Detection of microcalcifications in mammograms using error of prediction and statistical measures

Begoña Acha
Carmen Serrano
Universidad de Sevilla
Escuela Superior de Ingenieros
Departamento de Teoría de la Señal y Comunicaciones
Camino de los Descubrimientos s/n
41092-Sevilla, Spain
E-mail: cserrano@us.es

Rangaraj M. Rangayyan
J. E. Leo Desautels
University of Calgary
Schulich School of Engineering
Department of Electrical and Computer Engineering
2500 University Drive Northwest
Calgary, Alberta, Canada T2N 1N4

Abstract. *A two-stage method for detecting microcalcifications in mammograms is presented. In the first stage, the determination of the candidates for microcalcifications is performed. For this purpose, a 2-D linear prediction error filter is applied, and for those pixels where the prediction error is larger than a threshold, a statistical measure is calculated to determine whether they are candidates for microcalcifications or not. In the second stage, a feature vector is derived for each candidate, and after a classification step using a support vector machine, the final detection is performed. The algorithm is tested with 40 mammographic images, from Screen Test: The Alberta Program for the Early Detection of Breast Cancer with 50-μm resolution, and the results are evaluated using a free-response receiver operating characteristics curve. Two different analyses are performed: an individual microcalcification detection analysis and a cluster analysis. In the analysis of individual microcalcifications, detection sensitivity values of 0.75 and 0.81 are obtained at 2.6 and 6.2 false positives per image, on the average, respectively. The best performance is characterized by a sensitivity of 0.89, a specificity of 0.99, and a positive predictive value of 0.79. In cluster analysis, a sensitivity value of 0.97 is obtained at 1.77 false positives per image, and a value of 0.90 is achieved at 0.94 false positive per image.* © 2009 SPIE and IS&T.
[DOI: 10.1117/1.3099710]

1 Introduction: Calcifications in Mammograms

Mammography has been shown to be the most effective imaging modality for the early detection of breast cancer.[1] It has been shown that when mammograms are independently read by two radiologists (double reading), an increment of sensitivity of up to 15% can be obtained.[2,3] Because double reading by two radiologists skilled in mammography is difficult to realize in most radiological centers, in recent years several computerized systems have been developed to aid diagnosis by radiologists working in mammography.[4] In such computer-aided diagnosis (CAD) systems, it is of particular interest to detect microcalcifications due to the fact that, in studies analyzing the mammographic nature of missed cancers, clustered microcalcifications composed 19 to 31% of lesions missed in screening.[5–7] In the study by Bird, Wallace, and Yankasas,[5] 18% (six of 33) of the missed malignancies presented as clustered microcalcifications were simply overlooked by the radiologists. Microcalcifications may be inconspicuous, owing to their small size and/or obscuration by overlying fibroglandular tissues, and may be missed even by a diligent radiologist.

Current commercial CAD systems have been designed to include specific procedures for the detection of potential malignancies in the breast, and as such, the detection algorithms are heavily biased toward achieving high sensitivity, thereby sacrificing the specificity of any detection mark placed. This is clearly evident in the analysis performed by Freer and Ulissey,[8] where it was observed that 97.4% of all computer marks placed were dismissed by the radiologist, and, according to their study, the majority of such false positives is related to clustered microcalcifications.

The preceding arguments justify the significant effort that researchers have focused on the detection of microcalcifications. Due to sustained effort, the sensitivity of detection has improved, with some recent papers reporting more than 90% sensitivity. Nevertheless, the false-positive rate remains high, especially when detecting individual microcalcifications.[9]

Mammographic CAD systems are based on either a one-stage or two-stage model.[10] In a one-stage model, after a phase of feature extraction, the entire mammographic im-

age undergoes a classification phase to decide whether a region represents a microcalcification or not. Initial studies on the detection of microcalcifications, which date from the 1970s, are based on a one-stage model: hard decision rules have been applied to features that represent a given mammogram locally, such as local maxima, contrast, and compactness.[11] Methods proposed by Shen, Rangayyan, and Desautels,[12] Strickland and Hahn,[13] and Kim *et al.*[14] are examples of the one-stage model: they extract shape features, features in the wavelet domain, or statistical features, and then classify each region as a calcification or not.

The majority of recent methods for the detection of calcifications in mammograms apply a two-stage model: first, regions of interest (ROIs) are selected. Features are extracted from the ROIs, and then applied to a classifier to decide if the ROIs represent microcalcifications.

Among the existing methods to select ROIs in mammograms, due to the large number of related published papers, we limit our interest to those based on statistics derived from an analysis of the histogram;[15,16] analysis using Markov random fields;[17,18] and multiscale analysis.[19–21] Some recent works utilize mathematical morphology to detect ROIs.[10,22,23]

Cheng *et al.*,[9] in their review on computer-aided detection and classification of microcalcifications, classify the different features extracted from ROIs to detect microcalcifications into ten groups: features of individual microcalcifications;[24–26] co-occurrence features;[27] surround region dependence features;[28] gray-level run-length features;[29] gray-level difference features;[30] wavelet features;[31] Gabor filter bank features;[32] scale-space features;[19] fractal dimension;[33] and cluster features.[34]

The most significant features among a group of computed features are usually obtained via the sequential forward selection method or the sequential backward selection method.[10,35] Some works have employed genetic algorithms for feature selection, trying to maximize the area under the receiver operating characteristics (ROC) curve.[36] In this sense, Verma and Zhang[37] recently published a neural-genetic algorithm to find the most significant combination of features.

Efforts have also been directed toward the development of new classifiers to determine whether an ROI corresponds to a microcalcification or not. In this sense, Hall[38] proposed the use of fuzzy rules instead of hard decision rules. Grohman and Dhawan[39] developed a convex-set-based neuro-fuzzy algorithm for the classification of calcifications. Lee *et al.*[40] used a shape cognitron neural network to analyze the shape of microcalcifications and clusters, and classify them into the malignant or benign category. Tsujii, Freedman, and Mun[41] proposed a trend-oriented radial-basis-function neural network. There are also some interesting works that try to determine the best classifier among a set of classifiers,[42] or try to optimize a neural network to be used as the classifier.[43] In recent years, support vector machines (SVMs) have been utilized as classifiers.[44–46]

We present a two-stage method. First, the candidates for microcalcification are selected by means of a 2-D adaptive filter and a new statistical measure. Then, a feature extraction step is performed, followed by classification using an SVM. Classifiers based on the SVM have been utilized for the detection of microcalcifications.[44–46] The novelty of our
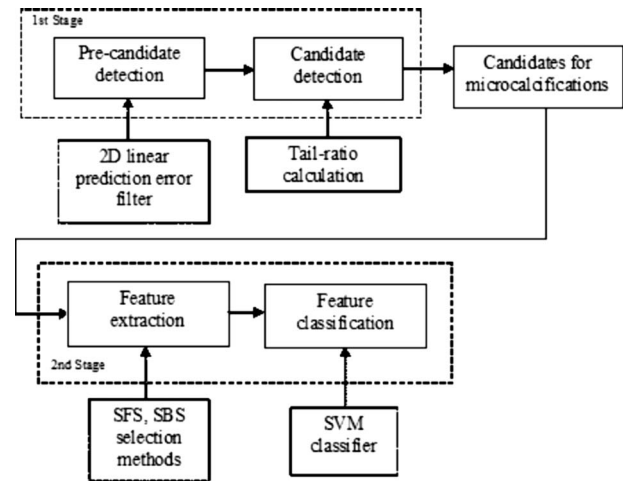
Fig. 1 Scheme of the proposed algorithm for the detection of microcalcifications in mammograms. SBS is sequential backward selection. SFS is sequential forward selection. SVM is support vector machine.

proposed method is not in the use of the SVM per se, but in the features employed to classify with the SVM. Gurcan *et al.*[47] used high-order statistics, including skewness and kurtosis. In the present work, we propose a new statistical feature called the tail ratio. Similar to the method of Gurcan, Yardimci, and Cetin,[48] the method proposed in this work utilizes the prediction error of an adaptive filter to detect microcalcifications; however, Gurcan *et al.* used the test of Gaussianity to classify calcifications, whereas we use an SVM. Furthermore, the algorithm for adaptive linear prediction to compute the prediction error used in the present work is different from that used by Gurcan, Yardimci, and Cetin,[48] as described in the following sections.

2 Methods: Detection of Microcalcifications in Mammograms

Our algorithm to detect microcalcifications follows the scheme shown in Fig. 1. The algorithm is divided into two stages. The first stage focuses on the detection of candidates for microcalcifications, and is divided into two parts: precandidate detection and candidate detection. The second stage consists of deciding whether a candidate is a microcalcification or not. This stage is performed in two steps: a feature extraction procedure and the final classification step. In the following sections, each part of the scheme in Fig. 1 is explained in detail.

2.1 First Stage of the Algorithm

2.1.1 Detection of precandidates

As a first step in the algorithm, a selection of *precandidates* for microcalcifications is performed. For this purpose, a 2-D linear prediction error filter[49,50] is applied, followed by a thresholding step. A pixel is selected as a precandidate for microcalcification if its prediction error is greater than an adaptively determined threshold.[51] This is based on the observation that a microcalcification can be seen as a point of nonstationarity in an approximately homogeneous region or neighborhood in a mammogram. Such a pixel cannot be
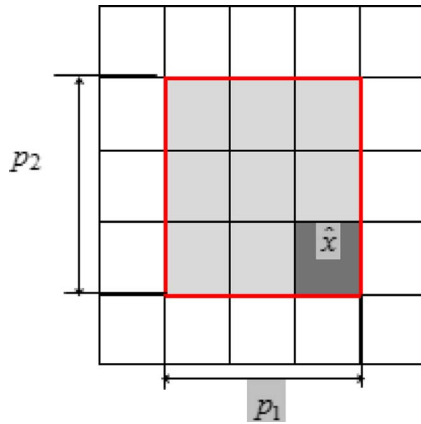
Fig. 2 Region of support (ROS) of size $p_1 \times p_2$ for the predicted pixel $\hat{x}(m,n)$.

Fig. 3 Initialization of prediction error vectors with the original image pixels, in an $M \times N$ subimage for a 3×3 predictor. Each vector $e_0^f(k) = e_0^b(k)$ (for $k = 0, \ldots$), of size $p_2 \times 1$, is initialized to the corresponding pixel values of the original image.

predicted well by the linear predictor, and hence leads to a high prediction error.

The given image is divided into $M \times N$ subimages that are assumed to be approximately homogeneous. The adaptive 2-D linear predictor represents each $M \times N$ subimage as a 2-D sequence $x(m,n)$, with $\{m,n\} = \{0, 1, \ldots, M-1; 0, 1, \ldots, N-1\}$. The basic assumption in linear predictive coding is that any pixel in an image may be predicted as a linear combination of a few neighboring pixels. The predicted value of a pixel $x(m,n)$ at (m,n) is given by

$$\hat{x}(m,n) = \sum_{(i,j) \in \text{ROS}} \sum a(i,j)x(m-i,n-j), \tag{1}$$

where $a(i,j)$ are known as the prediction coefficients and ROS represents the region of support of the predictor, as shown in Fig. 2. The prediction error is

$$e(m,n) = x(m,n) - \hat{x}(m,n), \tag{2}$$

and the mean-squared prediction error is

$$\varepsilon^2 = E[e^2(m,n)], \tag{3}$$

where $E[\cdot]$ is the expectation operator. The prediction error values tend to have a smaller dynamic range and a more nonuniform distribution than the original pixel values.[50]

The optimal prediction coefficients that minimize the mean-squared error are obtained by solving the 2-D Yule-Walker equations:

$$\sum_{(i,j) \in \text{ROS}} \sum a(i,j)r_{xx}(k-i,l-j) = \begin{cases} 0 & (k,l) \in \text{ROS} \\ \varepsilon^2 & (k,l) = (0,0) \end{cases}, \tag{4}$$

where $r_{xx}(k,l)$ is the 2-D autocorrelation function of the subimage, and ε^2 is the mean-squared prediction error.

The computation of the prediction coefficients from this relationship requires an estimate of the autocorrelation function of the image.[52,53] However, the coefficients may be obtained directly from the data with the multichannel Burg algorithm.[49] In addition to computing the 2-D linear prediction coefficients, the Burg algorithm may be used to compute the prediction errors directly from the data. This is advantageous in the present application, because the error
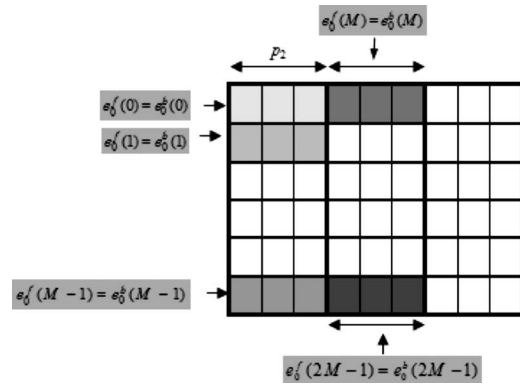
values provide the information required to select precandidates for microcalcifications (the actual prediction model is not required).

The multichannel version of the Burg algorithm calculates the optimal prediction coefficients for a $p_1 \times p_2$ 2-D predictor by computing the prediction errors of order p_1. The prediction errors are represented as vectors of size $p_2 \times 1$ and are computed recursively, beginning with the zero-order prediction errors. The error values are initialized to the original image, as shown in Fig. 3. The prediction errors of higher orders are calculated by repeating recursively the following three steps:[49–51]

1. Compute the covariance matrices for the forward and backward prediction errors.
2. Calculate the prediction coefficient matrix.
3. Compute the backward and forward prediction error vectors of the higher prediction order.

Once we know the prediction error values, we use them as a measure of the stationarity of the image at the corresponding points. Pixels where the prediction error is higher than a dynamically determined threshold are considered to be points of nonstationarity, and therefore, precandidates for microcalcifications. The threshold is adaptively determined based on the local average intensity of the subimage, local average value of the prediction error, and the average intensity of the whole image.

2.1.2 Detection of candidates

Microcalcifications represent small points of high intensity. Therefore, if we analyze the probability density function (PDF) of the pixels belonging to a neighborhood surrounding a microcalcification, it is expected to have the right tail (higher gray levels) longer than the left one. Based on this observation, we propose a parameter called the tail-ratio, TR, which is a relative measure of the tail length of a PDF. There are some descriptors in the literature that measure the heaviness of the tail of a PDF. The kurtosis coefficient is often regarded as a measure of the heaviness of a distribution relative to the normal distribution; however, its interpretation and use have been restricted to symmetric distributions, because of its intrinsic comparison with the

symmetric normal distribution. Another disadvantage with kurtosis is that it is sensitive to outliers in the data, because it is based on moments of the data. For any PDF $f(x)$ with finite moments, kurtosis is defined as:

$$k = \frac{E_F[X - E_F(X)]^4}{\{E_F[X - E_F(X)]^2\}^2}, \tag{5}$$

where the numerator and the denominator represent the second and the fourth central moments of F, respectively, and X are the data values.

Some authors have presented other robust measures of kurtosis, but defined only for symmetric distributions, or merely measuring the peakedness instead of the tail weight.[54,55]

Other works overcome the problems mentioned by introducing several measures of tail weight for univariate continuous distributions that can be applied to symmetric as well as asymmetric distributions.[56] They define left-tail and right-tail measures as measures of skewness that are applied to the half-portion of the probability mass lying to the left or the right, respectively, of the median of the distribution, denoted as $m_F = F^{-1}(0.5)$, where $F(x)$ is the distribution function. Nevertheless, we have not found these measures capable of solving the problem of detecting microcalcifications, because they are estimates of the heaviness of the tail but not its length. Microcalcifications, as described before, are characterized by a long right tail in the local histogram. Therefore, to characterize the right tail, we have developed the measure TR, defined as:

$$\text{TR} = \frac{\max(X) - F^{-1}(0.5)}{F^{-1}(0.5) - \min(X)}, \tag{6}$$

where $\max(X)$ and $\min(X)$ represent the maximum and minimum values of X, respectively, and F^{-1} is the inverse of the distribution function.

Because the size of a microcalcification is variable, the neighborhood around the precandidates to calculate the local histogram should be adaptive. According to the size of a microcalcification, with the diameter varying from 0.1 to 1 mm, and according to the resolution of the database of images to be used (50 μm in the present work), microcalcifications can occupy from 3.2 to 314 pixels. Therefore, we initially calculate the histogram and the parameter TR for a 3×3-pixel square around each precandidate. If TR is above the applicable threshold, the precandidate is considered to be a candidate for the subsequent steps for the detection of microcalcifications. Otherwise, the square box is increased in size, and TR is recalculated. This procedure is repeated until one of the following two possible conditions is fulfilled: 1. the selection box reaches its maximal area (specified as 20×20 pixels in the present work), or 2. TR is higher than the applicable threshold. This procedure is summarized in Fig. 4. After this step, the candidates are prepared for subsequent classification as a microcalcification or not.

2.2 Second Stage of the Algorithm

2.2.1 Feature extraction

For each candidate for microcalcification, statistical texture features are extracted that, after a selection procedure, will
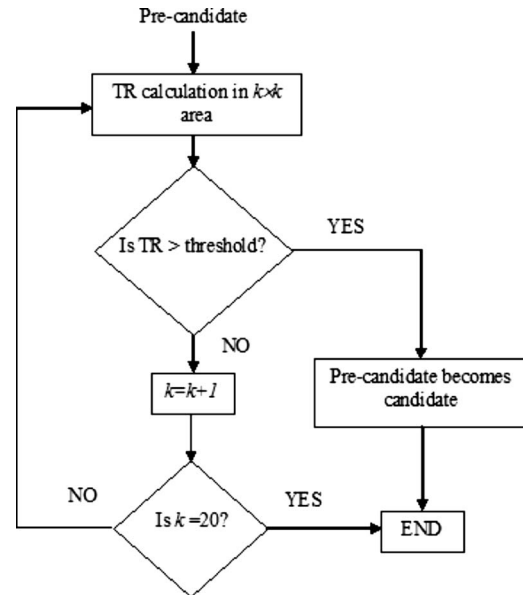

Fig. 4 Scheme to select the candidates for microcalcifications.

be the inputs to a classifier. To calculate these statistical features we have approximated the PDF with a normalized histogram, with bin size equal to 1. Specifically, the descriptors chosen for each candidate are:

- Size of the square neighborhood (k) where TR exceeded the threshold.
- TR value.
- Interdistance ID: the parameter ID is based on the observation that pixels that have high intensity values (above the 95th percentile) will belong to a microcalcification if they are close to one another. Therefore, we define a parameter ID calculated as:

$$\text{ID} = \frac{1}{n} \sum_{i=1}^{N} [(x_i - x_c)^2 + (y_i - y_c)^2]^{1/2}, \tag{7}$$

where N is the number of pixels above the 95'th percentile, (x_i, y_i) are the coordinates of the pixels selected, and (x_c, y_c) are the coordinates of the centroid of the selected pixels. The smallest neighborhood is 9 pixels. This will be the case for tiny microcalcifications. In such a case, there will be only 1 pixel above percentile 95. Therefore, ID=0. When microcalcifications are bigger, the neighborhood size grows. Then, the number of pixels above the percentile 95 will be higher. To maintain ID low, it is necessary for those pixels to be near one another. In other words, the highest values cannot be dispersed but together, constituting the center of a microcalcification.

- Average of the mean slopes (MS): in the four possible directions (North, South, East, and West), the mean descending slope from the pixel with the maximum value in the neighborhood $k \times k$ is calculated, and MS is obtained as the average of the four values, according to the equation

$$\text{MS} = \frac{1}{4} \cdot \frac{1}{(k/6)+1} \left\{ \sum_{n=n\,\max}^{n\,\max - k/6} [x(n-1,m) - x(n,m)] \right.$$

$$+ \sum_{n=n\,\max}^{n\,\max + k/6} [x(n+1,m) - x(n,m)]$$

$$+ \sum_{m=m\,\max}^{m\,\max - k/6} [x(n,m-1) - x(n,m)]$$

$$\left. + \sum_{m=m\,\max}^{m\,\max + k/6} [x(n,m+1) - x(n,m)] \right\}, \quad (8)$$

where $m\,\max$ and $n\,\max$ are the coordinates of the pixel with the maximum value within the neighborhood. The interval we have chosen to average the slope is $k/6$, because the estimated diameter of the microcalcification is $k/3$ in the $k \times k$ square neighborhood.

- Average of maximum slopes (max S): this feature represents the average of the maximum slopes in the North, South, East, and West directions of a candidate according to

$$\max S = \frac{1}{4} (\max_{n\,\max \geqslant n \geqslant n\,\max - k/6} [x(n-1,m) - x(n,m)]$$

$$+ \max_{n\,\max \leqslant n \leqslant n\,\max + k/6} [x(n+1,m) - x(n,m)]$$

$$+ \max_{m\,\max \geqslant m \geqslant m\,\max - k/6} [x(n,m-1) - x(n,m)]$$

$$+ \max_{m\,\max \leqslant m \leqslant m\,\max + k/6} [x(n,m+1) - x(n,m)]).$$

$$(9)$$

- Entropy (Ent), defined as $\text{Ent} = -\sum_{j=0}^{N-1} P(j) \log_2 [P(j)]$, where N represents the number of gray levels in the image and $P(j)$ is the probability of occurrence of gray level j.
- Average height (AH) of the histogram within the $k \times k$ neighborhood: this parameter is defined as

$$\text{AH} = \frac{\sum_{j=0}^{N-1} h(j)}{[\max(X) - \min(X)]},$$

where h represents the histogram of the data distribution X within the $k \times k$ neighborhood.

- Correlation with a Gaussian distribution (CG) with standard deviation equal to $k/6$: although microcalcifications vary in form, it is considered to be a plausible assumption that they have a circularly symmetric Gaussian distribution.[13,57] Therefore, if a microcalcification is present in the square region being analyzed, the correlation evaluated between the Gaussian and the image centered at the maximum within the square will be close to 1 if both signals have their energy normalized to 1.
- Contrast parameter (C), calculated as $\text{C} = \text{mean}_k - m / \text{mean}_k + m$, where mean_k is the average value of the pixels within the $k \times k$ square neighborhood, and m represents the mean value of the pixels belonging to the two-pixel-wide border of the square.

- Dynamic range (DR), obtained as $\text{DR} = \max(X) - \min(X)$, where X represents the image values within the $k \times k$ square neighborhood.

2.2.2 *Feature selection*

Not all of the ten features defined in Sec. 2.2.1 are used in the subsequent classification step: a larger number of features does not necessarily lead to better classification performance. Hence, a feature selection procedure is applied. In particular, we use the *sequential forward selection* (SFS) and *sequential backward selection* (SBS) methods[58] to obtain the subset of features that gives the minimum classification error. The classifier used is an SVM, which is explained in the next section.

SFS is a bottom-up search procedure where one feature at a time is added to the current feature set. At each stage, the feature to be included in the feature set is selected from the remaining available features that have not been added to the feature set yet, such that the new enlarged feature set yields a lower classification error as compared to the addition of any other single feature. The algorithm stops when the addition of a new feature leads to an increase in the classification error.

SBS is the top-down counterpart of the SFS method. It starts with the complete set of features, and at each stage, the feature that shows the least discriminative power is discarded. The algorithm stops when removing another feature leads to an increase in the classification error.

To analyze the two feature selection methods and to determine the optimal feature set, we used five mammograms and selected 230 microcalcifications and 400 points that were not microcalcifications. These five mammograms were not included in the experiments to obtain the final classification results presented in Sec. 4. The 400 points were all precandidates, that is, they had a prediction error higher than the applicable threshold. Of the selected points, 184 microcalcifications and 320 nonmicrocalcifications (that is, 80% of the data) were used as the training set for the classifier, and the remaining 46 microcalcifications and 80 nonmicrocalcifications were used as the validation set where the classification error is measured. The selection performance was evaluated by five-fold cross-validation (XVAL),[58] where the procedure is repeated five times, changing each time the training and the validation sets, and the average results are computed. In this manner, the disadvantage of the sensitivity of the SBS and SFS methods to the order of presentation of the training set is diminished.[58] To implement the SBS and SFS methods, a classifier is required; for this purpose, an SVM was used, as explained next.

The average error was calculated by counting the misclassifications and dividing by the total number of microcalcifications. SBS resulted in the feature set [tail ratio (TR), average height (AH), correlation with a Gaussian distribution (CG), contrast (C), and dynamic range (DR)] with an average error of 4.92%. SFS led to the best feature set [mean slope (MS), correlation with a Gaussian distribution (CG), contrast (C), dynamic range (DR), average height (AH), and interdistance (ID)], with an average error of 4.13%. By analyzing the aggregation order that features followed in the SFS method and how much the classifica-
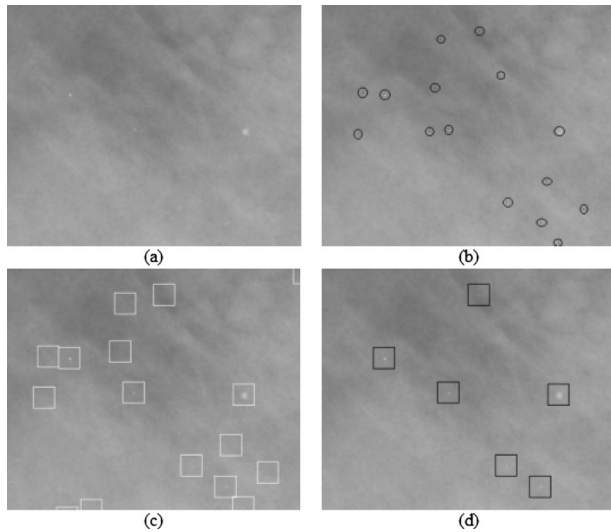
Fig. 5 Example of the results generated by the proposed algorithm at each step. (a) 415×342-pixel portion of an original mammogram. (b) Precandidates shown as black points surrounded by a circle. (c) Candidates (selected by the tail-ratio parameter) marked by white squares. (d) Final detected microcalcifications marked by black squares.

tion error decreases when adding a new feature, one can know the discriminative power of each feature. This order and error percentage is: mean slope (MS) 7.61%, correlation with a Gaussian distribution (CG) 6.34%, dynamic range (DR) 5.08%, contrast (C) 4.76%, average height (AH) 4.44%, and interdistance (ID) 4.13%.

2.2.3 *Feature classification*

The final step consists of the classification of each candidate as a microcalcification or not. For this purpose, a vector with the six features selected by SFS is prepared for each candidate pixel. This vector is applied as the input to an SVM classifier, whose outputs are "microcalcification" and "nonmicrocalcification."

Support vector algorithms[59] constitute one of the important advances in computational learning in the 1990s. The SVM is the final step in a long research study known as statistical learning, carried out mainly by Vapnik.[60] Vector machine formulation is based on the principle of structural risk minimization (SRM), which has been shown to be better than the principle of empirical risk minimization. Traditional classifiers, such as neural networks, suffer from the generalization ability problem, and thus can lead to overfitting, as a consequence of trying to minimize the empirical risk. The main idea of SVMs is to separate the classes with a surface that maximizes the margin between them. This is an approximate implementation of SRM.

The idea is to map the training points into a high-dimensional feature space where a separating hyperplane (w, b), with w as the normal and b as the bias to the hyperplane, can be found that maximizes the margin or distance to the closest data points. The closest data points are called support vectors. The optimal separating hyperplane can be represented as

Table 1 Average number of precandidates, candidates, and detected microcalcifications.

Number of precandidates	Number of candidates	Number of detected microcalcifications
1399	194	72

$$f(x) = \text{sign}\left[\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b \right], \quad (10)$$

where n is the number of training examples, y_i is the label value of example i, K represents the kernel, and the α_i coefficients must be found in a way to maximize a particular Lagrangian representation. Only those training examples that lie on the decision boundary, that is, the support vectors, have nonzero α_i.

If we admit that some vectors are misclassified, the optimal hyperplane equation can be defined as:

$$y_i(wx_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0. \quad (11)$$

Hence, to obtain the optimal hyperplane, the function that should be minimized is

$$\|w\|^2 + D \sum_{i=1}^{n} \xi_i, \quad (12)$$

where ξ represents a measure of the amount of violation of the constraints, and D is a parameter that controls the tradeoff between maximizing the margin and minimizing the training error.

There are no techniques available to select the kernel. In a large variety of applications, Gaussian kernels lead to
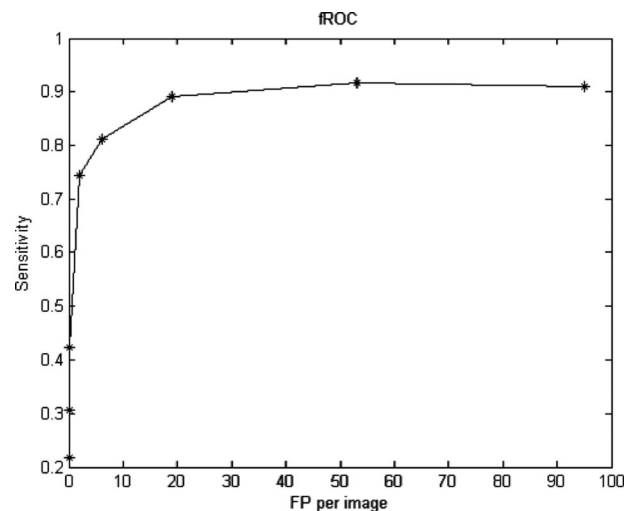
Fig. 6 Free-response receiver operating characteristics (fROC) curve with eight different values of the tail-ratio threshold (TR $=1.0, 1.5, 2.0, 2.5, 3.0, 6.0, 8.0, 10.0$) for individual microcalcification detection. FP is the false-positive individual microcalcifications per image.

Table 2 Sensitivity, specificity, and positive predictive values (PPV) for different tail-ratio (TR) thresholds.

	TR threshold							
	1.0	1.5	2.0	2.5	3.0	6.0	8.0	10.0
Sensitivity	0.91	0.92	0.89	0.81	0.75	0.42	0.31	0.22
Specificity	0.94	0.97	0.99	1.00	1.00	1.00	1.00	1.00
PPV	0.45	0.59	0.79	0.92	0.96	1.00	1.00	1.00

good results.[59] Thus, in this work, the kernel used is a Gaussian function with variance equal to 2.0.

The SVM was trained with the same dataset (five mammograms) that was used in the feature selection step.

3 Materials

A database of 45 digitized mammograms was obtained from Screen Test: The Alberta Program for the Early Detection of Breast Cancer.[61]. They were digitized with a Lumiscan 85 laser scanner (Lumisys, Sunnyvale, California) with a spatial resolution of 50 μm and 12 bits per pixel. The proposed algorithm was tested with 40 digitized mammograms from this database. The other five mamograms were discarded because microcalcifications from them were used to train the SVM. The remaining 40 images contain an average of 95 microcalcifications per image of different nature and diagnosis. The average area of the microcalcifications is 0.017 mm^2, the standard deviation is 0.0056 mm^2, the largest microcalcification has an area of 0.5625 mm^2, and the smallest one has an area equal to 0.01 mm^2. The results obtained with the proposed algorithms were examined by an expert radiologist specialized in mammography (Desavtels), to determine the accuracy of detection.

4 Results

Figure 5 shows examples of the results generated by the proposed algorithm at each step. Figure 5(a) shows a 415 ×342-pixel portion of an original mammogram. In Fig. 5(b), the precandidates obtained with the 2-D prediction error filter can be seen as small black points. In Fig. 5(c), we show the candidates (selected by the tail-ratio parameter) marked by white squares. In Fig. 5(d), the final result of the detection of microcalcifications is shown. For a further illustration of the whole procedure, in Table 1 the number of precandidates, candidates, and detected microcalcifications are shown.

We have preformed two different analyses on the dataset: in the first one, individual microcalcifications are detected, and in the second one, cluster analysis is performed.

For the individual microcalcification analysis, a free-response receiver operating characteristics (fROC) curve was generated, representing the sensitivity values versus the number of false positives per image. Eight points were calculated for the fROC curve, varying the threshold on the tail-ratio parameter as [1.0, 1.5, 2.0, 2.5, 3.0, 6.0, 8.0, 10.0]. The numbers of true and false positives were calculated by counting each individual microcalcification and not clus-

ters. We consider as true negatives those pixels that are precandidates but the algorithm has rejected them. The fROC curve obtained is shown in Fig. 6. In Table 2, the values obtained for sensitivity, specificity, and positive predictive value (PPV) are presented. The best performance was obtained with the TR threshold of 2.0, where a sensitivity value of 0.89, a specificity value of 0.99, and a PPV of 0.79 were obtained. Detection sensitivity values of 0.75 and 0.81 were obtained at 2.6 and 6.2 false positives per image, on the average, respectively. The computational cost of the algorithm, when run on a Pentium 4 at 3.06 GHz with 1-GB DDR, was a processing time of 25 s per mammogram, on the average.

To do the cluster analysis, we have considered as a cluster three or more microcalcifications within a cm^2. In cluster analysis, another fROC curve was generated, representing the sensitivity values versus the number of false-positive clusters per image. Six points were calculated for the fROC curve, varying the threshold on the tail-ratio parameter as [1.5, 2.0, 2.3, 2.5, 3.0, 6.0]. The TR value of 1.5 instead of 1.0 was chosen as the lowest limit, because for TR=1.0, the same sensitivity as for TR=1.5 was obtained, but with more false positives. Tail ratio equal to 6.0 was chosen as the highest limit, because no false positive was obtained for this value. The numbers of true and false posi-
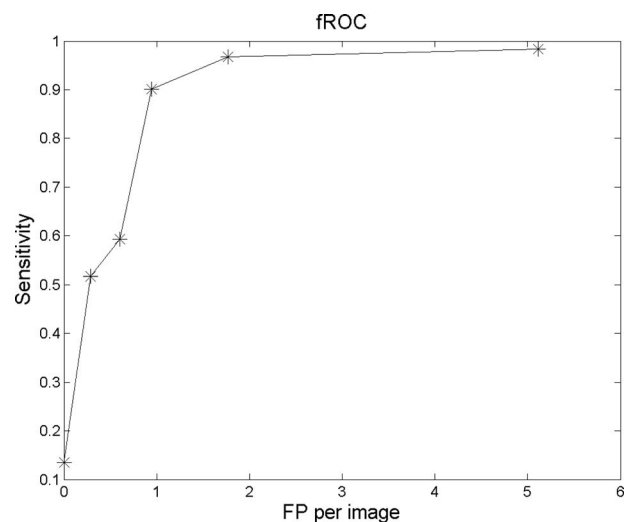
Fig. 7 Free-response receiver operating characteristics (fROC) curve with five different values of the tail-ratio threshold (TR =1.5,2.0,2.3,2.5,3.0,6.0) for cluster detection. FP is the false-positive clusters per image.

Table 3 True positives per image (TP), false positives per image (FP), false negatives per image (FN), sensitivity, and positive predictive value (PPV) for different tail-ratio (TR) thresholds.

	TR threshold					
	1.5	2.0	2.3	2.5	3.0	6.0
TP	1.71	1.69	1.57	1.00	0.89	0.23
FP	5.11	1.77	0.94	0.60	0.29	0.00
FN	0.03	0.06	0.17	0.69	0.83	1.46
Sensitivity	0.98	0.97	0.90	0.57	0.51	0.13
PPV	0.25	0.49	0.62	0.63	0.76	1.00

tives were calculated by counting clusters and not individual microcalcifications. The fROC curve is shown in Fig. 7. In Table 3, the values obtained for true positives per image, false positives per image, false negatives per image, sensitivity, and PPV are presented. The best performance was obtained with the TR threshold of 2.0, where a sensitivity value of 0.97 was obtained at 1.77 false positives per image, or with TR=2.3, where a sensitivity value of 0.90 was obtained at 0.94 false positives per image The best sensitivity value is 0.98, obtained with TR=1.5, at 5.11 false positives per image.

The method proposed in this work was tested independently by Martínez-Álvarez *et al.*[62] with a different dataset, including 94 images with an average of 111 microcalcifications per image. A sensitivity of 95.87% was obtained at the rate of 2.3 false positives per image for individual microcalcification detection.

5 Conclusions

We propose a method for the detection of microcalcifications in mammograms. The algorithm is divided into two stages. In the first stage, candidates for microcalcifications are obtained based on a 2-D linear prediction error filter and the calculation of a new statistical measure named tail ratio. The second stage consists of a feature extraction step and a subsequent classification task. The feature vector, after a selection procedure, is formed by six descriptors, which form the input to an SVM classifier, whose output gives the detected microcalcifications.

40 mammograms are used to validate the algorithm in this work. We perform two different analyses on the dataset: individual microcalcification detection and cluster analysis. The best performance obtained for individual microcalcification detection is a sensitivity value of 0.89, a specificity value of 0.99, and a PPV of 0.79. For cluster analysis, a sensitivity value of 0.97 is obtained at 1.77 false positives per image, and a sensitivity value of 0.90 is achieved at 0.94 false positives per image. This result is comparable with the results provided by other methods published in the literature that have achieved sensitivity of about 90% at a false-positive rate that varies from 0.5 to 3 false-positive clusters per image.[9] Nevertheless, for the range less than one false positive per image, our method should be improved to become closer to the best methods present in the literature.[63] In any case, it is important to note that this comparison is made with different databases.

References

1. S. A. Feig and R. E. Hendrick, "Risk, benefit and controversy in mammographic screening," *79th Scientific Assembly and Annual Meeting of RSNA*, A. G. Haus and M. Yaffe, eds., pp. 21–23 (1993).
2. E. Thurfjell, K. A. Lernevall, and A. Taube, "Benefit of independent double reading in a population-based mammography screening program," *Radiology* **191**, 241–244 (1994).
3. I. Anttinen, M. Pamilo, M. Soiva, and M. Roiha, "Double reading of mammography screening films- one radiologist or two?" *Clin. Radiol.* **48**, 414–421 (1993).
4. A. Lauria, R. Palmiero, G. Forni, M. E. Fantacci, M. Imbriaco, A. Sodano, and P. L. Indovina, "A study on two different CAD systems for mammography as an aid to radiological diagnosis in the search of microcalcification clusters," *Eur. J. Radiol.* **55**, 264–269 (2005).
5. R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613–617 (1992).
6. J. A. Harvey, L. L. Fajardo, and C. A. Innis, "Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation," *Am. J. Roentgenol.* **161**, 1167–1172 (1993).
7. L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "The potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* **215**, 554–562 (2000).
8. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: prospective study of, 12,860 patients in a community breast center," *Radiology* **220**(3), 781–786 (2001).
9. H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern Recogn.* **36**, 2967–2991 (2003).
10. J. C. Fu, S. K. Lee, S. T. C. Wong, J. Y. Yeh, A. H. Wang, and H. K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," *Comput. Med. Imaging Graph.* **29**, 419–429 (2005).
11. W. Spiesberger, "Mammogram inspection by computer," *IEEE Trans. Biomed. Eng.* **26**(4), 213–219 (1979).
12. L. Shen, R. M. Rangayyan, and J. E. L. Desautels, "Detection and classification of mammographic calcifications," *Int. J. Pattern Recognit. Artif. Intell.* **7**(2), 1403–1416 (1993).
13. R. N. Strickland and H. I. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms," *IEEE Trans. Med. Imaging* **15**(2), 218–229 (1996).
14. J. K. Kim, J. M. Park, K. S. Song, and H. W. Park, "Adaptive mammographic image enhancement using first derivative and local statistics," *IEEE Trans. Med. Imaging* **16**, 495–502 (1997).
15. M. Kallergi, K. Woods, L. P. Clarke, W. Qian, and R. A. Clark, "Image segmentation in digital mammography: comparison of local thresholding and region growing algorithms," *Comput. Med. Imaging Graph.* **16**, 323–331 (1992).

16. K. S. Woods, C. C. Doss, K. W. Bowyer, J. L. Solka, C. E. Priebe, and W. P. Kegelmeyer, "Comparative evaluation of pattern recognition techniques for detection of microcalcifications in mammography," *Int. J. Pattern Recognit. Artif. Intell.* **7**, 1417–1436 (1993).

17. N. Karssemeijer, "A stochastic model for automated detection of calcifications in digital mammograms," *12th Intl. Conf. IPMI*, pp. 227–238 (1992).

18. B. Caputo, E. La Torre, S. Bouattour, and G. E. Gigante, "A new kernel method for microcalcification detection: Spin-Glass-Markov random fields," *Stud. Health Technol. Inform.* **90**, 30–34 (2002).

19. T. Netsch and H. O. Peitgen. "Scale-space signatures for the detection of clustered microcalcifications in digital mammograms," *IEEE Trans. Med. Imaging* **18**(9), 774–786 (1999).

20. S. Dippel, M. Stahl, R. Wiemker, and T. Blaffert, "Multiscale contrast enhancement for radiographies: Laplacian pyramid versus fast wavelet transform," *IEEE Trans. Med. Imaging* **21**(4), 343–353 (2002).

21. M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters," *Med. Phys.* **31**(2), 314–326 (2004).

22. H. Takeo, K. Shimura, T. Imamura, A. Shimizu, and H. Kobatake, "Detection system of clustered microcalcifications on CR mammogram," *IEICE Trans. Inf. Syst.* **e88-d**(11), 2591–2602 (2005).

23. S. Halkiotisa, T. Botsisa, and M. Rangoussi, "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks," *Signal Process.* **87**, 1559–1568 (2007).

24. S. Bothorel, B. B. Meunier, and S. Muller, "A fuzzy logic based approach for semiological analysis of microcalcifications in mammographic images," *Int. J. Intell. Syst.* **12**, 819–848 (1997).

25. I. N. Bankman, J. Tsai, D. W. Kim, O. B. Gatewood, and W. R. Brody, "Detection of microcalcification clusters using neural networks," *Proc. 16th Annual Int1. Conf. IEEE EMBS* **1**, 590–591 (1994).

26. L. Zhang, R. Sankar, and W. Qian, "Advances in micro-calcification clusters detection in mammography," *Comput. Biol. Med.* **31**, 515–528 (2002).

27. D. Kramer and F. Aghdasi, "Texture analysis techniques for the classification of microcalcifications in digitized mammograms," *Proc. IEEE Africon* **1**, 395–400 (1999).

28. C. S. Lee, J. K. Kim, and H. W. Park, "Computer-aided diagnostic system for breast cancer by detecting microcalcification," *Proc. SPIE* **3335**, 615–626 (1998).

29. J. K. Kim and H. W. Park, "Statistical textural features for detection of microcalcifications in digitized mammograms," *IEEE Trans. Med. Imaging* **18**(3), 231–238 (1999).

30. J. S. Geronimo, D. P. Hardin, and P. R. Massopust, "Fractal functions and wavelet expansions based on several scaling functions," *J. Approx. Theory* **78**, 373–401 (1994).

31. A. P. Dhawan, Y. S. Chitre, C. Bonasso, and M. Moskowitz, "Analysis of mammographic microcalcifications using gray-level image structure features," *IEEE Trans. Med. Imaging* **15**(3), 245–259 (1996).

32. T. Bhangale, U. B. Desai, and U. Sharma, "An unsupervised scheme for detection of microcalcifications on mammograms," *IEEE Intl. Conf. Image Process.*, pp. 184–187 (2000).

33. G. L. Rogova, P. C. Stomper, and C. C. Ke, "Microcalcification texture analysis in a hybrid system for computer aided mammography," *Proc. SPIE* **3661**, 1426–1433 (1999).

34. L. P. Cordella, F. Tortorella, and M. Vento, "Combing experts with different features for classifying clustered microcalcifications in mammograms," *Proce. 15th Intl. Conf. on Patt. Recog.*, pp. 324–327 (2000).

35. W. J. H. Veldkamp and N. Karssemeijer, "Automated classification of clustered microcalcifications into malignant and benign types," *Med. Phys.* **27**(11), 2600–2608 (2000).

36. H. Soltanian-Zadeh, S. Pourabdollah-Nezhad, and F. Rafiee-Rad, "Shape-based and texture-based feature extraction for classification of microcalcifications in mammograms," *Proc. SPIE* **4322**, 301–310 (2001).

37. B. Verma and P. Zhang, "A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms," *Appl. Soft Comput.* **7**, 612–625 (2007).

38. L. O. Hall, "Learned fuzzy rules versus decision trees in classifying microcalcifications in mammograms," *Proc. SPIE* **2761**, 54–61 (1996).

39. W. M. Grohman and A. P. Dhawan, "Classification of difficult-to-diagnose microcalcifications using fuzzy neural network with convex sets," *Proc. First Joint BMES/EMBS Conf. Serving Humanity, Adv. Technol.*, pp. 13–16 (1999).

40. S. K. Lee, P. C. Cheng, C. I. Chang, C. S. Lo, T. Lee, G. C. Hsu, and C. W. Yang, "A shape cognitron neural network for breast cancer detection," *Proc. Intl. Joint Conf. Neural Net.* **1**, 822–827 (2002).

41. O. Tsujii, M. T. Freedman, and S. K. Mun, "Classification of microcalcifications in digital mammograms using trend-oriented radial basis function neural network," *Pattern Recogn.* **32**, 891–903 (1999).

42. L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Trans. Med. Imaging* **24**(3), 371–380 (2005).

43. M. N. Gurcan, B. Sahiner, H. P. Chan, L. Hadjiiski, and N. Petrick, "Optimal selection of neural network architecture for CAD using simulated annealing," *Proc. 22nd Ann. Intl. Conf. IEEE EMBS*, pp. 3052–3055 (2000).

44. S. Singh, V. Kumar, H. K. Verma, and D. Singh, "SVM based system for classification of microcalcifications in digital mammograms," *Proc. 28th Ann. Intl. Conf. IEEE EMBS*, pp. 4747–4750 (2006).

45. A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani, "An SVM classifier to separate false signals from microcalcifications in digital mammograms," *Phys. Med. Biol.* **46**, 1651–1663 (2001).

46. A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artif. Intell. Med.* **34**, 141–150 (2005).

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

48. M. N. Gurcan, Y. Yardimci, and A. E. Cetin, "Microcalcification detection using adaptive filtering and gaussianity tests," *Proc. 4th Intl. Workshop Digital Mammography*, pp. 157–164, Kluwer Academic, New York (1998).

49. G. R. Kuduvalli and R. M. Rangayyan, "An algorithm for direct computation of 2-d linear prediction coefficients," *IEEE Trans. Signal Process.* **41**(2), 996–999 (1993).

50. G. R. Kuduvalli and R. M. Rangayyan, "Performance analysis of reversible image compression techniques for high-resolution digital teleradiology," *IEEE Trans. Med. Imaging* **11**(3), 430–445 (1992).

51. B. Acha, C. Serrano, R. Rangayyan, and J. E. L. Desautels, "Detection of microcalcifications in mammograms," in *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*, J. S. Suri and R. M. Rangayyan, Eds., pp. 291–314, SPIE Press, Bellingham, WA (2006).

52. C. Serrano, J. Díaz, B. Acha, and R. M. Rangayyan, "Use of linear prediction error to detect microcalcifications in mammograms," *Proc. II Latinoamerican Conf. Biomed. Eng.* **1**, pp. 72–75 (2001).

53. B. Acha, C. Serrano, and R. M. Rangayyan, "Detection of microcalcifications in mammograms by seed selection and multitolerance region growing," *Proc. European Med. Biol. Eng. Conf. (EMBEC'99)* **37**(2), 984–985 (1999).

54. R. A. Groeneveld, "A class of quantile measures for kurtosis," *Statistician* **33**, 391–399 (1984).

55. F. Schmid and M. Trede, "Simple tests for peakedness, fat tails and leptokurtosis based on quantiles," *Comput. Stat. Data Anal.* **43**, 1–12 (2003).

56. G. Brys, M. Hubert, and A. Struyf, "Robust measures of tail weight," *Comput. Stat. Data Anal.* **50**(3), 733–759 (2005).

57. L. Bocchi, G. Coppini, J. Nori, and G. Valli, "Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks," *Med. Eng. Phys.* **26**, 303–312 (2004).

58. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Morgan Kaufmann (Academic Press), San Diego, CA (1990).

59. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.* **2**(2), 121–167 (1998).

60. V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York (1995).

61. Alberta Cancer Board, "Screen Test: The Alberta Program for the Early Detection of Breast Cancer," 2001/03 Biennial Report, Edmonton, Alberta, Canada (2004).

62. F. Martínez-Álvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar-Ruiz, "Detection of microcalcifications in mammographies based on linear pixel prediction and support-vector machines," *Proc. 20th IEEE Intl. Symp. Computer-Based Med. Syst.*, pp. 141–146 (2007).

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

Begoña Acha received her PhD in telecommunication engineering in July 2002. She has been working as associate professor since 1996 in the Signal Processing and Communications Department of the University of Seville, Spain. Her current research activities include work in the field of color image processing and its medical applications, in particular in segmentation, classification, and compression. She is the author of numerous papers both in journals and international conferences.

Carmen Serrano received her MS in telecommunication engineering from the University of Seville, Spain, in 1996 and her PhD in January 2002. In 1996, she joined the Signal Processing and Communication Department at the same university, where she is currently an associate professor. Her research interests concern image processing and, in particular, color image segmentation, classification, and compression, mainly with biomedical applications. She is the author of numerous papers both in journals and international conferences.

Rangaraj M. Rangayyan is a professor with the Department of Electrical and Computer Engineering and an adjunct professor of surgery and radiology at the University of Calgary, Alberta, Canada. He received his BEng degree in electronics and communication in 1976 from the University of Mysore at the People's Education Society College of Engineering, Mandya, Karnataka, India, and his PhD in electrical engineering from the Indian Institute of Science, Bangalore, Karnataka, India, in 1980. His research interests are in the areas of digital signal and image processing, biomedical signal analysis, biomedical image analysis, and computer-aided diagnosis. He has published more than 120 papers in journals and 200 papers in proceedings of conferences. His research productivity was recognized with the 1997 and 2001 Research Excellence Awards of the Department of Electrical and Computer Engineering, the 1997 Research Award of the Faculty of Engineering, and by appointment as a university professor in 2003 at the University of Calgary. He was recognized by IEEE with the Third Millennium Medal in 2000 and was elected as a Fellow of IEEE in 2001, Fellow of the Engineering Institute of Canada in 2002, Fellow of the American Institute for Medical and Biological Engineering in 2003, Fellow of SPIE in 2003, Fellow of the Society for Imaging Informatics in Medicine in 2007, and Fellow of the Canadian Medical and Biological Engineering Society in 2007. He has been awarded the Killam Resident Fellowship thrice (1998, 2002, and 2007) in support of his book-writing projects.

J.E. Leo Desautels obtained his MD from the University of Ottawa in 1955 and completed post-graduate training in radiology at the Henry Ford Hospital, Detroit, Michigan. He was a staff radiologist at the Foothills Hospital and a clinical professor with the Faculty of Medicine, the University of Calgary, Alberta, Canada, from 1970 to 1994. He served as a reference radiologist to the Alberta Program for the Early Detection of Breast Cancer until 2007. He is an adjunct professor of electrical and computer engineering at the University of Calgary. He is interested in computer applications in mammography.