digital database for screening mammography *Proc. 5th Int. Conf. on Digital Mammography* pp 212–8

Heath M D and Bowyer K W 2000 Mass detection by relative image intensity *5th Int. Workshop on Digital Mammography (Canada)* pp 219–25

Hong B W and Sohn B S 2010 Segmentation of regions of interest in mammograms in a topographic approach *IEEE Trans. Inform. Technol. Biomed.* **14** 129–39

Hupse R, Samulski M, Lobbes M, Heeten A D, Imhof-Tas M W, Beijerinck D, Pijnappel R, Boetes C and Karssemeijer N 2012 Standalone computer-aided detection compared to radiologists' performance for the detection of mammographic masses *Eur. Radiol.* **23** 93–100

Jain A, Nandakumar K and Ross A 2005 Score normalization in multimodal biometric systems *Pattern Recognit.* **38** 2270–85

Jesneck J L, Nolte L W, Baker J A, Floyd C E and Lo J Y 2006 Optimized approach to decision fusion of heterogeneous data for breast cancer diagnosis *Med. Phys.* **33** 2945–54

Ji S and Ye J 2008 Generalized linear discrminant analysis: a unified framework and efficient model selection *IEEE Trans. Neural Netw.* **19** 1768–82

Kallergi M, Carney G M and Gaviria J 1999 Evaluating the performance of detection algorithm in digital mammography *Med. Phys.* **26** 267–75

Karssemeijer N and Brake G M 1996 Detection of stellate distortions in mammograms *IEEE Trans. Med. Imaging* **15** 611–9

Kim D H, Choi J Y, Choi S H and Ro Y M 2012a Mammographic enhancement with combining local statistical measures and sliding band filter for improved mass segmentation in mammograms *SPIE Medical Imaging (USA)* pp 83151Z-1–6

Kim D H, Choi J Y and Ro Y M 2012b Region based stellate features for classification of mammographic spiculated lesions in computer-aided detection *ICIP: IEEE Int. Conf. on Image Processing* pp 2821–4

Kobatake H and Murakami M 1996 Adaptive filter to detect rounded convex regions: Iris filter *IEEE Int. Conf. on Pattern Recognition* pp 340–4

Kopans D B 2007 *Breast Imaging* 3rd edn (Philadelphia, PA: Williams & Wilkins)
Kuncheva L I 2004 *Combining Pattern Classifiers: Methods and Algorithms* (New York: Wiley)
Kuncheva L I and Whitaker C J 2003 Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy *Mach. Learn.* **51** 181–207
Kupinski M A 1997 Investigation of regularized neural networks for the computerized detection of mass lesions in digital mammograms *EMBS: IEEE Int. Conf. on Engineering in Medicine and Biology Society* vol 3 pp 1336–9
Kupinski M A and Giger M L 1998 Automated seeded lesion segmentation on digital mammograms *IEEE Trans. Med. Imaging* **17** 510–7
Lai S M, Li X and Bischof W F 1989 On techniques for detecting circumscribed masses in mammograms *IEEE Trans. Med. Imaging* **8** 377–86
Mudigonda N R, Rangayyan R M and Desaultels J E L 2000 Gradient and texture analysis for the classification of mammographic masses *IEEE Trans. Med. Imaging* **19** 1032–4
Mudigonda N R, Rangayyan R M and Desaultels J E L 2001 Detection of breast masses in mammograms by density slicing and texture flow-field analysis *IEEE Trans. Med. Imaging* **20** 1215–27
Nishikawa R M 2007 Current status and future directions of computer-aided diagnosis in mammography *Comput. Med. Imaging Graph.* **31** 224–35
Nishikawa R M, Giger M L, Doi K, Metz C E and Yin F-F 1994 Effect of case selection on the performance of computer-aided detection scheme *Med. Phys.* **21** 265–9
Oliver A, Freixenet J, Martí J, Pérez E, Pont J, Denton E R and Zwiggelaar R 2010 A review of automatic mass detection and segmentation in mammographic images *Med. Image Anal.* **14** 87–110
Pereira C S, Fernandes H, Mendonça A M and Campilho A 2007 Detection of lung nodule candidates in chest radiographs *LNCS* **4478** 170–7
Pisano E D *et al* 2005 Diagnostic performance of digital versus film mammography for breast cancer screening *New Engl. J. Med.* **353** 1773–83
Ranawana R and Palade V 2009 Multi-classifier systems—a review and roadmap for developers *Inf. Sci.* **179** 1298–318
Raudys S J and Jain A K 1991 Small sample size effects in statistical pattern recognition: recommendations for practitioners *IEEE Trans. Pattern Anal. Mach. Intell.* **13** 252–64
Sahiner B, Chan H P, Petrick N, Helvie M A and Goodsitt M M 1998a Design of a high-sensitivity classifier based on a genetic algorithm: application to computer-aided diagnosis *Med. Phys.* **43** 2853–71
Sahiner B, Chan H P, Petrick N, Helvie M A and Hadjiiski L M 1998b Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis *Med. Phys.* **25** 516–26
Sahiner B, Chan H P, Petrick N, Helvie M A and Hadjiiski L M 2001 Improvement of mammographic mass characterization using speculation measures and morphological features *Med. Phys.* **28** 1455–65
Sahiner B, Chan H P, Petrick N, Wei D, Helvie M A, Adler D D and Goodsitt M M 1996 Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images *IEEE Trans. Med. Imaging* **15** 598–609
Sampat M P 2005 Computer-aided detection and diagnosis in mammography 2nd edn ed A C Bovik *Handbook of Image and Video Processing* (New York: Academic) pp 1195–217
Samulski M and Karssemeijer N 2011 Optimizing case-based detection performance in a multiview CAD system for mammography *IEEE Trans. Med. Imaging* **30** 1001–9
Santo M D, Molinara M, Tortorella F and Vento M 2003 Automatic classification of clustered microcalcifications by a multiple expert system *Pattern Recognit.* **36** 1467–77
Shavlik J W, Mooney R J and Towell G G 1991 Symbolic and neural learning algorithms: an experimental comparison *Mach. Learn.* **6** 111–43
Suri J S and Rangayyan R M 2006 *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer* (Washington: SPIE)
Tang J, Rangayyan R M, Xu J, Naqa I E and Yang Y 2009 Computer-aided detection and diagnosis of breast cancer with mammography: recent advances *IEEE Trans. Inform. Technol. Biomed.* **13** 236–51
Vapnik V N 1998 *Statistical Learning Theory* (New York: Wiley)
Way T W, Sahiner B, Hadjiiski L M and Chan H P 2010 Effect of finite sample size on feature selection and classification: a simulation study *Med. Phys.* **37** 907–20

Wei D, Chan H P, Helvie M A, Sahiner B, Petrick N, Adler D D and Goodsitt M M 1995 Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis *Med. Phys.* **22** 1501–13

Wei D, Chan H P, Petrick N, Sahiner B, Helvie M A, Adler D D and Goodsitt M M 1997 False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis *Med. Phys.* **24** 903–14

Wei J, Chan H P, Sahiner B, Hadjiiski L M, Helvie M A, Roubidoux M A, Zhou C and Ge J 2006 Dual system approach to computer-aided detection of breast masses on mammograms *Med. Phys.* **33** 4157–68

Weiss G M and Provost F 2003 Learning when training data are costly: the effect of class distribution on tree induction *J. Artif. Intell. Res.* **19** 315–54

Yoon S J and Kim S J 2008 AdaBoost-based multiple SVM-RFE for classification of mammograms in DDSM *IEEE Int. Conf. on Bioinformatics and Biomedicine Workshops* pp 75–82