

# A completely automated CAD system for mass detection in a large mammographic database

R. Bellotti

*Dipartimento di Fisica, Università di Bari, Sezione INFN di Bari, Italy  
and Center of Innovative Technologies for Signal Detection and Processing (TIRES), Bari, Italy*

F. De Carlo, S. Tangaro,<sup>a)</sup> G. Gargano, and G. Maggipinto

*Dipartimento di Fisica, Università di Bari and Sezione INFN di Bari, Italy*

M. Castellano

*Dipartimento di Elettronica ed Elettrotecnica, Politecnico di Bari, Italy*

R. Massafra

*Dipartimento di Fisica, Università di Bari and Sezione INFN di Bari, Italy*

D. Cascio, F. Fauci, R. Magro, and G. Raso

*Dipartimento di Fisica e Tecnologie Relative, Università di Palermo and Sezione INFN di Catania, Italy*

A. Lauria and G. Forni

*Dipartimento di Scienze Fisiche, Università di Napoli and Sezione INFN di Napoli, Italy*

S. Bagnasco and P. Cerello

*Sezione INFN di Torino, Italy*

E. Zanon

*Ospedale Valdese di Torino and INFN, Sezione di Torino, Italy*

S. C. Cheran

*Dipartimento di Informatica, Università di Torino, Associazione Sviluppo Piemonte  
and Sezione INFN di Torino, Italy*

E. Lopez Torres

*CAEDEN, Habana, Cuba*

U. Bottigli, G. L. Masala, and P. Oliva

*Struttura Dipartimentale di Matematica e Fisica, Università di Sassari and Sezione INFN di Cagliari, Italy*

A. Retico and M. E. Fantacci

*Dipartimento di Fisica, Università di Pisa and Sezione INFN di Pisa, Italy*

R. Cataldo

*Dipartimento di Scienza dei Materiali, Università di Lecce and Sezione INFN di Lecce, Italy*

I. De Mitri

*Dipartimento di Fisica, Università di Lecce and Sezione INFN di Lecce, Italy*

G. De Nunzio

*Dipartimento di Scienza dei Materiali, Università di Lecce and Sezione INFN di Lecce, Italy*

(Received 4 June 2005; revised 18 May 2006; accepted for publication 18 May 2006;  
published 28 July 2006)

Mass localization plays a crucial role in computer-aided detection (CAD) systems for the classification of suspicious regions in mammograms. In this article we present a completely automated classification system for the detection of masses in digitized mammographic images. The tool system we discuss consists in three processing levels: (a) Image segmentation for the localization of regions of interest (ROIs). This step relies on an iterative dynamical threshold algorithm able to select iso-intensity closed contours around gray level maxima of the mammogram. (b) ROI characterization by means of textural features computed from the gray tone spatial dependence matrix (GTSDM), containing second-order spatial statistics information on the pixel gray level intensity. As the images under study were recorded in different centers and with different machine settings, eight GTSDM features were selected so as to be invariant under monotonic transformation. In this way, the images do not need to be normalized, as the adopted features depend on the texture only, rather than on the gray tone levels, too. (c) ROI classification by means of a neural network, with supervision provided by the radiologist's diagnosis. The CAD system was evaluated on a large database of 3369 mammographic images [2307 negative, 1062 pathological (or positive), containing at least one confirmed mass, as diagnosed by an expert radiologist]. To assess the performance of the system, receiver operating characteristic (ROC) and free-response ROC analysis were

employed. The area under the ROC curve was found to be  $A_z=0.783\pm 0.008$  for the ROI-based classification. When evaluating the accuracy of the CAD against the radiologist-drawn boundaries, 4.23 false positives per image are found at 80% of mass sensitivity. © 2006 American Association of Physicists in Medicine. [DOI: 10.1118/1.2214177]

Key words: breast cancer, mammographic mass detection, image processing, computer-aided detection (CAD), textural features, neural network

## I. INTRODUCTION

The analysis of medical images is gathering, in the last years, a growing interest from the scientific community working at the crossover point among physics, engineering, and medicine. The development of computer-aided detection (CAD) systems for the automated search for pathologies could be very useful for the improvement of physicians' diagnosis.

A typical example is the analysis of mammographic images, which are widely recognized as the only imaging modality for an early detection of breast neoplasia.<sup>1,2</sup> Breast cancer is reported as the leading cause of woman cancer deaths in both the United States and Europe. At present, screening programs are the best known method for an early diagnosis in asymptomatic women, thus allowing a reduction of the mortality.<sup>3,4</sup> Screening programs are based on a double visual inspection of the mammographic images, since double reading increases the diagnostic accuracy.<sup>5</sup> From this point of view, the use of a CAD system could provide valuable assistance to the radiologist.

In the present paper, a CAD system for mass detection will be discussed. Masses, as well as microcalcification clusters, are often clear marks of a breast neoplasia. While microcalcifications are small ( $d\approx 0.1\div 1.0$  mm of diameter) and brilliant objects, masses are rather large ( $d\approx 1$  cm of diameter) objects with variable shape and show up with faint contrast. These textural characteristics can be exploited both in the definition of a ROI hunter procedure and in the choice of the proper features to identify positive regions of the mammogram.

A number of CAD systems have already been proposed to the attention of the scientific community. Just to mention only few recent examples, Timp *et al.*<sup>6</sup> discuss an automated technique based on dynamic programming that allows to segment mass lesions from surrounding tissue. In addition, an efficient algorithm is proposed to guarantee that the resulting contour is closed. In Baydush *et al.*,<sup>7</sup> subregion Hotelling observers in conjunction with linear discriminants are investigated for an automated mass detection. The Hotelling observer is claimed to be a good linear detector when information about signal, background, and noise correlation are known. A knowledge databank of ROIs with known ground truth is used for a template matching approach in Tourassi *et al.*:<sup>8</sup> the mutual information is adopted as a similarity metric to determine if a query depicts a true mass. In Catarious *et al.*,<sup>9</sup> a CAD system consisting of filtration, suspicious region localization, feature extraction, feature selection, and classification stages is presented. This algorithm has been improved by adding a new iterative gray level linear segmen-

tation procedure.<sup>10</sup> All these studies report results that are difficult to compare since performances of the different algorithms strongly depend on the database.

In a previous study, our group proposed a CAD scheme based on ROI localization, feature extraction, and neural network classification.<sup>11</sup> Suspect regions were detected by searching for local intensity maxima in rings whose radius was increased until the average intensity decreased to a pre-defined fraction of the local maximum. The ROIs thus obtained were described in terms of statistical features like average, variance, skewness, and kurtosis of the intensity distributions at different fractions of the ROI radius. This scheme relied on a simplified and rough description of the ROI, which was modeled as a round region. An improvement was achieved by implementing a new edge-based segmentation algorithm where ROIs are defined by iso-intensity contours rather than circles.<sup>12</sup> In this paper we retain the improved version of the segmentation step and replace the feature set with Haralik's one.<sup>13</sup> The choice for texture-based features is justified by the successful application of such features to the detection of pathologies in medical image analysis.<sup>14,15</sup>

Comparing our approach with the previously mentioned ones, two main aspects should be stressed. Some algorithms lack an automatic localization of the suspicious regions, rather they make use of manually selected ROIs. Other algorithms, whose scheme includes a computerized ROI hunter, lack a large and heterogeneous database to test the performance in screening-like conditions. Both these points should be taken into account in view of the development a completely automated CAD system, which should assist the radiologists in the framework of a large scale screening program. Our CAD meets both these requirements as it fits in the more general framework of the MAGIC-5 Project (Medical Application on a Grid Infrastructure Connection) which focuses on the development of software tools for biomedical image analysis and their use on distributed image database by means of the GRID technologies.<sup>16,17</sup> Image collection in a screening program intrinsically creates a distributed database, as it involves many hospitals and/or screening centers in different locations. The amount of data generated by such periodic examinations would be so large that it would not be efficient to concentrate them in a single computing center. In addition, it would linearly increase with time, and a full transfer over the network from the collection centers to a central site would likely saturate the available connections. However, making the whole database available to authorized users, regardless of the data distribution, would provide several advantages. The best way to tackle these demands is to

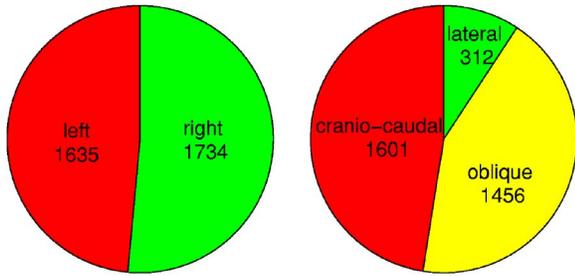


FIG. 1. Database partition. Left: number of left/right breast images. Right: number of cranio-caudal/oblique/lateral views.

use GRID services to manage distributed databases and to allow real time remote diagnosis. This approach would provide access to the full database from any site.

**II. THE IMAGE DATABASE**

The mammograms used in this study were collected in a network of hospitals that collaborate with the MAGIC-5 project.<sup>18,19</sup> Images were acquired using different mammographic screen/film systems and settings (all with molybdenum anode) and in the framework of different applications, including both clinical routine carried out on symptomatic women and screening programs addressed to asymptomatic women. All the images were digitized with a CCD scanner at a pixel size of  $85 \times 85 \mu\text{m}^2$  with 12 bit resolution.<sup>20</sup> Each image is thus  $2657 \times 2067$  pixels with  $G=2^{12}=4096$  gray level tones. No normalization is applied to the images. The database consists of 3369 mammograms from 967 analyzed subjects. Some of the mammograms are different views (cranio-caudal, lateral, oblique) from the same subject and are treated as different samples in the analysis. The pie diagrams reported in Fig. 1 show the partition of the database in left/right breast images (left) and cranio-caudal/oblique/lateral views (right).

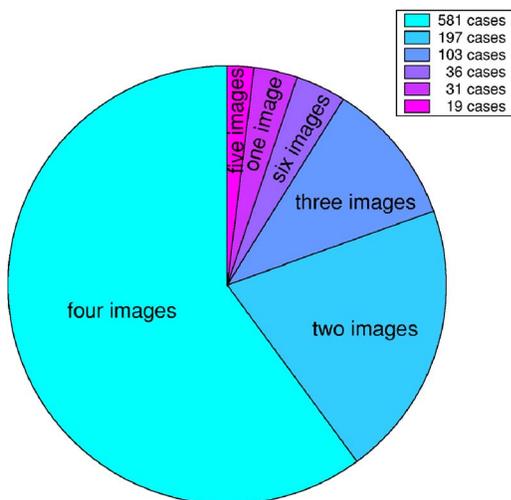


FIG. 2. Number of cases with one to six images.

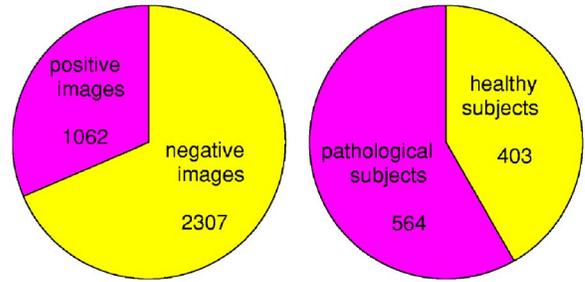


FIG. 3. Database composition: images (left) and subjects (right).

Figure 2 displays the number of cases in the database with one to six images.

We consider positive images the ones that contain at least one mass, as diagnosed by an expert radiologist and confirmed by biopsy; images with no mass at the first exam and after a follow-up of at least 3 years are considered as negative, even if they contain some other pathology (e.g., microcalcifications). The breakdown of the cases is displayed in Fig. 3 for both the images (positive/negative) and the analyzed subjects (pathological/healthy). Each positive image comes with a description of the lesion including radiological diagnosis, histological data, and type of mass, as shown in Fig. 4. The total number of masses is 1236. The location and size of a mass is defined by a radiologist-drawn circle, characterized by center coordinates  $(X_{\text{rad}}, Y_{\text{rad}})$  and radius  $(R_{\text{rad}})$ , which fully contains the mass. The radius size of the masses ranges from 3.1 to 47.2 mm with a mean size of 11.7 mm. Figure 5 shows some images of the database.

As far as the breast background is concerned, we adopt a tissue classification which is used as a standard by many Italian radiologists:<sup>21,22</sup>

- fibro-adipose tissue: indicates a fat breast with little fibrous connective tissue;

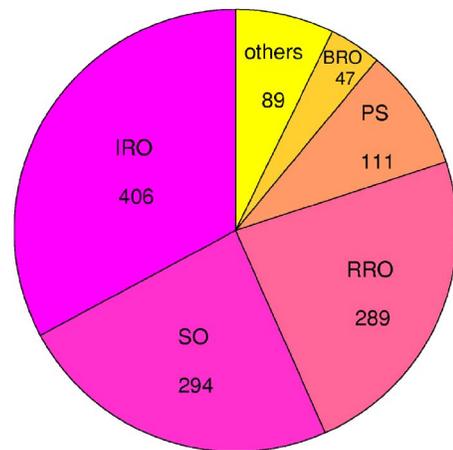


FIG. 4. Different kinds of masses present in the database. Legend: IRO =irregular round opacity, SO=spiculated opacity, RRO=regular round opacity, PS=parenchymal distortion, and BRO=blurred round opacity; “others” include a combination of the above mentioned kinds.

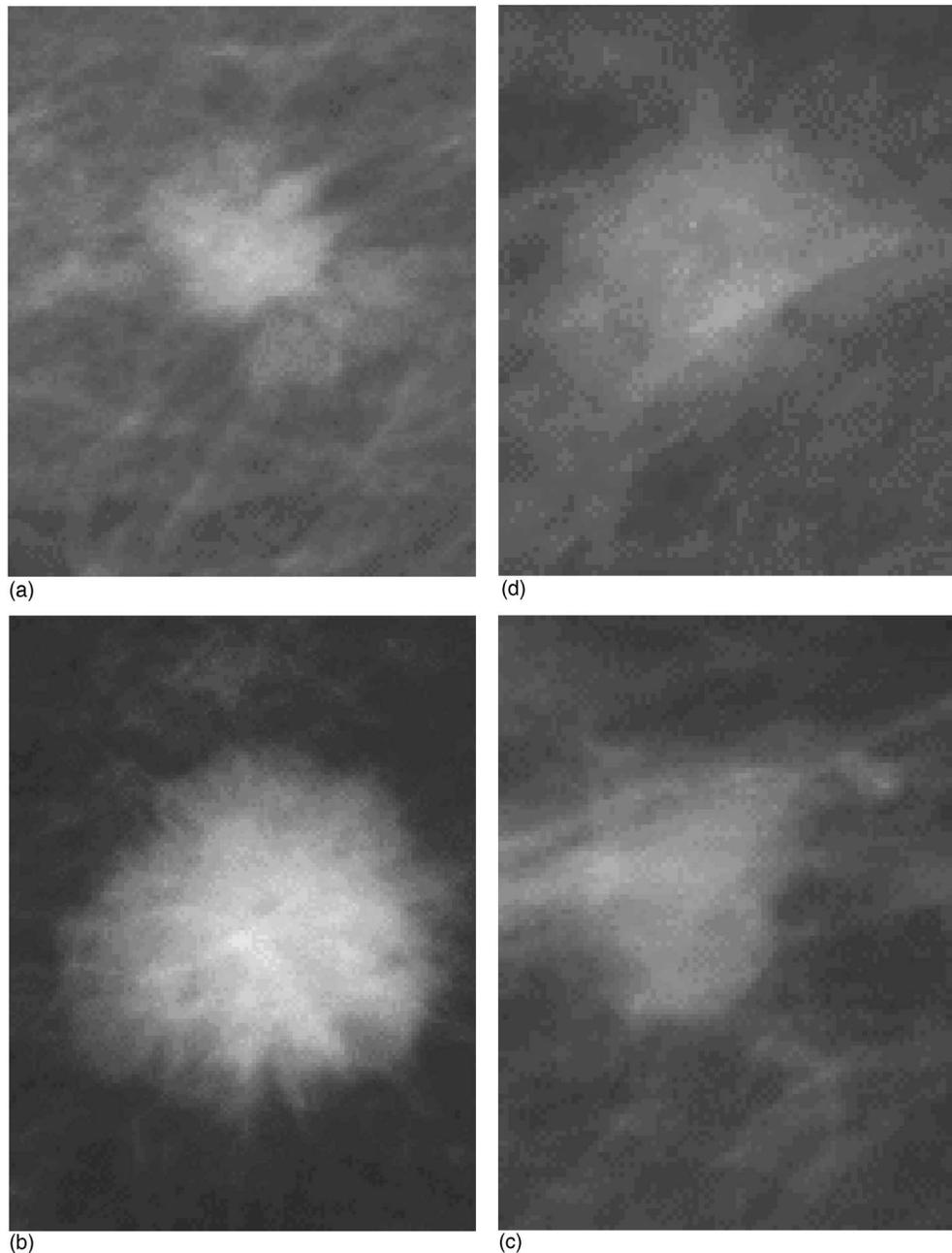


FIG. 5. Some ROIs present in the database: (a) spiculated (star-like) opacity; round opacities with regular (b), irregular (c), and blurred (d) edge.

- glandular tissue: indicates the presence of prominent duct patterns;
- dense tissue: indicates a dense breast parenchyma.

The breast background classification is based only on the appearance of the parenchyma, without any reference to skin, vascularity, presence/absence of masses, calcifications, lymph nodes, nor to parity, history of breast disease, age, and family history.

Figure 6 reports the background composition of the database. Most of the images are glandularlike: the detection of pathological structures in this kind of images is a quite hard task, since the target is surrounded by a “noisy” environment.

### III. METHODS

The CAD system consists of three main steps: (1) segmentation, (2) feature extraction, and (3) classification.

The goal of the segmentation step is to locate, within the image, the suspicious regions, or ROIs, which are likely to contain a mass. All the detected ROIs are characterized by a proper set of features providing texture information on the pixel intensity. The feature vector describing the ROI will be identified as “pattern.” In general, a number of ROIs can be detected with different degrees of superimposition on the same mass, though not overlapping among them. So, ROI-to-mass is not a one-to-one mapping. A tagging criterion, relying on the superimposition with the radiologist-drawn

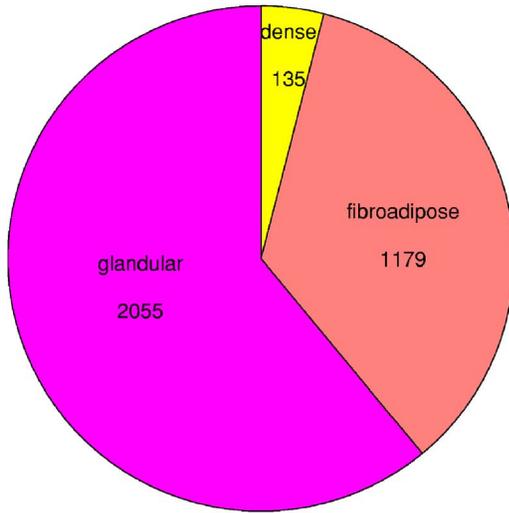


FIG. 6. Breast background composition of the database.

boundary, is adopted to define the true positive (TP) ROIs. These ROIs are used to train the neural network with a ground truth based on the radiologist's diagnosis. In the following subsections each of the processing steps is described in detail.

### A. Segmentation method

Prior to the processing, the images are made anonymous and the borders of the breast are extracted by means of a threshold algorithm. As the mammograms are digitized with the breast on the right part of the image, moving on a row of the mammogram in the right-left direction, a fixed length segment with pixel intensity less than  $I=16$  is searched, and all the pixels at the left side of this segment are set at zero value. By repeating the routine row by row, the contour of the breast is found, and noninteresting portions of the mammogram are cut off. The accuracy of the border extraction has been verified case by case.

An edge-based segmentation algorithm for the selection of the suspicious regions has been implemented.<sup>12</sup> The algorithm works as follows:

- the mammogram is divided into square cells of size  $S$  and a relative gray level maximum  $I_M$  (initial center for the candidate lesion) is searched in each cell, starting from the right top cell;
- an iso-intensity contour, including the pixel of the intensity relative maximum, is drawn at a threshold value  $I_{th}=I_M/2$ , thus delimiting a ROI with area  $A_R$ , which can exceed the cell dimension;
- the threshold  $I_{th}$  is increased/decreased by an amount which is one half of the previous one if the ROI area  $A_R$  is, respectively, greater/smaller than a limit area  $A_L$ ; the iteration is stopped when the difference between two consecutive thresholds is less than  $I_\Delta$  gray levels;

- the ROI is removed and stored for feature extraction and classification;
- the processing is repeated for the following square cell.

The partition of the image into cells of size  $S$  does not limit the search to ROIs smaller than  $S \times S$ , as the ROI area depends only of the final value of the threshold  $I_{th}$ . The values of the  $S$ ,  $A_L$ , and  $I_\Delta$  parameters are defined in order to maximize the efficiency of the ROI hunter algorithm (percentage of massive lesions correctly identified among those selected by the radiologists). The best values thus obtained are  $S=200$  pixels,  $A_L=500 \times 500$  pixels, and  $I_\Delta=4$  gray levels.

Figure 7 shows ROIs selected with the segmentation algorithm (right), together with the original image (left).

The number of ROIs detected from each image is related to the texture properties of the mammogram. All the ROIs extracted from negative images are tagged as negatives, while the ROIs from positive images can be labelled as true positive (TP) or false positive (FP) depending if they meet or not the following criterion: a minimal rectangle, fully containing the ROI, is drawn with parallel sides with respect to the ones of the image. Let  $(2L_x, 2L_y)$  be the sides of the rectangle,  $(X_{cad}; Y_{cad})$  the center coordinates of the ROI detected by the CAD, and  $\{(X_{rad}; Y_{rad}), R_{rad}\}$  the center coordinates and the radius of the radiologist-drawn boundary. We consider the following tagging condition:

$$|X_{rad} - X_{cad}| < \max(R_{rad}, L_x)$$

AND

$$|Y_{rad} - Y_{cad}| < \max(R_{rad}, L_y). \quad (1)$$

For the following classification step only TP and negative ROIs are used for both training and testing the neural network, according to the cross-validation technique (see Sec. III C), while the FP ROIs are used for validation purposes only. This is why we adopt a very strict condition like (1) to define the TP ROIs. Looser criteria were tested and the performance of the overall system worsened. In fact, looser conditions lead to tag as TP many ROIs that are not well superimposed to the radiologist-drawn region. As a consequence, the texture-based features used to describe the ROIs (see Sec. III B) are not so accurate, thus causing the neural network to work badly.

The efficiency of the ROI hunter, computed as the percentage of masses correctly detected among those found by the radiologist, is 83.1%, corresponding to 1027 detected masses, with respect to 1236 radiologist-drawn boundaries. Correspondingly, 6.27 FPP are obtained at this level, and the average area of the selected ROIs is 15% of the total area of the image. Table I reports the breakdown of the selected ROIs.

### B. Feature extraction

Texture analysis can be used either to segment the image into areas indicating the mass or to measure textural features to classify possible pathological regions. The successful ap-

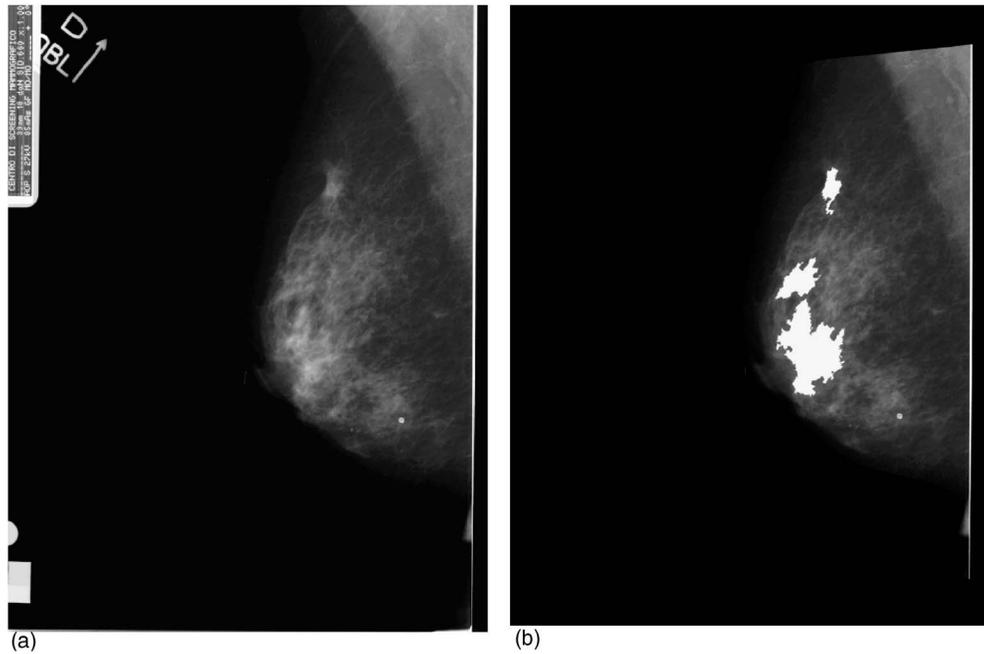


FIG. 7. Left: the original image. Right: the segmented image.

plications of such features to the detection of pathologies in medical image analysis<sup>14,15</sup> led us to adopt a similar approach in our CAD for the characterization of the selected ROIs. The focus of the analysis is the computation of the gray level co-occurrence matrix (GLCM),<sup>13</sup> also known as spatial gray level dependence (SGLD) matrix.<sup>23</sup> To this purpose, we consider the minimal rectangular portion of the image, which fully includes the ROI. As the name suggests, the GLCM is constructed from the image by estimating the pairwise statistics of pixel intensity, thus relying on the assumption that the texture content information of an image is contained in an overall or average spatial relationship between pairs of pixel intensities.<sup>13</sup> A co-occurrence matrix  $\mathcal{M}$  is a  $G \times G$  matrix, whose rows and columns are indexed by the image gray levels  $i=1, \dots, G$ , where  $G=2^n$  for a  $n$ -bit image. Each element  $p_{ij}$  represents an estimate of the probability that two pixels with a specified polar separation  $(d, \theta)$  have gray levels  $i$  and  $j$ . Coordinates  $d$  and  $\theta$  are, respectively, the distance and the angle between the two pixels  $i$  and  $j$ . In their seminal paper, Haralik *et al.*<sup>13</sup> considered only displacements  $d=1$  at quantized angles  $\theta=k\pi/4$ , with  $k=0, 1, 2, 3$ , thus having  $\mathcal{M}_{d,\theta}(j, i) = \mathcal{M}_{d,\theta+\pi}(i, j)$ . Symmetry is achieved by averaging the GLCM with its transpose, thus leading to invariance under  $\pi$  rotations, too. Textural features can be derived from the GLCM and used in texture classification in place of the single GLCM elements. Fourteen features are introduced, related to textural properties of the im-

age such as homogeneity, contrast, presence of organized structure, complexity, and nature of gray tone transitions. The values of these features are sensitive to the choice of the direction  $\theta$ , given that the parameter  $d$  is fixed to 1. Invariance under rotation should be restored in order to avoid describing two images, one obtained by rotating the other, with different feature sets. This is achieved by considering average and range of each feature value over the  $\theta$  angles, thus obtaining 28 textural variables, even if only a few of them are used as inputs to the classifier.<sup>23-25</sup>

As the texture is gray tone independent, either the image must be normalized or one should choose features that are invariant under monotonic gray level transformation. As said in Sec. II, the images of our database come from different centers and no kind of normalization is applied. For this reason we select, among all the GLCM features, the ones that are invariant under monotonic gray tone transformation:

(1) energy:

$$f_1 = \sum_{ij} p_{ij}^2; \quad (2)$$

(2) entropy:

$$f_2 = - \sum_{ij} p_{ij} \ln(p_{ij}); \quad (3)$$

(3) information measures of correlation:

$$f_3 = \frac{f_2 - H_1}{\max\{H_x, H_y\}}; \quad (4)$$

$$f_4 = [1 - \exp\{-2(H_2 - f_2)\}]^{1/2}, \quad (5)$$

where

TABLE I. Breakdown of the selected ROIs.

TP	FP	Negative	Total
1207	7642	13473	22322

TABLE II. Breakdown of the patterns for the cross validation: first set A is used for training and set B for testing, then vice versa.

	Set A	Set B	Validation
TP ROI	603	604	...
Negative ROI	604	603	13473
FP ROI	...	...	7642
Total	1207	1207	21115

$$P_x(i) = \sum_j p_{ij}, \quad (6)$$

$$P_y(j) = \sum_i p_{ij}, \quad (7)$$

$$H_1 = - \sum_{ij} p_{ij} \ln\{P_x(i)P_y(j)\}, \quad (8)$$

$$H_2 = - \sum_{ij} P_x(i)P_y(j) \ln\{P_x(i)P_y(j)\}, \quad (9)$$

$$H_x = - \sum_i P_x(i) \ln\{P_x(i)\}, \quad (10)$$

$$H_y = - \sum_j P_y(j) \ln\{P_y(j)\}. \quad (11)$$

For each of the above-mentioned features  $\{f_i\}_{i=1,\dots,4}$ , average and range are computed for the angles  $\theta = k\pi/4$ , with  $k=0,1,2,3$  and  $d=1$ , thus obtaining eight textural features.

### C. Classification

A number of classifiers based on linear discriminant analysis,<sup>26,27</sup> artificial neural networks,<sup>28-30</sup> and rule-based methods<sup>31,32</sup> have shown effectiveness in detection and diagnostic systems. We used a supervised two-layered feed-

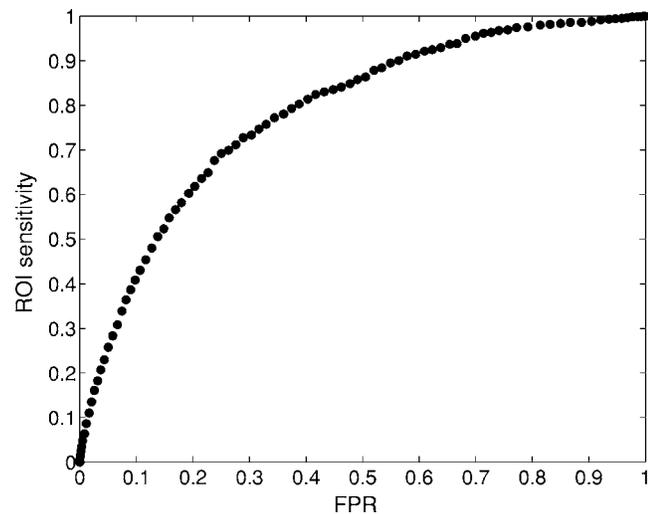


FIG. 8. ROC curve for ROI-based classification. The area under the curve (AUC) is  $A_z=0.783 \pm 0.008$ .

TABLE III. Values of the AUC at different number  $N_h$  of the hidden neurons.

	$N_h=8$	$N_h=10$	$N_h=12$	$N_h=14$	$N_h=16$
$A_z$	0.777	0.781	0.781	0.783	0.783

forward neural network, trained with the gradient descent learning rule<sup>33</sup> for the ROI pattern classification:

$$\Delta w_{ij}(\tau) = -\eta \frac{\partial E(\tau)}{\partial w_{ij}} + \alpha \Delta w_{ij}(\tau-1), \quad (12)$$

$$E(\tau) = \frac{1}{2} \sum_{\mu} (t^{\mu} - y^{\mu})^2, \quad (13)$$

where the  $E(\tau)$  function measures the error of the network outputs  $y^{\mu}$  in reproducing the targets  $t^{\mu}=1,0$ , at iteration  $\tau$ , and  $w_{ij}$  are the network weights. The second term in (12), known as *momentum*,<sup>34</sup> represents a sort of inertia which is added to quickly move along the direction of decreasing gradient, thus reducing the computational time to the solution. Different values of the momentum parameter  $\alpha$  were tested and the best trade-off between performance and computational time was reached for  $\alpha=0.1 \div 0.2$ . The learning rate was  $\eta=0.01$ . A sigmoid transfer function was used:

$$g(x) = \frac{1}{1 + e^{-\beta x}}, \quad (14)$$

with *gain factor*  $\beta=1$ .

The network architecture consisted of  $N_i=8$  input neurons and one output neuron. The size of the hidden layer was tuned in the range  $[N_i-1, 2N_i+1]$  to optimize the classification performance. All the TP ROIs ( $N_{TP}=1207$ ) and as many negative ones were used to train the neural network. To make sure that the negative training patterns were representative, they were selected with a probability given by the distribution of the whole negative ROI set, in the eight-dimensional feature space. With a random procedure we build up two sets

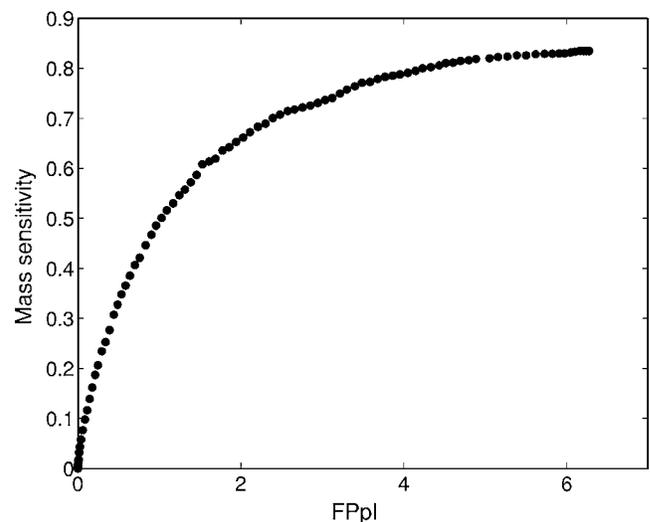


FIG. 9. CAD FROC curve for mass-based classification.

TABLE IV. Maximal and mean values of the AUC for mass-based classification.

	SO	BRO	IRO	RRO	PS	Others
Maximal AUC value	0.711	0.749	0.719	0.554	0.718	0.699
Mean AUC value	0.703	0.745	0.714	0.547	0.712	0.696

(A and B), each one made of 1207 patterns, which are used, in turn, for both training and test, according to the cross-validation technique:<sup>35</sup> first, the network is trained with set A and tested with set B, then the two sets are reversed. In addition, we take care that the occurrence for each kind of mass and tissue present in the database is balanced in set A and set B, in order to train and test the network in the most complete and correct way.

All the other patterns (negative ROIs not selected for the training stage and FP) are used for validation only. The results presented in the following section (see ROC curves in Sec. IV) refer to the classification of all the patterns at our hand. The breakdown of the patterns for the cross validation is reported in Table II.

#### IV. RESULTS AND DISCUSSION

The results are provided in terms of both ROC and FROC curves. The ROC curve is particularly suitable when testing a binary hypothesis:<sup>36</sup> it is obtained by plotting the ROI sensitivity against the ROI false positive rate (FPR), at different values of the decision threshold on the neural network output. While the ROC curve displays the neural network performance in classifying the ROI patterns, the FROC curve provides the performance of the overall CAD system in detecting the masses, as it reports the mass sensitivity against the number of false positive masses per image (FPpI).

Figure 8 displays a typical ROC curve obtained for the pattern classification. The area under the curve (AUC) is  $A_z = 0.783 \pm 0.008$ , where the error is computed as reported in Hanley *et al.*<sup>37</sup> The results are quite insensitive to the number  $N_h$  of the hidden neurons as the different values of the AUC obtained for different sizes of the hidden layer lie in the error range (see Table III).

As said in Sec. III, a number of ROIs can be superimposed to the same mass, according to the condition (1), though not overlapping among them. For this reason, it should be more useful to provide the results in terms of the mass sensitivity, defined as the fraction of masses correctly detected by the CAD with respect to the total number of

TABLE V. Maximal and mean values of the AUC for tissue-based classification.

	Dense	Glandular	Fibro-adipose
Maximal AUC value	0.749	0.704	0.654
Mean AUC value	0.744	0.697	0.648

TABLE VI. FPpI for the main kind of masses present in the database.

	SO	BRO	IRO	RRO	PS	Others
Sensitivity	80%	80%	80%	70%	80%	80%
FPpI	3.40	2.76	3.30	5.16	2.58	3.40

radiologist-drawn boundaries. In this way, the accuracy of the overall CAD system can be assessed against the radiologist's diagnosis. To this purpose, the following prescription is adopted: a mass is correctly detected by the CAD system if at least one ROI, among the ones superimposed to that mass, is classified as positive by the neural network.

A free-response ROC (FROC) curve can be drawn (see Fig. 9), which reports the mass sensitivity of the overall system against the number of false positive per image (FPpI): 80% of mass sensitivity is achieved with 4.23 FPpI.

Both the neural network classification and the overall CAD performance were evaluated for different kinds of masses and tissues. The classification parameters were set in order to maximize the performance, though changes in the results are within the error range. Tables IV and V report the maximal and the mean AUC values for both a mass-based and a tissue-based classification. The values of the FPpI and mass sensitivity for each of the main kinds of masses and tissues present in the database are reported in Tables VI and VII, respectively. When detecting RRO the performance of the CAD decreases as a maximum of 70% sensitivity is achieved with 5.16 FPpI.

As reported by the previous tables, the CAD is robust against different kinds of both masses and tissues. As far as the neural performance is concerned, as our AUC value  $A_z = 0.783$  lies in the range  $0.7 < AUC < 0.9$ , the classifier can be considered accurate enough.<sup>36</sup> Each ROI is scored with the neural output that represents the probability to contain a mass. Our CAD system was tested on a digitized mammographic image database with a size much greater than the ones<sup>6-10</sup> described in the Introduction. Though direct comparison is unfeasible, since the performances of the algorithms strongly depends on the database, it should be remarked that the high image statistics reduces the error in the performance evaluation. Since our database spans a wide range of image and mass kinds, this allows the following:

- (1) to approach a situation which is close to the real clinic one,
- (2) to check the flexibility of the methods on images of different quality, and
- (3) to test the CAD with different cases to analyze.

TABLE VII. FPpI at 80% sensitivity for each kind of tissue of the images of the database.

	Dense	Glandular	Fibro-adipose
FPpI	3.22	3.59	4.89

## V. CONCLUSIONS

It is well known that mammogram interpretation is a very difficult task even for experienced radiologists; from this point of view, CAD systems can be a useful tool to help them.

Two main characteristics distinguish our CAD from the other approaches: (1) the complete automatization of the computation chain (segmentation  $\rightarrow$  feature extraction  $\rightarrow$  classification), and (2) the large database of mammographic images used to evaluate the CAD performance. In fact, the system is able to automatically select suspicious portions of the mammogram that are more likely to contain a mass and provide the cancer probability (i.e., mass sensitivity) at a certain value of FPPi. In addition, the performance of the system was evaluated on a large database of mammographic images that were collected in different hospitals and with different settings. The database includes breast masses spanning a wide range of shapes and sizes, as well as the main types of breast tissue, thus reproducing at best the typical situation of a screening program.

## ACKNOWLEDGMENTS

We thank Professor V. Lattanzio (Policlinico di Bari, Italy), Professor Bazzocchi (Dipartimento di Ricerche Mediche e Morfologiche, Università di Udine, Italy), Professor A. Sodano (Dipartimento di Scienze Biomorfologiche e Funzionali, Università Federico II, Napoli, Italy), Professor R. Lagalla, and Dr. R. Ienzi (Istituto di Radiologia P. Cignolini, Università di Palermo) for the medical support in the data acquisition.

<sup>a)</sup>Address for correspondence: Dipartimento di Fisica, Università di Bari, Via Amendola 173, 70126 Bari, Italy. Electronic mail: sonia.tangaro@ba.infn.it

<sup>1</sup>L. W. Bassett, R. H. Gold, and S. C. Kimme, "History of the technical development of mammography, Syllabus: A Categorical course in Physics," RSNA (1994).

<sup>2</sup>L. Tabar *et al.*, "Reduction in mortality from breast cancer after mass screening with mammography," *Lancet* **8433**, 829–832 (1985).

<sup>3</sup>S. Feig, "Increased benefit from shorter screening mammography intervals for women ages 40–49 years," *Cancer* **80**, 2035–2039 (1997).

<sup>4</sup>R. Bird, T. Wallace, and B. Yankaskas, "Analysis of cancer missed at screening mammography," *Radiology* **184**, 613–617 (1992).

<sup>5</sup>N. Karssemeijer, J. D. Otten, A. L. Verbeek, J. H. Groenewoud, H. J. de Koning, J. H. Hendriks, and R. Holland, "Computer-aided detection versus independent double reading of masses in mammograms," *Radiology* **227**, 192–200 (2003).

<sup>6</sup>S. Timp and N. Karssemeijer, "A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography," *Med. Phys.* **31**, 958–971 (2004).

<sup>7</sup>A. H. Baydush, D. M. Catarious, Jr., C. K. Abbey, and C. E. Floyd, "Computer aided detection of masses in mammography using subregion Hotelling observers," *Med. Phys.* **30**, 1781–1787 (2003).

<sup>8</sup>G. D. Tourassi, R. Vargas-Voracek, D. M. Catarious, Jr., and C. E. Floyd, Jr., "Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information," *Med. Phys.* **30**, 2123–2130 (2003).

<sup>9</sup>D. M. Catarious, Jr., A. H. Baydush, C. K. Abbey, and C. E. Floyd, Jr., "A mammographic mass CAD system incorporating features from shape, fractal, and channelized Hotelling observer measurements: preliminary results," in *SPIE Medical Imaging 2003*, San Diego, CA San Diego, CA, 2003, p. 1927.

<sup>10</sup>D. M. Catarious, Jr., A. H. Baydush, and C. E. Floyd, Jr., "Incorporation of an iterative, linear segmentation routine into a mammographic mass

CAD system," *Med. Phys.* **31**, 1512–1520 (2004).

<sup>11</sup>F. Fauci, G. Raso, R. Magro, G. Forni, A. Lauria, S. Bagnasco, P. Cerello, S. C. Cheran, E. Lopez Torres, R. Bellotti, F. De Carlo, G. Gargano, S. Tangaro, I. De Mitri, G. De Nunzio, and R. Cataldo, "A massive lesion detection algorithm in mammography," *Phys. Medica* **XXI**(1), 21–28 (2005).

<sup>12</sup>F. Fauci, S. Bagnasco, R. Bellotti, D. Cascio, S. C. Cheran, F. De Carlo, G. De Nunzio, M. E. Fantacci, G. Forni, A. Lauria, E. Lopez Torres, R. Magro, G. L. Masala, P. Oliva, M. Quarta, G. Raso, A. Retico, and S. Tangaro, "Mammogram segmentation by contour searching and massive lesion classification with neural network," *Proc. IEEE Medical Imaging Conference*, 16–22 October 2004, Rome, Italy.

<sup>13</sup>R. M. Haralik, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 610–621 (1973).

<sup>14</sup>D. Wei, H.-P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," *Med. Phys.* **24**, 903–914 (1997).

<sup>15</sup>G. S. Cox, F. J. Hoare, and G. de Jager, "Experiments in lung cancer nodule detection using texture analysis and neural network classifiers," in *Third South African Workshop on Pattern Recognition*, 1992.

<sup>16</sup>R. Bellotti, S. Bagnasco, U. Bottigli, M. Castellano, R. Cataldo, E. Catanzariti, P. Cerello, S. C. Cheran, F. De Carlo, P. Delogu, I. De Mitri, G. De Nunzio, M. E. Fantacci, F. Fauci, G. Forni, G. Gargano, B. Golosio, P. L. Indovina, A. Lauria, E. Lopez Torres, R. Magro, D. Martello, G. L. Masala, R. Massafra, P. Oliva, R. Palmiero, A. Preite Martinez, R. Prevette, M. Quarta, L. Ramello, G. Raso, A. Retico, M. Santoro, M. Sitta, S. Stumbo, S. Tangaro, A. Tata, and E. Zanon, "The MAGIC-5 Project: Medical applications on a grid infrastructure connection," *Proc. IEEE Nuclear Science Symposium*, 16–22 October 2004, Rome, Italy.

<sup>17</sup>C. S. Cheran, R. Cataldo, P. Cerello, F. De Carlo, F. Fauci, G. Forni, B. Golosio, A. Lauria, E. Lopez Torres, D. Martello, G. Masala, G. Raso, A. Retico, and A. Tata, "Detection and classification of microcalcification clusters in digital mammograms," *Proc. IEEE Medical Imaging Conference*, 16–22 October 2004, Rome, Italy.

<sup>18</sup>M. Bazzocchi, I. Facecchia, C. Zuiani, V. Londero, S. Smania, U. Bottigli, and P. Delogu, "Application of a computer-aided detection (CAD) system to digitalized mammograms for identifying microcalcifications," *Radiol. Med. (Torino)* **101**, 334–340 (2001).

<sup>19</sup>M. E. Fantacci, U. Bottigli, P. Delogu, F. Fauci, B. Golosio, A. Lauria, R. Palmiero, G. Raso, S. Stumbo, and S. Tangaro, "Search of microcalcifications clusters with the CALMA CAD Station," *Physics of the Medical Imaging*, San Diego, USA, 24–26 February, 2002.

<sup>20</sup>S. R. Amendolia *et al.*, "Comparison of imaging properties of several digital radiographic system," *Nucl. Instrum. Methods Phys. Res. A* **466**, 95–98 (2001).

<sup>21</sup>J. N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern," *Cancer* **37**, 2486–2492 (1976).

<sup>22</sup>V. Lattanzio and G. Simonetti, "Mammography-guide to interpretation, reporting and auditing mammographic images Re. Co. R. M." Springer-Verlag ISBN 3-540-20018-5.

<sup>23</sup>R. W. Connors and C. A. Harlow, "A theoretical comparison of texture algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.* **2**, 204–222 (1980).

<sup>24</sup>R. W. Connors, M. M. Trivedi, and C. A. Harlow, "Segmentation of a high-resolution urban scene using texture operators," *Comput. Vis. Graph. Image Process.* **25**, 273–310 (1984).

<sup>25</sup>J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Syst. Man Cybern.* **6**, 269–285 (1976).

<sup>26</sup>H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space," *Phys. Med. Biol.* **40**, 857–876 (1995).

<sup>27</sup>H. P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. Sanjay-Gopal, "Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study," *Radiology* **212**, 817–827 (1999).

<sup>28</sup>G. te Brake and N. Karssemeijer, "Segmentation of suspicious densities in digital mammograms," *Med. Phys.* **28**, 259–266 (2001).

<sup>29</sup>Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to deci-

- sion making in the diagnosis of breast cancer," *Radiology* **187**, 81–87 (1993).
- <sup>30</sup>B. Zheng, Y. Chang, W. F. Good, and D. Gur, "Performance gain in computer-assisted detection schemes by averaging scores generated from artificial neural networks with adaptive filtering," *Med. Phys.* **28**, 2302–2308 (2001).
- <sup>31</sup>B. Zheng, Y. H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," *Acad. Radiol.* **2**, 959–966 (1995).
- <sup>32</sup>L. Li, Y. Zheng, L. Zheng, and R. A. Clark, "False-positive reduction in CAD mass detection using a competitive classification strategy," *Med. Phys.* **28**, 250–258 (2001).
- <sup>33</sup>J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, MA, 1991).
- <sup>34</sup>D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing* (MIT, Cambridge, MA, 1986), Vol. I.
- <sup>35</sup>M. Stone, "Cross-validators choice and assessment of statistical predictions," *J. R. Stat. Soc. Ser. B. Methodol.* **36**(1), 111–147 (1974).
- <sup>36</sup>J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science* **240**, 1285–1293 (1988).
- <sup>37</sup>J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology* **143**, 29–36 (1982).