

RESEARCH

Open Access



Microcalcification detection in full-field digital mammograms with PFCM clustering and weighted SVM-based method

Xiaoming Liu^{1,2*}, Ming Mei¹, Jun Liu^{1,2} and Wei Hu^{1,2}

Abstract

Clustered microcalcifications (MCs) in mammograms are an important early sign of breast cancer in women. Their accurate detection is important in computer-aided detection (CADE). In this paper, we integrated the possibilistic fuzzy *c*-means (PFCM) clustering algorithm and weighted support vector machine (WSVM) for the detection of MC clusters in full-field digital mammograms (FFDM). For each image, suspicious MC regions are extracted with region growing and active contour segmentation. Then geometry and texture features are extracted for each suspicious MC, a mutual information-based supervised criterion is used to select important features, and PFCM is applied to cluster the samples into two clusters. Weights of the samples are calculated based on possibilities and typicality values from the PFCM, and the ground truth labels. A weighted nonlinear SVM is trained. During the test process, when an unknown image is presented, suspicious regions are located with the segmentation step, selected features are extracted, and the suspicious MC regions are classified as containing MC or not by the trained weighted nonlinear SVM. Finally, the MC regions are analyzed with spatial information to locate MC clusters. The proposed method is evaluated using a database of 410 clinical mammograms and compared with a standard unweighted support vector machine (SVM) classifier. The detection performance is evaluated using response receiver operating (ROC) curves and free-response receiver operating characteristic (FROC) curves. The proposed method obtained an area under the ROC curve of 0.8676, while the standard SVM obtained an area of 0.8268 for MC detection. For MC cluster detection, the proposed method obtained a high sensitivity of 92 % with a false-positive rate of 2.3 clusters/image, and it is also better than standard SVM with 4.7 false-positive clusters/image at the same sensitivity.

Keywords: Microcalcification detection; Microcalcification cluster; Computer-aided diagnosis; Possibilistic fuzzy *c*-means; Support vector machine

1 Introduction

Breast cancer is the most frequent form of cancer in women and is also the leading cause of mortality in women each year. The World Health Organization estimated that 521,907 women worldwide died in 2012 due to breast cancer [1]. Studies have indicated that early detection and treatment improve the survival chances of the patients. In order to detect it in its early stage, many countries have established screening programs. Among all the diagnostic methods currently available for detection of

breast cancer, mammography is regarded as the only reliable and practical method capable of detecting breast cancer in its early stage [2].

The screening programs generate large volumes of mammograms to be analyzed. However, due to the complexity of the breast structure, low disease prevalence (approximately 0.5 % [3]), and radiologist fatigue, abnormalities are often ignored. It is reported that about 10–25 % abnormal cases shown in mammography have been wrongly ignored by radiologists [4]. Double reading can improve the detection rate, but it is too expensive and time consuming. Thus, computer-aided cancer detection technologies have been investigated. The adoption of a computer-aided detection (CAD) system could reduce

* Correspondence: lxmspace@gmail.com

¹College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China

²Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System, Wuhan, China

the experts' workload and can improve the early cancer detection rate [5].

Various types of abnormalities can be observed in mammograms, such as microcalcification clusters and mass lesion, distortion in breast architecture, and asymmetry between breasts which are the most dangerous ones. Microcalcification clusters and mass [5–7] are the most common signs of breast cancer, and microcalcification (MC) clusters appear in 30–50 % of diagnosed cases. MCs are calcium deposits of very small dimension and appear as a group of granular bright spots in a mammogram. A typical mammogram with microcalcification clusters is shown in Fig. 1a and the full view of a cluster of microcalcifications in Fig. 1b. Individual MCs are sometimes difficult to detect because of the surrounding breast tissue, their variation in shape, and small dimensions.

Computer-aided detection of microcalcification clusters has been investigated using many different techniques [5, 8]. Roughly speaking, the methods can be classified as traditional enhancement-based method, multiscale analysis, and classifier-based methods. Kim et al. [9] enhanced the mammographic images based on the first derivative (such as Sobel operators and compass operators) and the local statistics.

Laine et al. [10] investigated wavelet multiresolution for mammography contrast enhancement. Three over-complete multiresolution representations were investigated. Contrast enhancement was applied for each level separately, and edge features were extracted in each level and decomposition coefficients were modified based on edge features. Final images were reconstructed with modified coefficients. Improved contrast for irregular structures such as microcalcification was observed on their experiments. In our previous work [11], a new

wavelet-based image enhanced method is proposed. A multiscale measure which matches the human vision system is proposed and used to modify wavelet coefficients. The degree of enhancement can be adjusted by manipulating a single parameter. Ramirez-Cobo et al. [12] used 2D wavelet-based multifractal spectrum for malignant and normal classification.

Several machine learning methods have been used for microcalcification detection. El-Naqa et al. [13] investigated the support vector machine (SVM) classifier for MC cluster detection, and a successive enhancement learning scheme was proposed to improve the performance. On a set of 76 mammogram images containing 1120 MCs, their method obtained a sensitivity of 94 % with an error rate of one false-positive cluster per image. Ge et al. [14] proposed a system to identify microcalcification clusters on full-field digital mammograms (FFDMs) with convolution neural network. The system includes six stages: preprocessing, image enhancement with box-rim filter, segmentation of microcalcification candidates, false-positive (FP) reduction for individual microcalcifications with convolution neural network, region clustering, and FP reduction for clustered microcalcifications. On a dataset of 96 cases with 192 images, they obtained a cluster-based sensitivity of 70, 80, and 90 % at 0.12, 0.61, and 1.49 FPs/image, respectively. Tiedeu et al. [15] detected microcalcifications by integrating image enhancement and the threshold-based segmentation method. Several features were extracted for each region from the enhanced image, and by embedding feature clustering in the segmentation, their method obtained much less false positives than other methods. Experiments were performed on a dataset of 66 images containing 59 MC clusters and 683 MCs, and high sensitivity (100 %) was obtained, balanced by a

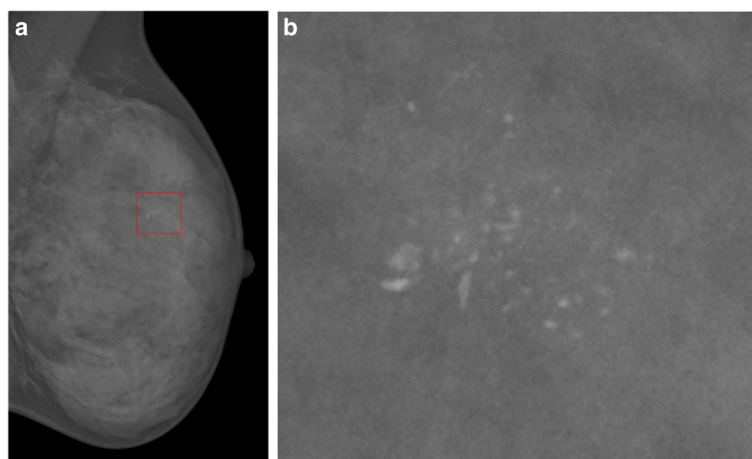


Fig. 1 Mammogram with a microcalcification cluster. **a** Original image. **b** Enlarged view of the microcalcification part of **a** outlined with a red rectangle

lower specificity (87.77 %). In the work of Oliver et al. [16], the individual microcalcification detection is based on local image features of the microcalcifications from a bank of filters. A pixel-based boosting classifier is then trained, and salient features were selected. Clusters are found by inspecting the local neighborhood of each microcalcification. Malar et al. [17] utilized wavelet-based texture features and the extreme learning machine (ELM) for microcalcification detection and classification. One hundred and twenty regions of interest (ROIs; with 32×32 pixels) extracted from the MIAS [18] dataset are used for the experiments. They obtained a classification accuracy of 94 %.

Most of the previous microcalcification detection works have been performed on film-scanned mammograms. With the development of the imaging technique, FFDMs have been widely deployed, and they have better image quality than film-scanned images. We will concentrate on FFDM images.

In this paper, we proposed a novel weighted support vector machine-based microcalcification cluster detection method for FFDM images. Inspired by the work in [19], the possibilistic fuzzy c-means (PFCM) clustering algorithm is used to derive weights for the samples. Several features are extracted and used to train SVM. The proposed method is evaluated on a publically available FFDM dataset [20], consisting of 410 images.

The contributions of the paper are as follows: (1) The weighted SVM is for the first time introduced for MC and MC cluster detection. (2) A novel weighting scheme based on PFCM clustering is introduced to assign weights to samples; unlike the traditional transductive learning-based “pseudo training dataset generating” method, this integration is more simple and principled. (3) A mutual information criterion-based feature selection is investigated for the MC detection in FFDM, and until now, only very few MC detection works have been done on the FFDM dataset.

The rest of the paper is organized as follows: Section 2 introduces the related PFCM technique used, and our method is introduced in Section 3. Experimental results are shown in Section 4. The conclusions and discussions are provided in Section 5.

2 Brief introduction of PFCM clustering

The PFCM is a recently developed clustering algorithm, which has the advantages of the fuzzy c-means (FCM) as well as the possibilistic c-means (PCM) [21] algorithm. Outlier sensitivity is one shortcoming of the FCM clustering. The PCM clustering algorithm [21] can overcome the shortcoming, and it can identify the degree of typicality that a sample has with respect to the group to which it belongs. However, sometimes the prototypes of PCM clusters can coincide, and the PCM will fail in

these cases. Pal et al. proposed a hybridized PFCM clustering [22] to cope with the above shortcomings.

For an unlabeled dataset $X = \{x_1, x_2, \dots, x_n\} \in R^p$, a c -partition of X is a set of (cn) values $\{u_{ik}\}$ that can be written as a $(c \times n)$ matrix $U = [u_{ik}]$, $i = 1, \dots, c$, $k = 1, \dots, n$. The possibilistic and fuzzy c -partitions of X are defined as [22]

$$M_{pcn} = \{U \in R^{cn} : 0 \leq u_{ik} \leq 1 \forall i, k; \forall k \exists i \ni u_{ik} > 0\} \quad (1)$$

$$M_{fcn} = \left\{ U \in M_{pcn} : \sum_{i=1}^c u_{ik} = 1 \forall k; \sum_{k=1}^n u_{ik} > 0 \forall i \right\} \quad (2)$$

The optimization function in PFCM is formulated as [22]

$$J_{pfcM}(X; U, T, V) = \sum_{i=1}^c \sum_{k=1}^n (au_{ik}^m + bt_{ik}^\eta) \times \|x_k - v_i\|^2 + \sum_{i=1}^c \gamma_i \sum_{k=1}^n (1-t_{ik})^\eta \quad (3)$$

with constraints $\sum_{i=1}^c u_{ik} = 1 \forall k$; $0 \leq u_{ik}, t_{ik} \leq 1$, and the constants $a > 0$, $b > 0$, $m > 1$, and $\eta > 1$. $v_i \in R^p$ is the center of the i -th cluster, and x_k is the k -th data sample. The values of a and b represent the relative importance of membership and typicality values in the computation of the prototypes. The parameters m and η represent the absolute weight of the membership value and typicality value, respectively. One can set $b > a$ and $m > \eta$ to reduce the sensitivity to outliers.

The probability (memberships, or relative typicalities, used in FCM) and possibilities (or absolute typicalities, used in PCM) are different. The membership u_{ik} for data x_k belongs to a class c_i which is a function of x_k and all c centroids $\{v_1, \dots, v_k\}$, while the typicality value is a function of x_k and the center v_i alone, as shown below. For example, in a two-class clustering problem, for a noise data point, which is far away from both clusters, the membership to both cluster u_{ik} will be about 0.5 (required by $\sum_{i=1}^2 u_{ik} = 1$), while the typicality values to both cluster t_{ik} will be near zero. For another point which lies between the two clusters (not far away from the centers), the u_{ik} will also be about 0.5, but the typicality values t_{ik} will be a positive number not approaching zero. The probability and typicality can convey different information about the dataset.

PFCM theorem [22]: If $D_{ikA} = \|x_k - v_i\| > 0$, for every i , k , $m > 1$, $\eta > 1$, and if X contains at least c distinct data points, then $(U, T, V) \in M_{fcn} \times M_{pcn} \times \mathcal{R}^{c \times n}$ may minimize J_{pfcM} only if

$$u_{ik} = \left(\sum_{j=1}^c \left(\frac{D_{ikA}}{D_{jkA}} \right)^{2/(m-1)} \right)^{-1}, \quad 1 \leq i \leq c; \quad 1 \leq k \leq n \quad (4)$$

$$t_{ik} = \frac{1}{1 + \left(\frac{D_{ikA}^2}{\mu_{ik}^m} \right)^{1/(\eta-1)}}, \quad 1 \leq i \leq c; \quad 1 \leq k \leq n \quad (5)$$

$$v_i = \sum_{k=1}^n (au_{ik}^m + bt_{ik}^\eta) x_k / \sum_{k=1}^n (au_{ik}^m + bt_{ik}^\eta), \quad 1 \leq i \leq c \quad (6)$$

$$\gamma_i = K \frac{\sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2}{\sum_{k=1}^n \mu_{ik}^m} \quad (7)$$

The iterative process of the algorithm is presented in [22].

3 Proposed method

The workflow of the proposed method is shown in Fig. 2. For each image, suspicious MC regions are extracted with active contour segmentation. Then geometry and texture features are extracted for each suspicious MC, a

mutual information-based supervised criterion is used to select important features, and then PFCM is applied to cluster the samples into two clusters. Weights of the samples are calculated based on possibilities and typicality values from the PFCM, and the ground truth labels. A weighted nonlinear SVM is trained. During the test process, when an unknown image is presented, a similar process is performed. Suspicious regions are located by active contour segmentation, selected features are extracted, and the suspicious MC regions are classified by the more powerful weighted nonlinear SVM. Finally, the MC regions are analyzed with spatial information to locate MC clusters.

3.1 Level set-based MC segmentation

The segmentation of MC consists of two steps: firstly, several edge points are detected and used to initialize the MC segmentation, and then an active contour is used to refine the initial segmentation. The initial step follows the method proposed in [23]. For a given image $f(x, y)$, the edge of a microcalcification to be segmented is a closed contour around a known pixel (x_0, y_0) , which is the location of the local highest grayscale value pixel. For each pixel, a slope value $s(x, y)$ referred to $f(x_0, y_0)$ is defined as [23]

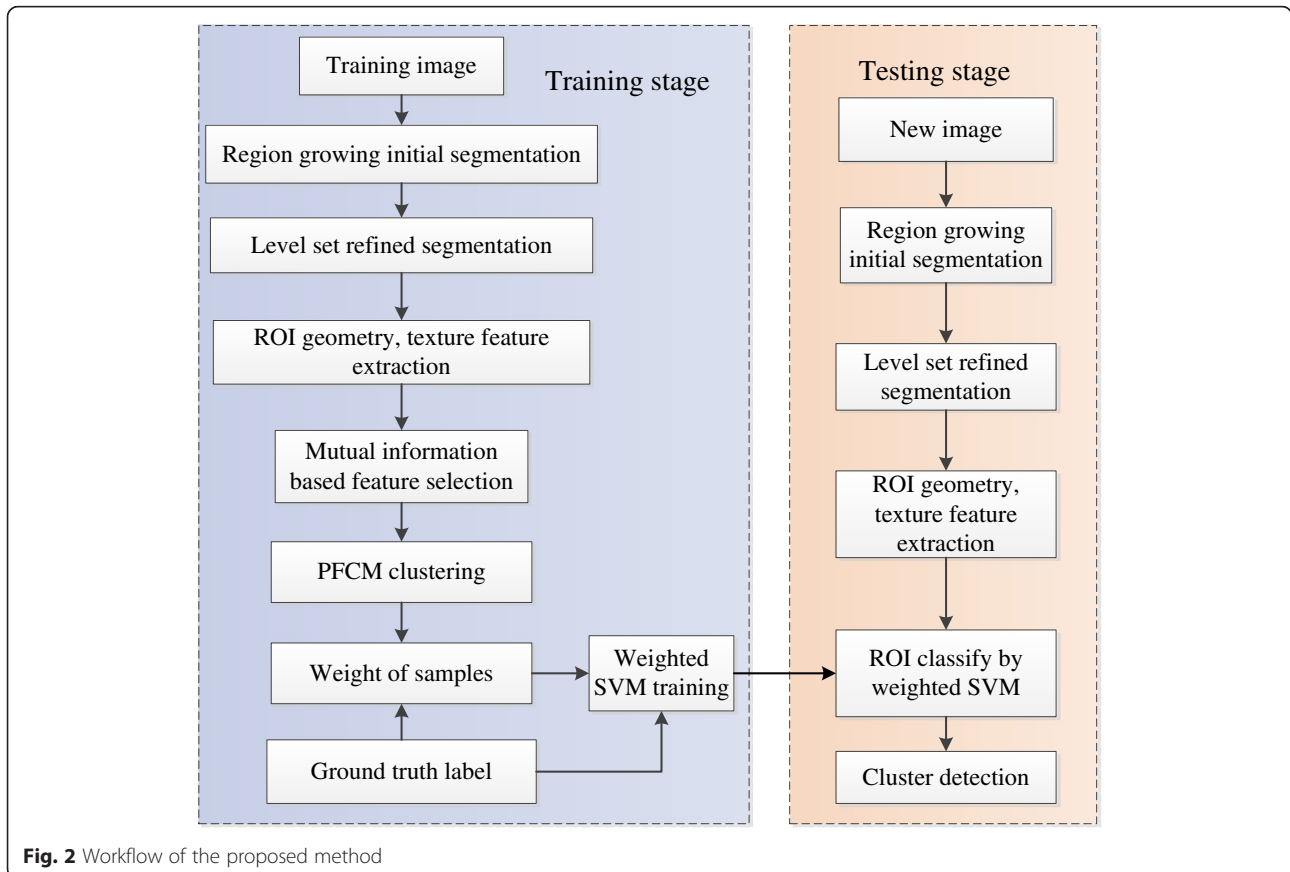


Fig. 2 Workflow of the proposed method

$$s(x, y) = \frac{f(x_0, y_0) - f(x, y)}{d(x_0, y_0, x, y)} \quad (8)$$

where $d(x_0, y_0, x, y)$ is the Euclidean distance between the local maximum pixel (x_0, y_0) and pixel (x, y) . A pixel is considered on the edge if $s(x, y)$ is maximal along a line segment originating from (x_0, y_0) . The length of the considered line segment is chosen as 15 (approximately 1 mm as the spatial resolution in INbreast is 70 μm per pixel). The line search is applied in 16 equally spaced directions originating from the seed pixel. Thus, with each local maximal, 16 edge points are located. Note that the segmentation step will encounter some difficulties for small MC, and the approach used in [24] is adopted to overcome the problem. That is, the segmentation step is performed on the upscaled region containing a possible MC.

For a given local maximal with 16 edge points, a circle is fitted to the points. And the circle is used as the initialization of the level set-based segmentation. Level set originates from the active contour model (snake) [25]. Snake can be edge based or region based.

In this paper, we used the segmentation method we have proposed previously [7, 26]. The final energy functional we used is

$$\begin{aligned} E(\phi, f_1, f_2) = & \lambda_1 \int \left(\int K_\sigma(x-y) |I(y) - f_1(x)|^2 H(\phi(y)) dy \right) dx \\ & + \lambda_2 \int \left(\int K_\sigma(x-y) |I(y) - f_2(x)|^2 (1 - H(\phi(y))) dy \right) dx \\ & + \gamma_1 \int |I(x) - c_1|^2 (1 - H(\phi(x))) dx + \gamma_2 \int |I(x) - c_2|^2 H(\phi(x)) dx \\ & + \mu \int |\nabla H(\phi(x))| dx + \nu \int g\delta(\phi) |\nabla \phi| dx + w \int \frac{1}{2} (|\nabla \phi(x)|^2 - 1) dx \end{aligned} \quad (9)$$

For details about the above function and the numerical implementation, please refer to [7, 26, 27] (see Fig. 3 for an illustration of the segmentation steps).

3.2 Feature extraction from ROI

After segmenting suspicious MC from the ROI, we compute a set of geometry and texture features related to the boundary and the region. Several features used here

have been used in our previous work [7] for mass diagnosis.

3.2.1 Geometry features

Fourteen geometry features are considered in the study, including area (denoted as GF1, where GF means geometry feature), perimeter (GF2), compactness (C, GF3), normalized distance moment (NDM2, NDM3, NDM4, GF4-F6) [28], Fourier feature (FF, GF7) [28], normalized radial length (NRL)-based features (μ_{NRL} , σ_{NRL} , E_{NRL} , AR_{NRL} , GF8-GF11) [29], and relative gradient orientation (RGO)-based features (μ_{RGO} , σ_{RGO} , E_{RGO} , GF12-GF14) [30]. The area is computed by the pixels in the segmented region, and the perimeter is computed by the number of pixels on the boundary. C is a measure of contour complexity versus the enclosed area and is defined as $C = 1 - \frac{4\pi \times (\text{area})}{(\text{perimeter})^2}$. For details about geometry features, please see our previous work [7] and the references therein. The 14 geometry features are listed in Table 1.

3.2.2 Texture features

Besides the shape information of a MC contour, the texture information of the region surrounding the suspicious MC boundary also contains important information for MC analysis [8]. Thus, texture features are also used for MC detection. For each suspected MC, a patch with size 16×16 is extracted [14, 31], whose center is determined by the center of the suspected MC. Besides the average grayscale in the segmented region in a block (denoted as TF1, where TF means texture feature) (16×16 window) and the grayscale difference between the average suspicious region and background (TF2) (the window without taking into account the segmented region), the gray level co-occurrence matrix (GLCM) [32, 33] and wavelet texture features are also extracted.

GLCM has been widely used in mammographic microcalcifications [8] and masses [34]. We use several GLCM features, including autocorrelation (TF3), contrast (TF4), correlation (TF5), cluster prominence (TF6), cluster shade (TF7), energy (TF8), entropy (TF9), homogeneity (TF10), maximum probability (TF11), sum of squares

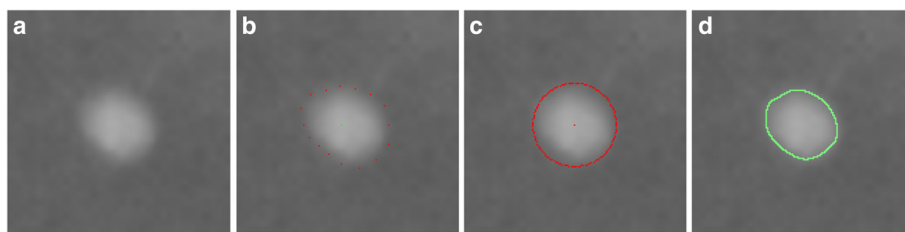


Fig. 3 An example of the segmentation steps. **a** An image patch showing a subtle microcalcification, **b** 16 edge points determined by the algorithm, **c** fitting edge points with a circle, and **d** the result of active contour segmentation

Table 1 Geometry features extracted from a MC boundary

Feature index	Feature	Note
GF1	Area of the MC	Number of pixels in the segmented region
GF2	Perimeter of the MC	Number of pixels on the border of the MC
GF3	Compactness	Calculated with $C = 1 - \frac{4\pi \times (\text{area})}{(\text{perimeter})^2}$
GF4	NDM2	Normalized distance-based moments
GF5	NDM3	
GF6	NDM4	
GF7	Fourier feature	Fourier feature calculated taking boundary pixel as a complex number
GF8	NRL mean	Statistical values from normalized radial length
GF9	NRL standard deviation	
GF10	NRL entropy	
GF11	NRL area ratio	
GF12	RGO mean	Statistical values from relative gradient orientation, measure spiculation, and defined as the acute angle θ between radial direction of a point on the contour and the gradient direction of the point
GF13	RGO standard deviation	
GF14	RGO entropy	

(TF12), sum average (TF13), sum variance (TF14), sum entropy (TF15), difference variance (TF16), difference entropy (TF17), information measure of correlation (TF18, TF19), inverse difference normalized (TF20), and inverse difference moment normalized (TF21).

Besides GLCM-based features, we have also extracted several wavelet-based features. Multiscale representations have been widely used in image processing applications. Wavelet analysis is the most common way to generate such a representation [35]. We used undecimated wavelet transform with the Daubechies 4 filter for each suspicious MC patch (a 16×16 window) in the paper. The entropy and energy of each sub-band are used as features. For an $N \times N$ ($N = 16$) sub-image, normalized energy and entropy are computed as follows [24]:

$$\text{Energy} = \frac{\sum_i \sum_j x_{ij}^2}{N^2} \quad (10)$$

$$\text{Entropy} = - \frac{\sum_i \sum_j \left[\frac{x_{ij}^2}{\text{norm}^2} \right] \log_2 \left[\frac{x_{ij}^2}{\text{norm}^2} \right]}{\log_2 N^2} \quad (11)$$

where x_{ij} is the ij -th pixel value of the sub-images, and

$$\text{norm}^2 = \sum_i \sum_j x_{ij}^2 \quad (12)$$

Twelve sub-images are generated for each ROI with a three-level wavelet decomposition, and as the first-level

decomposition consists of mostly noise, features are extracted from the levels 2 and 3 sub-bands. Thus, 16 wavelet features are extracted for each ROI and are denoted with TF22-TF23 from the energy and entropy feature of the level 2 approximation coefficient matrix, and TF24-TF25, TF26-TF27, and TF28-TF29 from horizontal, vertical, and diagonal coefficient matrices, respectively. TF30-TF37 are defined similarly for level 3 decomposition.

3.3 Feature selection based on mutual information

With the above procedure, a lot of features are extracted to represent the possible MC. However, not every feature is useful to discriminate non-MC and MC. Each feature used here has its physical meaning, and it is important to preserve the intelligibility of the result; several methods, such as principal component analysis (PCA) and linear discriminant analysis (LDA) [36], are not applicable.

Feature selection methods can be categorized into two types: filter methods and wrapper methods [37]. The performance of wrapper methods is dependent on the specific classifiers, while the performance of filter methods is usually independent of the classifiers. In this paper, we concentrate on the mutual information (MI)-based filter feature selection method.

Positive (MC) and negative (non-MC) samples are needed to select features. The positive samples are obtained with the annotation in the image database, and the negative samples are those ROIs segmented by active contour but do not contain MC. In this way, the selection of non-MC samples is tuned with the whole detection procedure. It is an advantage compared with other commonly used random sample methods, as used in [13].

In information theory, MI calculates the statistical dependence between two random variables and can be used to measure the relative utility of each feature to a classification problem. The MI between two random variables X and Y is defined as

$$I(X, Y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (13)$$

where $p(x, y)$ is the joint probability density function of continual random variables, and $p(x)$ and $p(y)$ are the marginal probability density functions. MI can also be defined with Shannon entropy

$$\begin{aligned} I(X, Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y) \end{aligned} \quad (14)$$

where $H(x) = - \int p(x) \log p(x) dx$, $H(y|x) = - \iint p(x, y) \log p(y|x) dx dy$, and $H(x, y) = - \iint p(x, y) \log p(x, y) dx dy$ are the

Shannon entropies. An explanation of MI for feature selection is as follows: Let Y be a variable representing the class label (e.g., MC or non-MC) and X a variable denoting a feature. The entropy $H(Y)$ is known to be a measure of the amount of uncertainty about Y , while $H(Y|X)$ is the amount of uncertainty left in Y when knowing an observation X . Therefore, MI can be seen as the amount of information that the measure at X has about the class label Y . Thus, MI measures the capability of this feature to predict the class label.

As is known, the best k single features are usually not the best k combined features, since there may exist redundancy between these features. The minimum redundancy maximum relevance (mRMR) method [38] considered this problem, and it selects features that have the highest relevance with the target class and are also minimally redundant. Denote the i -th feature as f_i and the class variable as c . The maximum relevance criterion selects the top m features in the descent order of $I(f_i, c)$, i.e., the best m individual features correlated to the class labels:

$$\max_S \frac{1}{|S|} \sum_{f_i \in S} I(f_i; c) \quad (15)$$

where S is the set of selected features, and $|S|$ is the cardinality of the set.

Due to feature correlations among features, the m best separated features are not the best combined m features. The minimum redundancy criteria are introduced to remove the redundancy among features:

$$\min_S \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i, f_j) \quad (16)$$

The final features are selected sequentially, to select the m -th feature after obtaining the $m-1$ features S_{m-1} , by solving the following optimization problem [38]:

$$\max_{f_j \in F - S_{m-1}} \left[I(f_j; c) - \frac{1}{m-1} \sum_{f_i \in S_{m-1}} I(f_j; f_i) \right] \quad (17)$$

3.4 Clustering with PFCM and weight samples

With the obtained MC and non-MC samples, using the features selected by the mRMR criterion, PFCM is applied to cluster the samples. Each sample with selected feature values is regarded as a data point. As shown above, for each sample after PFCM clustering, it has a probability and a typicality value.

Let y_i denote the label of sample i , and let $y_i \in \{+1, -1\}$ denote the class variable (MC or non-MC) which we can obtain by the doctor's manual annotation. Let MU_i denote the probability of sample i belonging to

calcification, let MT_i^{+1} denote the typicality value of sample i belonging to MC, and let MT_i^{-1} denote the typicality value of it belonging to non-MC. MU_i , MT_i^{+1} , and MT_i^{-1} can be obtained by PFCM clustering, and their value ranges are between 0 and 1.

We want to give more weights to the samples with higher confidence and define the weight $W1_i$ as

$$W1_i = \frac{1+y_i}{2} \times MU_i + \frac{1-y_i}{2} \times (1-MU_i) \quad (18)$$

If a sample is MC ($y_i = +1$), we can simplify it to $W1_i = MU_i$, and if it is non-MC ($y_i = -1$), $W1_i$ can be simplified to be $1 - MU_i$. In this way, if a sample is MC and the possibility value of it belongs to MC obtained by PFCM is high, the weight $W1_i$ is high; otherwise, the weight is low.

Besides the confidence value, we also used the typicality values outputted by PFCM. The weight term considering the typicality value is defined as follows:

$$W2_i = \frac{1+y_i}{2} \times MT_i^1 \times (1-MT_i^{-1}) + \frac{1-y_i}{2} \times (1-MT_i^1) \times MT_i^{-1} \quad (19)$$

For a typical sample belonging to MC or non-MC, its weight $W2_i$ is high. For example, for a typical MC sample x_i , the first term in $W2_i$ approaches 1, since $y_i = 1$, $MT_i^1 \approx 1$, $MT_i^{-1} \rightarrow 0$. For a typical non-MC, $W2_i$ is also large due to the second term, while $W2_i$ is small for a noise point, since in this case, both MT_i^{+1} and MT_i^{-1} approach 0.

We take both possibility information and typical information into consideration, and the final weight of sample i we defined is

$$W3_i = W1_i * W2_i \quad (20)$$

3.5 Weighted SVM-based classification

Given a set of vectors (x_1, \dots, x_n) and their corresponding labels (y_1, \dots, y_n) with $y_i \in \{+1, -1\}$, the SVM classifier defines a hyperplane (w, b) in kernel space that separates the training data by a maximal margin.

For the weighted SVM, each sample consists of a data vector x_i , a label y_i as in standard SVM; besides, a sample also contains a confidence value v_i . Define the effective weighted functional margin of weighted sample (x_i, y_i, v_i) with respect to a hyperplane (w, b) and a margin normalization function f to be $f(v_i)y_i(\langle w \cdot x_i \rangle + b)$, where f is a monotonically decreasing function. To tolerate noise and outliers, more training samples than just those close to the boundary need to be considered.

Definition (margin slack variable) [39]: Given a value $\gamma > 0$, the margin slack variable of a sample (x_i, y_i) with

respect to the hyperplane (w, b) and target margin γ is defined to be

$$\xi_i = \max(0, \gamma - y_i(\langle w \cdot x_i \rangle + b)) \quad (21)$$

The quantity measures how much a point fails to have a margin γ from the hyperplane (w, b) . If x_i is misclassified by (w, b) , then $\xi_i > 0$. To generalize the soft margin classifier to the weighted soft margin classifier, the weighted version of the slack variable is introduced.

Definition (effective weighted margin slack variable) [39]: The effective weighted margin slack variable of a sample (x_i, y_i, v_i) with respect to a hyperplane (w, b) and margin normalization function f , slack normalization function g , and target margin γ is defined as

$$\begin{aligned} \xi_i^w &= g(v_i) \max(0, \gamma - y_i f(v_i)(\langle w \cdot x_i \rangle + b)) \\ &= g(v_i) \xi_i \end{aligned} \quad (22)$$

where f is a monotonically decreasing function such that $f(\cdot) \in (0, 1]$, and g is a monotonically increasing function such that $g(\cdot) \in (0, 1]$.

The weighted SVM optimization problem can be formulated as follows: Given a training sample set $S = ((x_1, y_1, v_1), \dots, (x_n, y_n, v_n))$, the hyperplane (w, b) that solves the following optimization problem

$$\begin{aligned} &\text{minimize } \langle w \cdot w \rangle + C \sum_{i=1}^n g(v_i) \xi_i \\ &\text{s.t. } y_i(\langle w \cdot x_i \rangle + b) f(v_i) \geq 1 - \xi_i, i = 1, \dots, n \\ &\quad \xi_i \geq 0, i = 1, \dots, n \end{aligned} \quad (23)$$

realizes the maximal weighted soft margin hyperplane. If both functions f and g are set to be constant at 1, then the WSVM coincides with standard SVM.

In the above formulation, the final decision plane will be less affected by those margin-violating samples with low confidence, and samples with high confidence have higher impact on the final decision plane. The optimization problem can be solved using the sequential minimal optimization technique as in standard SVM. The function $f(x)$ and $g(x)$ are set as used in [39].

4 Experimental results

4.1 Mammogram database

The proposed method was tested on the publically available INbreast database [20]. The database was acquired from the Breast Centre in CHSJ, Porto, between April 2008 and July 2010, and the acquisition equipment was the MammoNovation Siemens FFDM. The database has a total of 115 cases (410 images), from which 90 cases are from women with both breasts affected (four images per case) and 25 cases are from mastectomy patients (two images per cases). Several types of lesions (masses, calcifications, architectural distortion) were included. The pixel size is 70 μm , with 14-bit resolution. The image

matrix was 3328×4084 or 2560×3328 pixels. All images were saved in DICOM format. The database has a large portion of calcifications. Among the 410 images, calcifications presented in 301 images, and 27 sets of microcalcification clusters occurred in 21 images (≈ 1.3 clusters per image). A total of 6880 microcalcifications were individually identified in 299 images (≈ 23.0 calcifications per image).

For our investigation, very small MCs (number of pixels less than 3) are ignored and treated as normal. Note that the tiny MC can be detected with techniques such as wavelet transform. With such criterion, we obtained 2748 MCs on 232 images (≈ 11.8 MCs per image). Since the MC cluster in this criterion in the dataset is small, and a cluster can contain dozens of MC, we used a variant criterion about the MC cluster [40]. That is, a group of objects classified as MCs is considered to be a true-positive (TP) cluster only if at least three true calcifications should be detected by an algorithm within an area of 1 cm^2 . A group of objects classified as MCs is labeled as a FP cluster provided that the objects satisfy the cluster requirement but do not contain true MCs. In this way, 76 MC clusters are defined.

4.2 Results

4.2.1 Segmentation results

Our method first extracted suspicious MC regions, and then use WSVM to reduce the false positives. If a MC is missed in the segmentation step, then it will not show up in the final detection. Visual inspection of the output images and their corresponding annotations showed that all the MC clusters had been detected in the first segmentation stage. Figure 4 shows the segmentation stage of an image. In Fig. 4a, the MC cluster is circled. Figure 4b showed the output mask of the first stage, and it can be seen that the MC cluster has been detected correctly. The enlarged parts are shown in Fig. 4c, d.

To quantitatively evaluate the segmentation results, we used Dice coefficient D , which has been widely used for segmentation evaluation. The value of D ranges from 0 (no overlap) to 1 (perfect overlap) and is defined by $D = \frac{2(A \cap G)}{(A \cap G) + A \cup G} \times 100\%$, where A is the region segmented by a method, and G is the manually labeled region. The averaged value for D on 100 MC images was 93.8 %, and it indicates that the proposed method is accurate for MC segmentation.

4.2.2 Selection of features

As there are 2748 MCs in total, about half of them are used as training dataset, and 1382 MCs from the 116 images containing MC are used in the training as positive samples. In addition, twice as many non-MC examples were selected from the 117 images (one contains

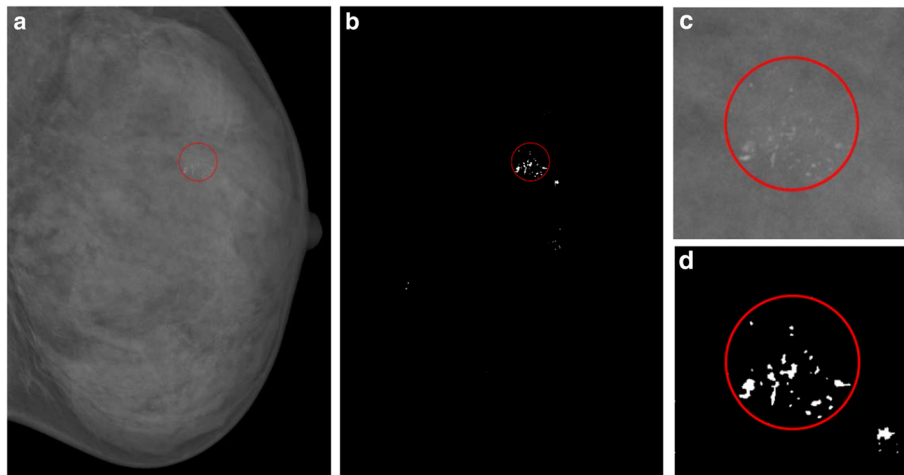


Fig. 4 Segmentation of suspected of MC. **a** Original image; the *red circle* is the annotation of MC cluster. **b** Segmentation mask of possible MC. **c** Enlarged part of the MC cluster in **a**. **d** Enlarged part of segmentation mask of **b**

only MC cluster) and 88 images not containing MC or MC cluster. That is, in total, 205 images are used for training (containing 1382 MCs, 38 MC clusters, and 2764 non-MCs), and the remaining 205 images (117 images containing MC or MC cluster and 88 images not containing MC or MC cluster) are used for testing; the test dataset contains 1366 MCs on 116 images and 38 MC clusters on 10 images.

Unlike usually random selection, the non-MC examples are selected considering the initial detection procedure. That is, the non-MC examples are selected from the level set segmented regions not containing MC. Thus, the negative examples are specific to the used segmentation method, which is an advantage over the traditional random selection method. There were 4146 (1382×3) training examples in total.

Each MC or non-MC was covered by a 16×16 (about $1.12 \text{ mm} \times 1.12 \text{ mm}$ with pixel spatial resolution of 0.07 mm) window whose center coincided with the center of the suspected MC. Geometry features are extracted from the level set segmentation, and GLCM and wavelet texture features are extracted from the window.

With the extracted features and known class label, MI is used to select the important features. We have extracted 51 features (14 geometry features, 2 grayscale features, 19 GLCM features, and 16 wavelet features) to represent MC and non-MC, and the top 30 features ranked by MI are shown in Table 2.

We used a fivefold cross-validation method to select the number of features; the training samples were equally split into five subsets, four subsets were used as the training dataset and the remaining one subset was used for testing. The averaged performances were recorded to set parameter values. For the classifier here, we used a standard SVM (with radial basis function kernel) without weights;

more specifically, the LIBSVM toolbox [41] was used. The parameters C and σ in the SVM were obtained with cross-validation from set $\{2^{-5}, 2^{-4}, \dots, 2^0, \dots, 2^5\}$. Denote the true-positive number of a classifier as TP, the false-positive number as FP, the true-negative number as TN, and the false-negative number as FN. Then the TPR (true-positive rate), TNR (true-negative rate), and accuracy are defined as $\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$, $\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}$, and $\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}$.

Figure 5 shows the classification accuracy with different numbers of features. This figure shows that the accuracy increases with the added features initially, but it begins to decrease after several features are selected, which indicate that some features may degrade the classifier's performance. The best number of feature used here is 22, and we will use these features in the following experiments. From the features listed in Table 2, we can see selected features including both geometry features and texture features, which indicate that both geometry and texture features are useful to separate MC from non-MC. The top feature is the grayscale difference feature, which is in accordance with the typical characteristic that MC is brighter than the background. The second top feature is the compactness geometry feature, which is also useful to distinguish MC from other bright regions, for example, vessel, since the MC is typically compact while the vessel region is elongated. While the other used features may not have a direct explanation, they contain discriminating information for false-positive reduction.

4.2.3 Detection results and comparison with unweighted SVM

After deciding the number of selected features, each data sample is represented by a 22-feature vector. Then the

Table 2 Top 30 features ranked by mutual information filter

Order	Feature no. and name	Order	Feature no. and name	Order	Feature no. and name
1	TF2, grayscale difference	11	TF31, entropy in approximate matrix of level 3 with wavelet	21	GF11, NRL area ratio
2	GF3, compactness	12	TF3, autocorrelation with GLCM	22	GF13, RGO standard deviation
3	TF17, difference entropy with GLCM	13	GF10, NRL entropy	23	TF20, inverse difference normalized with GLCM
4	TF23, entropy in approximate matrix of level 2 with wavelet	14	TF6, cluster prominence with GLCM	24	TF22, energy in approximate matrix of level 2 with wavelet
5	TF4, contrast with GLCM	15	GF8, NRL mean	25	TF9, entropy with GLCM
6	GF1, area of the MC	16	GF4, NDM2	26	TF33, entropy in horizontal coefficient matrix of level 3 with wavelet
7	GF7, Fourier feature	17	TF25, entropy in horizontal coefficient matrix of level 2 with wavelet	27	TF12, sum of squares with GLCM
8	TF1, the average grayscale in segmented region	18	TF36, energy in diagonal coefficient matrix of level 3 with wavelet	28	TF10, homogeneity with GLCM
9	GF12, RGO mean	19	TF15, sum entropy with GLCM	29	TF28, energy in diagonal coefficient matrix of level 2 with wavelet
10	GF2, perimeter of the MC	20	GF5, NDM3	30	TF32, energy in horizontal coefficient matrix of level 3 with wavelet

samples in the training are clustered and weighted as in Section 3.4. A weighted SVM is trained as introduced in Section 3.5 and then used for testing.

We measured both MC detection and MC cluster detection for the experiments. The MC cluster was identified by grouping the objects that have been determined by the algorithm to be MC. The receiver operating characteristic (ROC) curves are used to evaluate the performance of MC detection, and the free-response receiver operating characteristic (FROC) curves are used

to evaluate the performance of MC cluster detection. The standard SVM without weighting is used to evaluate the effect of the PFCM-based weighting scheme. An ROC curve is a plot of operating points which can be considered as a plot of true-positive rate as a function of false-positive rate. The curve is generated by thresholding the output possibilities of MC of the classifier. A FROC curve is a plot of the correct detection rate (true-positive rate) achieved by a classifier versus the average number of false positives (FPs) per image varied over the

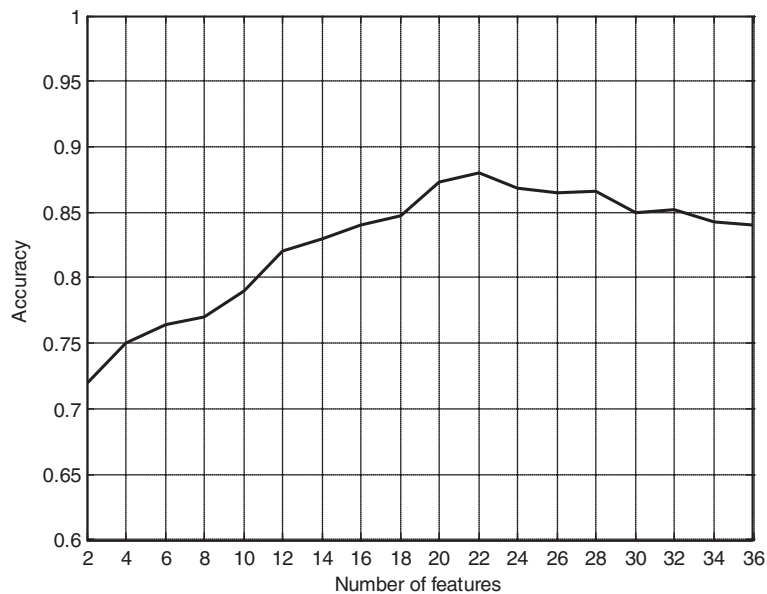


Fig. 5 Accuracy of the SVM classification versus the number of selected features

decision threshold. A FROC curve can provide a summary of the trade-off between detection sensitivity and specificity.

We compared the performance of the standard unweighted SVM and the proposed PFCM clustering-based weighted SVM on MC classification. The test set contains 1366 true MCs, and we selected 2732 non-MC samples from the segmentation on test images, similar to the training step. The performance of standard unweighted SVM and our weighted SVM is shown in Fig. 6 with a ROC curve. The AUC (A_z), which is the area under the ROC curve, is used to compare the performance of the two classification methods. The AUC for standard unweighted SVM is 82.68 %, and the AUC of the proposed PFCM-based weighted SVM is 86.76 %. We can see that with the same training samples and test samples, the proposed weight SVM achieved better performance than the standard unweighted SVM.

The performances of the proposed weighted SVM approach, along with the standard unweighted SVM, are also presented for MC cluster detection with the FROC curve, as shown in Fig. 7. The proposed method obtained a high sensitivity of 92 % with a FP rate of 2.3 FP clusters/image. At a similar sensitivity, the FP rate of the standard unweighted SVM obtained 4.7 FP clusters/image. It can be seen that the proposed weighted SVM also outperformed the standard unweighted SVM for MC cluster detection.

5 Discussion and conclusion

Automatic detection of microcalcification in mammograms has been investigated by many researchers in the

past two decades. In [15], Tiedeu et al. segmented microcalcifications with an adaptive threshold method on the enhanced image, and a set of moment-based geometrical features were used for false-positive reduction. On a dataset of 66 images containing 59 MC clusters and 683 MCs, they obtained a sensitivity of 100 % with a low specificity of 87.77 %. They also performed benign/malignant classification. Oliver et al. [16] extracted image features with a bank of filters, and a boosting method is used to separate MC from non-MC. The dataset they used included the MIAS dataset (322 mammograms) and another 280 FFDM mammograms. Their method's performance for MC is $A_z = 0.85$ with ROC analysis, and for MC clusters, the result is 80 % sensitivity at one false-positive cluster per image.

In [42], Nunes et al. obtained $A_z = 0.93$ for MC detection on a database of 121 mammograms by combining three contrast enhancement techniques. Papadopoulos et al. [43] investigated five image enhancement techniques and obtained $A_z = 0.92$ for MC detection on a database consisting of 60 mammograms from the MIAS and Nijmegen databases. Linguraru et al. [44] proposed MC cluster detection based on a biologically inspired contrast detection algorithm, integrated with a preprocessing step (curvilinear structure removal and image enhancement). They obtained a 95 % sensitivity with 0.4 false positives per image on a small subset of the Digital Database for Screening Mammography (DDSM) dataset [45] (82 images, 58 of which contain microcalcification and the other 24 were normal ones, the number of MC cluster is 82). Ge et al. [46] developed two systems to detect microcalcification clusters, one for FFDM and the other one for screen-film

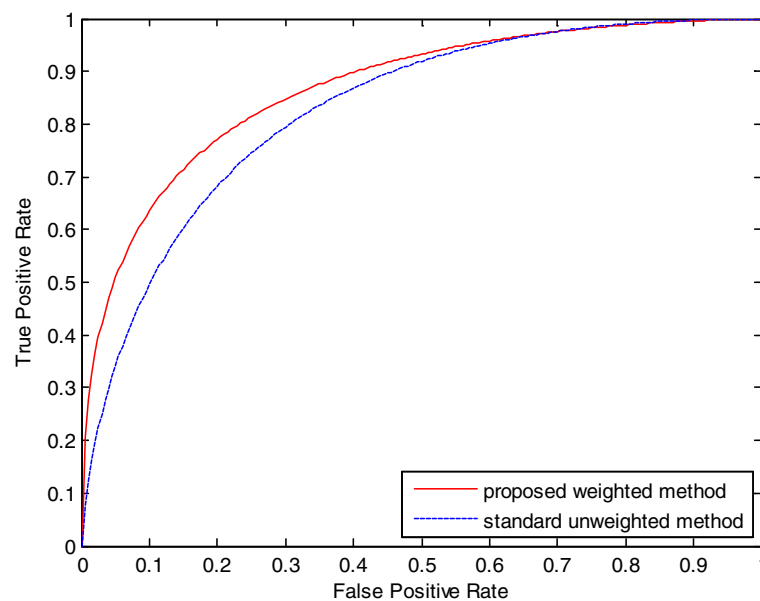


Fig. 6 ROC curve of the standard unweighted SVM and the proposed PFCM-based weighted SVM

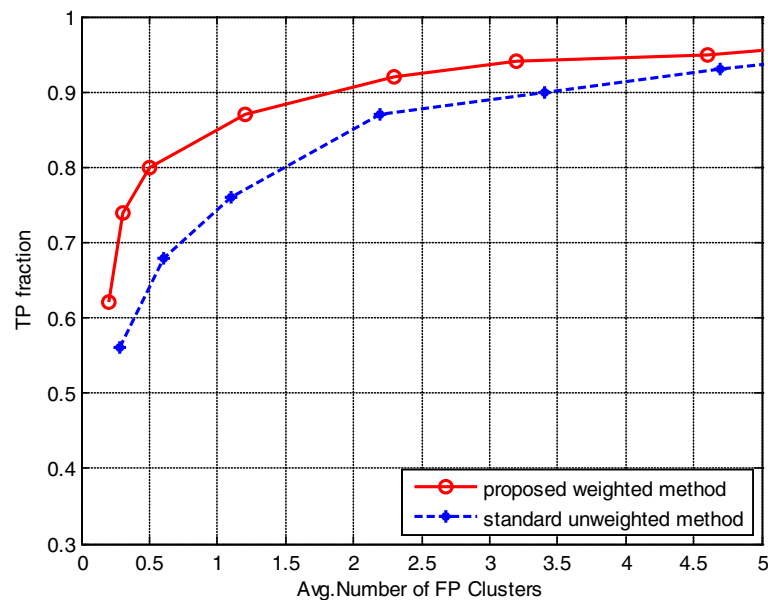


Fig. 7 FROC curves of the standard unweighted SVM and the proposed PFCM-based weighted SVM

mammograms (SFM). They obtained an average of 0.96 or 2.52 false positives per image at 90 % sensitivity on FFDMs and SFMs, respectively. The FFDM dataset they used includes 96 mammograms with microcalcifications and 108 normal mammograms.

It is hard to directly compare different methods, since the used datasets are different, and the definition for MC cluster sometimes is also different. On the DDSM dataset [45], usually different subsets are used in investigations. Most of the above techniques are developed for film-scanned mammograms. Here we developed the method on the FFDMs, as images from FFDM have better image quality than film-scanned images and are also widely deployed. From the above results, it can be seen that the performance of our method with 92 % sensitivity at 2.3 false-positive clusters per image is better or similar to the above results. It should be noted that our method is investigated on a larger dataset.

In this paper, we proposed a weighted SVM technique for detection of MC clusters in FFDM. In this approach, suspicious MC regions are first segmented with active contour, and then the regions are classified by a trained weighted SVM. The non-MC training samples are selected from the segmented regions, which can be better tuned to the whole procedure than random sampling. Mutual information criterion was used to select important features, and among the extracted 51 features, 22 features are selected and used in the training and test process. The training samples are weighted with the possibility and typicality value of a sample belonging to MC output by the novel introduction of possibilistic fuzzy c-means (PFCM) clustering. Experimental results with

ROC and FROC analysis using a set of 410 FFDM mammograms demonstrated that the proposed method outperformed the standard unweighted SVM. The proposed weight scheme may be also applicable to other classifiers, such as random forest, and we will investigate these problems in the future. Besides, we will try to adapt the method for traditional film-scanned mammograms.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (No. 61403287, No. 61472293, No. 31201121) and the Natural Science Foundation of Hubei Province (No. 2014CFB288).

Received: 1 March 2015 Accepted: 13 July 2015

Published online: 12 August 2015

References

1. GLOBOCAN 2012 v1.0, *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]* (International Agency for Research on Cancer, Lyon, France, 2013). Available from: <http://globocan.iarc.fr>
2. CH Lee, DD Dershaw, D Kopans, P Evans, B Monsees, D Monticciolo, RJ Brenner, L Bassett, W Berg, S Feig, E Hendrick, E Mendelson, C D'Orsi, E Sickles, LW Burhenne, Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J. Am. Coll. Radiol.* **7**, 18–27 (2010)
3. M Giger, Current issues in CAD for mammography, in *Digital Mammography* (Elsevier, Philadelphia, 1996), pp. 53–59
4. SV Destounis, P DiNitto, W Logan-Young, E Bonaccio, ML Zuley, KM Willison, Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience. *Radiology* **232**, 578–584 (2004)
5. J Tang, RM Rangayyan, J Xu, I El Naqa, Y Yang, Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans. Inf. Technol. Biomed.* **13**, 236–251 (2009)

6. X Liu, Z Zeng, A new automatic mass detection method for breast cancer with false positive reduction. *Neurocomputing* **152**, 388–402 (2015)
7. X Liu, J Tang, Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method. *IEEE Syst. J.* **8**, 910–920 (2014)
8. H-D Cheng, X Cai, X Chen, L Hu, X Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Patt. Recog.* **36**, 2967–2991 (2003)
9. JK Kim, JM Park, KS Song, H Park, Adaptive mammographic image enhancement using first derivative and local statistics. *IEEE Trans. Med. Imaging* **16**, 495–502 (1997)
10. AF Laine, S Schuler, J Fan, W Huda, Mammographic feature enhancement by multiscale analysis. *IEEE Trans. Med. Imaging* **13**, 725–740 (1994)
11. J Tang, X Liu, Q Sun, A direct image contrast enhancement algorithm in the wavelet domain for screening mammograms. *IEEE J. Sel. Top. Sign. Proces.* **3**, 74–80 (2009)
12. P Ramírez-Cobo, B Vidakovic, A 2D wavelet-based multiscale approach with applications to the analysis of digital mammograms. *Comput. Stat. Data. An.* **58**, 71–81 (2013)
13. I El-Naqa, Y Yang, MN Wernick, NP Galatsanos, RM Nishikawa, A support vector machine approach for detection of microcalcifications. *IEEE Trans. Med. Imaging* **21**, 1552–1563 (2002)
14. J Ge, B Sahiner, LM Hadjiiski, HP Chan, J Wei, MA Helvie, C Zhou, Computer aided detection of clusters of microcalcifications on full field digital mammograms. *Med. Phys.* **33**, 2975–2988 (2006)
15. A Tiedeu, C Daul, A Kentsop, P Graebing, D Wolf, Texture-based analysis of clustered microcalcifications detected on mammograms. *Digit. Signal Process.* **22**, 124–132 (2012)
16. A Oliver, A Torrent, X Lladó, M Tortajada, L Tortajada, M Sentís, J Freixenet, R Zwiiggelaar, Automatic microcalcification and cluster detection for digital and digitised mammograms. *Knowl.-Based Syst.* **28**, 68–75 (2012)
17. E Malar, A Kandaswamy, D Chakravarthy, A Giri Dharan, A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine. *Comput. Biol. Med.* **42**, 898–905 (2012)
18. J Suckling, J Parker, DR Dance, S Astley, I Hutt, C Boggis, I Ricketts, E Stamatakis, N Cerneaz, SL Kok, P Taylor, D Betal, J Savage, The mammographic image analysis society digital mammogram database, in *Excerpta Medica. International Congress Series*, (Excerpta Medica, Amsterdam 1994), pp. 375–378
19. J Quintanilla-Domínguez, B Ojeda-Magaña, A Marcano-Cedeño, MG Cortina-Januchs, A Vega-Corona, D Andina, Improvement for detection of microcalcifications through clustering algorithms and artificial neural networks. *EURASIP J. Adv. Sig. Proc.* **2011**, 91 (2011)
20. IC Moreira, I Amaral, I Domingues, A Cardoso, MJ Cardoso, JS Cardoso, INbreast: toward a full-field digital mammographic database. *Acad. Radiol.* **19**, 236–248 (2012)
21. R Krishnapuram, JM Keller, A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* **1**, 98–110 (1993)
22. NR Pal, K Pal, JM Keller, JC Bezdek, A possibilistic fuzzy c-means clustering algorithm. *IEEE Trans. Fuzzy Syst.* **13**, 517–530 (2005)
23. IN Bankman, T Nizialek, I Simon, OB Gatewood, IN Weinberg, WR Brody, Segmentation algorithms for detecting microcalcifications in mammograms. *IEEE Trans. Inf. Technol. Biomed.* **1**, 141–149 (1997)
24. H Soltanian-Zadeh, F Rafiee-Rad, S Pourabdollah-Nejad D, Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms. *Patt. Recog.* **37**, 1973–1986 (2004)
25. M Kass, A Witkin, D Terzopoulos, Snakes: active contour models. *Int. J. Comput. Vision* **1**, 321–331 (1988)
26. J Tang, X Liu, Classification of breast mass in mammography with an improved level set segmentation by combining morphological features and texture features, in *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies* (Springer, New York, 2011), pp. 119–135
27. C Li, C-Y Kao, JC Gore, Z Ding, Minimization of region-scalable fitting energy for image segmentation. *IEEE Trans. Image Processing* **17**, 1940–1949 (2008)
28. L Shen, RM Rangayyan, JL Desautels, Application of shape analysis to mammographic calcifications. *IEEE Trans. Med. Imaging* **13**, 263–274 (1994)
29. J Kilday, F Palmieri, M Fox, Classifying mammographic lesions using computerized image analysis. *IEEE Trans. Med. Imaging* **12**, 664–669 (1993)
30. B Sahiner, H-P Chan, N Petrick, MA Helvie, LM Hadjiiski, Improvement of mammographic mass characterization using spiculation measures and morphological features. *Med. Phys.* **28**, 1455–1465 (2001)
31. J Fu, S-K Lee, ST Wong, J-Y Yeh, A-H Wang, H Wu, Image segmentation feature selection and pattern classification for mammographic microcalcifications. *Comput. Med. Imaging Graph.* **29**, 419–429 (2005)
32. RC Gonzalez, RE Woods, *Digital Image Processing* (Prentice Hall, New Jersey, 2002)
33. R Haralick, I Dinstein, K Shanmugam, Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **3**, 610–621 (1973)
34. X Liu, J Liu, D Zhou, J Tang, A benign and malignant mass classification algorithm based on an improved level set segmentation and texture feature analysis, in *2010 4th International Conference on Bioinformatics and Biomedical Engineering (ICBBE)*, Chengdu, 18–20 June 2010, pp. 1–4
35. S Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, New York, 1999)
36. RO Duda, PE Hart, DG Stor, *Pattern Classification* (Wiley, New York, 2001)
37. R Kohavi, GH John, Wrappers for feature subset selection. *Artif. Intell.* **97**, 273–324 (1997)
38. H Peng, F Long, C Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Patt. Anal. Mach. Int.* **27**, 1226–1238 (2005)
39. X Wu, R Srihari, Incorporating prior knowledge with weighted margin support vector machines, in *Proceedings of the Tenth ACM SIGKDD*, 2004, pp. 326–333
40. M Kallergi, GM Carney, J Gavria, Evaluating the performance of detection algorithms in digital mammography. *Med. Phys.* **26**, 267–275 (1999)
41. C-C Chang, C-J Lin, LIBSVM: a library for support vector machines. *ACM T. Intel. Syst. Tec.* **2**, 1–27 (2011)
42. FL Nunes, H Schiabel, CE Goes, Contrast enhancement in dense breast images to aid clustered microcalcifications detection. *J. Digit. Imaging* **20**, 53–66 (2007)
43. A Papadopoulos, DI Fotiadis, L Costaridou, Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. *Comput. Biol. Med.* **38**, 1045–1055 (2008)
44. MG Linguraru, K Marias, R English, M Brady, A biologically inspired algorithm for microcalcification cluster detection. *Med. Image Anal.* **10**, 850–862 (2006)
45. M Heath, K Bowyer, D Kopans, R Moore, P Kegelmeyer, The digital database for screening mammography, in *Proceedings of the 5th International Workshop on Digital Mammography*, 2000, pp. 212–218
46. J Ge, LM Hadjiiski, B Sahiner, J Wei, MA Helvie, C Zhou, HP Chan, Computer-aided detection system for clustered microcalcifications: comparison of performance on full-field digital mammograms and digitized screen-film mammograms. *Phys. Med. Biol.* **52**, 981–1000 (2007)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com