# 10.4

# Computer-Aided Detection and Diagnosis in Mammography

Mehul P. Sampat,
Mia K. Markey,
and Alan C. Bovik
*The University of Texas at Austin*



## 1 Introduction

The American Cancer Society estimates that 215,990 women will be diagnosed with breast cancer in the United States in 2004 [1]. Another 40,110 women will die of the disease. In the United States, breast cancer is the most common form of cancer among women and is the second leading cause of cancer deaths after lung cancer [1]. Women in the United States have about a 1 in 8 lifetime risk of developing invasive breast cancer [2, 3]. Early detection of breast cancer increases the survival rate and increases the treatment options.

Screening mammography, radiographic imaging of the breast, is currently the most effective tool for early detection of breast cancer. Screening mammographic examinations are performed on asymptomatic woman to detect early, clinically unsuspected breast cancer. Two views of each breast are recorded; the craniocaudal (CC) view, which is a top-to-bottom view, and a mediolateral oblique (MLO) view, which is a side view taken at an angle. Examples of the MLO and CC views are shown in Fig. 1.

Radiologists visually search mammograms for specific abnormalities. Some of the important signs of breast cancer that radiologists look for are clusters of microcalcifications, masses, and architectural distortions. A mass is defined as a space-occupying lesion seen in at least two different projections [4]. Masses are described by their shape and margin characteristics. Calcifications are tiny deposits of calcium, which appear as small bright spots on the mammogram. They are characterized by their type and distribution properties. An architectural distortion is defined as follows: ''The normal architecture is distorted with no definite mass visible. This includes spiculations radiating from a point, and focal retraction or distortion of the edge of the parenchyma'' [4]. A typical example of each of these abnormalities is shown in Fig. 2.

Breast lesions are described and reported according to the Breast Imaging Reporting and Data System (BI-RADS$^{TM}$) [4]. BI-RADS$^{TM}$ is a mammography lexicon developed by the American College of Radiology (ACR) for the description of mammographic lesions. The BI-RADS$^{TM}$ lexicon includes

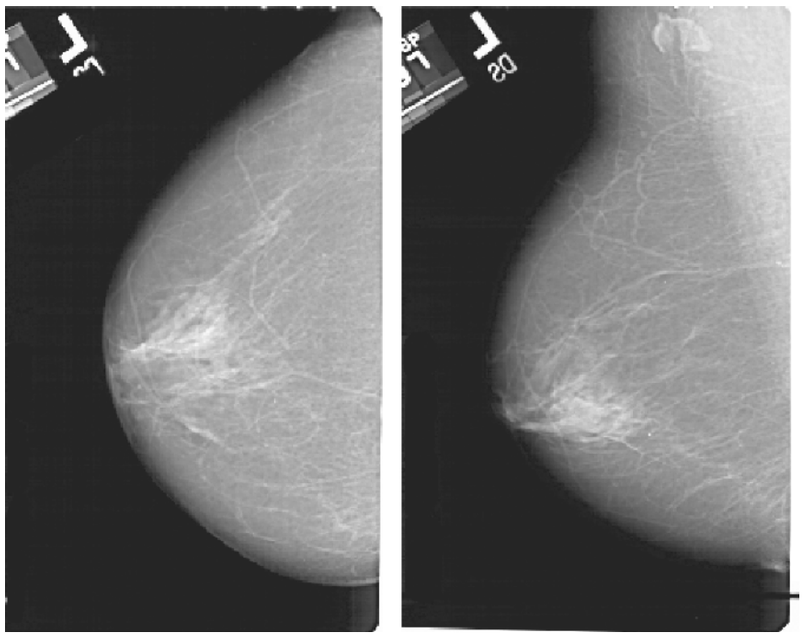

**FIGURE 1** In screening mammography two views of each breast are recorded; the craniocaudal (CC) view (**left**), which is a top-to-bottom view, and a mediolateral oblique (MLO) view (**right**), which is a side view taken at an angle. The images were obtained from the Digital Database for Screening Mammography (DDSM) [126].

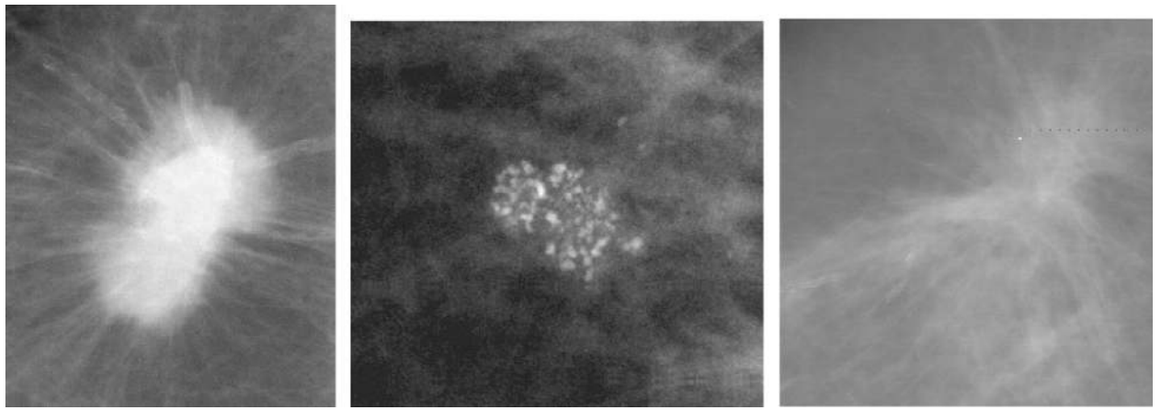

**FIGURE 2** Examples of a spiculated mass (**left**), cluster of microcalcifications (**center**), and architectural distortion (**right**). The images were obtained from Digital Database for Screening Mammography (DDSM) [126].

descriptors such as the margin of a mass and the distribution of calcifications and it defines final assessment categories to describe the radiologist's level of suspicion about the mammographic abnormality. It has been demonstrated that the BI-RADS$^{TM}$ final assessment rating is an indicator of the likelihood of malignancy [5]. If a suspicious abnormality is detected, a diagnostic mammographic examination is carried out to decide the future course of action required. Based on the level of suspicion of the abnormality following the diagnostic examination, a recommendation is made for routine follow-up, short-term follow-up, or biopsy.

Early detection via mammography increases breast cancer treatment options and the survival rate [6]. However, mammography is not perfect. Detection of suspicious abnormalities is a repetitive and fatiguing task. For every thousand cases analyzed by a radiologist, only three to four are cancerous and thus an abnormality may be overlooked. As a result, radiologists fail to detect 10% to 30% of cancers [7–9]. Approximately two thirds of these false-negative results are due to missed lesions that are evident retrospectively [10]. Due to the considerable amount of overlap in the appearance of malignant and benign abnormalities, mammography has a

positive predictive value (PPV) of less than 35% [11], where the PPV is defined as the percentage of lesions subjected to biopsy that were found to be cancer. Thus, a high proportion of biopsies are performed on benign lesions. Avoiding benign biopsies would spare women anxiety, discomfort, and expense.

Computer-aided detection (CAD) systems have been developed to aid radiologists in detecting mammographic lesions that may indicate the presence of breast cancer. These systems act only as a second reader and the final decision is made by the radiologist. Recent studies have also shown that CAD detection systems, when used as an aid, have improved radiologists' accuracy of detection of breast cancer [12–15]. Computer-aided diagnosis (CADx) systems for aiding in the decision between follow-up and biopsy are still in development. It is important to realize that mammographic image analysis is an extremely challenging task for a number of reasons. First, since the efficacy of CAD/CADx systems can have very serious implications, there is a need for near perfection. Second, the large variability in the appearance of abnormalities makes this a very difficult image analysis task. Finally, abnormalities are often occluded or hidden in dense breast tissue, which makes detection difficult.

The organization of the chapter is as follows. Section 2 discusses CAD algorithms while Section 3 discusses CADx algorithms. A number of commercial CAD systems have been approved by the Food and Drug Administration (FDA) for the detection task. Section 4 discusses the commercial CAD systems currently available. Finally, Section 5 discusses future areas of research for the breast cancer CAD/CADx community.

## 2 Computer-Aided Detection of Mammographic Abnormalities

The goal of the detection stage is to assist radiologists in locating abnormalities on the mammogram. A flowchart showing the different steps involved in detection algorithms is shown in Fig. 3. The metrics used to report the performance of detection algorithms are sensitivity (Equation 1) and the number of false positives per image (FPI; Equation 2). A true-positive mark is a mark made by the CAD system that corresponds to the location of a lesion. A false-positive mark is a mark made by the CAD system that does not correspond to the location of a lesion. A plot of sensitivity versus FPI is called a free-response receiver operating characteristic (FROC) plot and this is generally used to report the performance of the detection algorithm. An example of an FROC plot is shown in Fig. 4.

There is some disagreement regarding the manner in which detection results should be reported. While most authors report the performance in terms of the detection of any "actionable" objects, some report it terms of how many malignant masses were detected, since they believe that

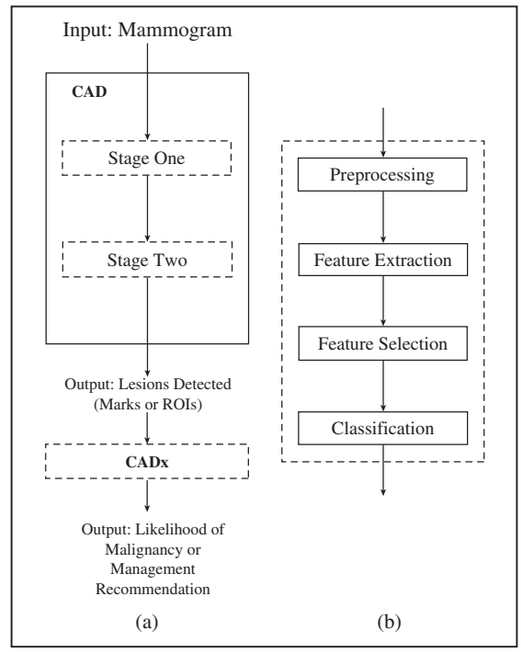


**FIGURE 3** A flowchart showing the main steps involved in the computer-aided detection (CAD) and computer-aided diagnosis (CADx) of mammographic abnormalities. Most detection algorithms consist of two stages. In stage 1, the aim is to detect suspicious lesions at a high sensitivity. In stage 2, the aim is to reduce the number of false positives without decreasing the sensitivity drastically. The steps that are involved in designing algorithms for both stages are shown in (**B**). We note that in some approaches some of the steps may involve very simple methods or be skipped entirely. For example, in stage 1, the classification step often is a simple size criteria (i.e., if the size of potential lesion is suspicious only if its size is greater than $N$ pixels). Most diagnosis algorithms (CADx) begin with a region of interest (ROI) containing the abnormality. Again, the steps typically involved in design such a system are shown in (**B**). The output of a CADx system may be the likelihood of malignancy or a management recommendation. Different research groups have worked on different components of the problem and human interaction may occur at various stages. For example, many CADx algorithms start with manually segmented ROIs.

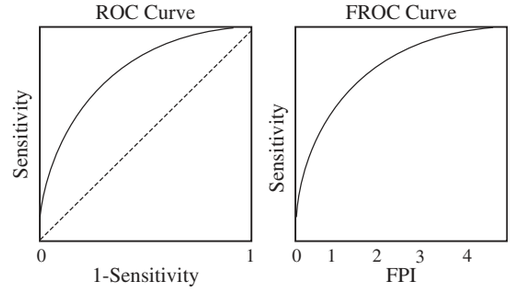


**FIGURE 4** Two plots illustrating receiver operating characteristic (ROC) and free-response receiver operating characteristic (FROC) curves. In an ROC curve, sensitivity is plotted on the y-axis and 1-Specificity or FPF is plotted along the x-axis. The dotted line in the ROC curve represents chance performance. In an FROC curve, sensitivity is plotted on the y-axis and the number of FPI is plotted along the x-axis. ROC curves are used for diagnosis studies and FROG curves are used for detection studies.

detection of malignant mass is most important. Whatever the methodology used, it is necessary for researchers to clearly state the reporting method that has been adopted.

Until recently, FROC analysis has been limited by the fact that the statistical analysis of FROC curves was less developed than that of traditional receiver operating characteristic (ROC). Major advances have recently been made in FROC analysis, particularly by Chakraborty and Berbaum [16]. However, despite the consistent use of evaluation methods in the literature, direct comparison of systems for detecting mammographic abnormalities is difficult because few studies have been reported on a common database. In Sections 2.1 and 2.2 we describe the different mass and calcification detection algorithms, respectively.

$$\text{Sens} = \frac{\text{Number of True-Positive Marks}}{\text{Number of Lesions}} \tag{1}$$

$$\text{FPI} = \frac{\text{Number of False-Positive Marks}}{\text{Number of Images}} \tag{2}$$

## 2.1 Detection of Masses

A mass is defined as a space-occupying lesion seen in at least two different projections [4]. Radiologists characterize masses by their shape and margin properties.

A number of researchers have worked on methods for detecting masses in mammograms. Masses with spiculated margins have a very high likelihood of malignancy and thus some methods have been developed specifically for the detection of spiculated masses. A spiculated mass is characterized by lines radiating from the margins of a mass [17]. However, since not all malignant masses are spiculated, the detection of nonspiculated masses is also important. Most mass detection algorithms consist of two stages: (a) detection of suspicious regions on the mammogram and (b) classification of suspicious regions as mass or normal tissue. These are described in Sections 2.1.1 and 2.1.2, respectively.

### 2.1.1 Stage 1: Detection of Suspicious Regions

The first stage is designed to have a very high sensitivity and a large number of false positives are acceptable since they are expected to be removed in stage 2. Algorithms for stage 1 detection can generally be considered to be of two types, pixel based or region based [18].

***2.1.1.1 Pixel-based Detection Methods.*** In pixel-based methods, features are extracted for each pixel and they are then classified as suspicious or normal. The terminology ''pixel-based'' is misleading since for every pixel, features are extracted from the local neighborhood of the pixel. This is followed by a classification step in which pixels are classified as suspicious or not. This may be done by simply applying a threshold to the feature image or by using sophisticated classification techniques. Finally, suspicious pixels are grouped together into regions, generally by collecting connected pixels. It is important to emphasize that regions labeled as suspicious by the detection algorithms are not necessarily malignant. The classification of detected regions into malignant or benign categories is a different problem. A brief summary of pixel based mass detection methods follows.

A number of detection methods have targeted particular subsets of masses. For example, some researchers have focused on the detection of spiculated masses because of their high likelihood of malignancy. The main idea behind the detection of spiculated masses is as that since spiculated masses are characterized by spicules radiating in all directions, one should compute the edge orientations at each pixel. Thus, each pixel is represented by a feature vector that represents the strongest edge orientation at the pixel. The edge orientation can be computed in a variety of different ways.

Kegelmeyer et al. [19] developed a method to detect spiculated masses using a set of five features for each pixel. They used the standard deviation of a local edge orientation histogram (ALOE) and the output of four spatial filters that are a subset of Laws texture features. The idea of using the ALOE feature is that a normal mammogram exhibits a tissue structure that radiates in a particular orientation (from the nipple to the chest). A spiculated mass would change this trend and thus normal tissue would have edge orientations in a particular direction, whereas in suspicious regions containing spiculated lesions, edges would exist in many different orientations. To detect this difference Kegelmeyer et al. computed edge orientations in a window around each pixel and then generated a histogram of edge orientations. This idea is depicted in Fig. 5. The ALOE feature was then defined as the standard deviation of the bin heights of the histogram and is described by Equation 3. Where $hist_{ij}$ is the histogram of edge orientations in a window around the pixel located at $(i, j)$, and $\text{hist}(i, j)$ is the average bin height of the histogram $hist_{ij}$. A binary decision tree was used to classify each pixel. The neighborhood size for computing the ALOE was chosen to be 4 cm so that it would encompass all of the spiculated masses in the dataset.

$$\text{ALOE feature } (\partial_{ij}) = \sqrt{\frac{\sum_{n=0}^{255} (hist_{ij}(n) - \overline{\text{hist}}(i,j))^2}{255}} \tag{3}$$

Karssemeijer and te Brake [20] detected stellate distortions by a statistical analysis of a map of pixel orientations. The orientation at each pixel was computed from the response of three filter kernels, which are second-order, directional derivatives of a Gaussian kernel in the directions $(0, \pi/3, 2\pi/3)$. These filters form a nonorthogonal basis and are shown in

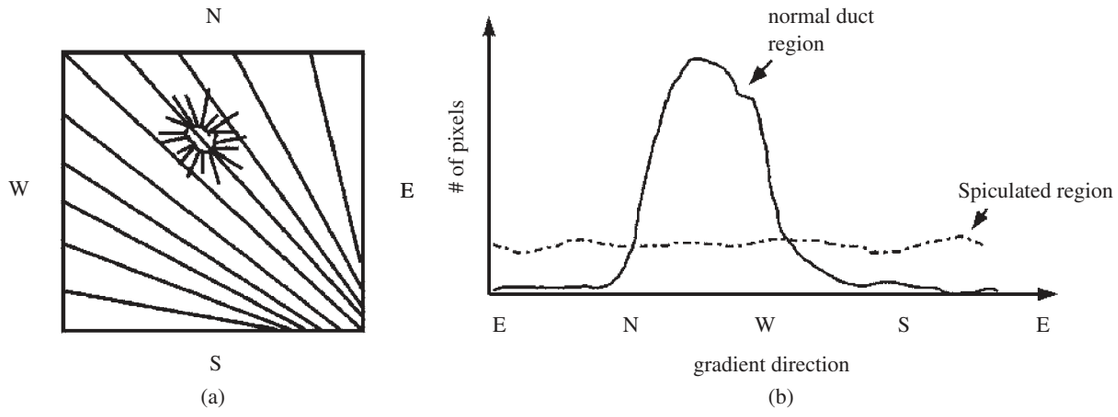


**FIGURE 5** (**A**) Directions of spicules of a spiculated lesion differ from the directions of normal linear markings in a mammogram and the (**B**) standard deviation of the gradient orientation histogram differentiates the area near a spiculated lesion from normal. The figure was obtained from [21] (©2004 IEEE).

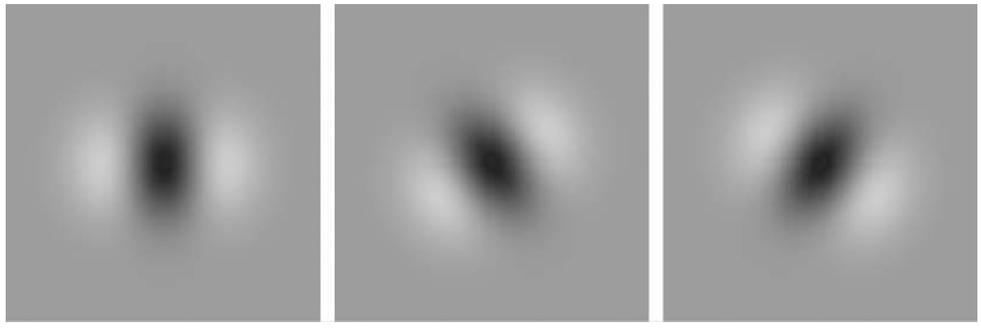

**FIGURE 6** Three-directional second-order Gaussian derivatives used for estimation of line orientation. The figure was obtained from Prof. Nico Karssemeijer [20] (©2004 IEEE).

Fig. 6. They used the relation that at a particular scale ($\partial$), the output at any orientation $W_\partial(\theta)$ can be expressed as a weighted sum of the responses of the filters. This is described in Equation 4, where $W_\partial(0)$, $W_\partial(\pi/3)$, and $W_\partial(2\pi/3)$ are the responses of the three filters. This relation was used to determine the orientation at each pixel and two features for each pixel were derived by a statistical analysis of these pixel orientation maps. The pixels were then classified as suspicious or normal. To account for the range of sizes of spiculations in their dataset, edge orientations were computed at three spatial scales ($\partial = 1, 2, 3$) and the one with the maximum magnitude was used. We note that this is equivalent to choosing local neighborhoods of varying sizes.

$$W_\partial(\theta) = \frac{1}{3}[1 + 2\cos(2\theta)]W_\partial(0) + \frac{1}{3}[1 - \cos(2\theta) + \sqrt{3}\sin(2\theta)]W_\partial(\pi/3) + \frac{1}{3}[1 - \cos(2\theta) - \sqrt{3}\sin(2\theta)]W_\partial(2\pi/3) \quad (4)$$

Liu et al. [21] point out that in general, it is difficult to estimate the size of the neighborhood that should be used to compute the local features of spiculated masses. Small masses may be missed if the neighborhood is too large and parts of large masses may be missed if the neighborhood is too small. To address this problem Liu et al. [21] developed a multiresolution algorithm for the detection of spiculated masses. They generated a multiresolution representation of a mammogram using the discrete wavelet transform. A detail description of the wavelet transform can be obtained in Chapter 4.2. They extracted four features at each resolution for each pixel. One of the features they used was the ALOE feature described in Equation 3. Pixels were then classified using a binary classification tree.

The detection was carried out in a top-down manner from the coarsest resolution to the finer resolutions. If a positive detection was made and a pixel was classified as abnormal, no feature extraction and detection were needed at the corresponding pixels at all finer resolutions. This approach reduced the number of pixels to be classified.

In all of the three methods described above, the focus was on developing sophisticated stage 1 detection techniques. These methods used very simple techniques for the stage 2 task. For example, Karssemeijer and te Brake [20] grouped suspicious regions and discarded regions that were smaller 500 pixels.

Other researchers have not restricted their efforts to the detection of spiculated masses since many malignant masses are not spiculated. Li et al. [22] developed a two-step process for detection of masses. In the first step, adaptive gray-level thresholding was used to obtain an initial segmentation of suspicious regions. The segmentation was iteratively improved using a multiresolution Markov random field (MRF)–based segmentation method. The algorithm was first applied at the coarsest resolution and the output was refined at the next-finer resolution. This strategy helps to reduce the computational complexity as mentioned above. A detailed description of MRF can be obtained in Chapter 4.3. In the second stage, a fuzzy binary decision tree was used to classify the segmented regions as masses or normal tissue using features based on shape, region size, and contrast.

Matsubara et al. [23] developed an adaptive thresholding technique for the detection of masses. They used histogram analysis techniques to divide mammograms into three categories ranging from fatty to dense tissue. Potential masses were detected using multiple threshold values based on the category of the mammogram. A number of features such as circularity, area, and standard deviation were used to reduce the number of false positives.

Li et al. [24] developed a method for lesion site selection using morphologic enhancement and stochastic model–based segmentation technique. A finite generalized Gaussian mixture distribution was used to model histograms of mammograms. The expectation maximization algorithm [25] was used to determine the parameters of the model. The segmentation was achieved by classifying pixels using a new Bayesian relaxation labeling technique. An underlying motivation for this technique was that it could incorporate neighborhood information into the classification process and that this would help improve the process. They argued that for the purpose of lesion site selection, sensitivity should be the sole criterion for evaluation and thus did not incorporate a false-positive detection step.

The primary advantage of using pixel-based methods is that one has a large number of samples to train a classifier. However, this class of methods also has inherent disadvantages. It does not take into account the spatial arrangement of the pixels, which is a very important factor to discriminate masses from normal tissue. A different set of features would be required to describe different mass types. It is computationally intensive and hence, most pixel-based methods must subsample images before detection.

The advantage of having many pixels per image available for use in training supervised learning methods should not be overstated. There are two problems regarding the use of multiple pixels. First, pixels at the periphery of a mass and at the center of the mass belong to the same class, but are not always homogenous and maybe represented by different feature values. This is a major limitation as ideally one would want samples of a particular class to possess similar feature values. Second, multiple pixels from a single mass represent only one particular lesion example. This does not eliminate the need for a comprehensive database containing masses that encompasses the range of natural variability of masses.

***2.1.1.2 Region-based Detection Methods.*** In region-based detection methods, regions of interest are first extracted by a segmentation or filtering technique. Features are then extracted for each region and the region is classified as suspicious or otherwise. These features are designed to describe important diagnostic information like shape and texture of the extracted regions.

A number of these methods are based on the idea of matched filtering. In these approaches, the image is filtered with a filter that is used as a model for a mass. The idea is that the output of the filtered image will be high near the center of the tumor masses. Often the $N$ largest outputs are selected as possible suspicious regions. This is followed by the extraction of ROIs around the $N$ largest peaks. Features are extracted from the ROI, and the ROIs are classified as containing a mass or normal tissue. Here lies the main difference between pixel- and region-based detection methods. In the pixel-based methods, features were extracted for each pixel, whereas in the region-based methods, features are extracted for each region. A brief description of the region-based methods that used a matched filtering approach is given below.

Kobatake et al. [26] modeled masses as rounded convex regions and based on this idea, developed an ''iris filter'' to enhance and detect masses. The iris filter was applied to a gradient image that was generated by Perwitt-type operators (see Chapter 4.13). The output of the filter was computed by measuring the average convergence of the gradient over the region of support of the filter. The peaks of the output of the filter were selected as centers of tumor candidates. The filter was then reapplied locally to detect the boundaries of candidate masses. Finally, texture features were computed from the candidates and were used to reduce false-positives. The authors showed that one of the advantages of using this filter was that the output of the filter would be constant regardless of the contrast between a rounded convex region and the background.

Petrick et al. [27] developed a two-stage algorithm for the enhancement of suspicious objects. In the first stage, they proposed an adaptive density-weighted contrast-enhancement (DWCE) filter to enhance objects and suppress background structures. The central idea of this filtering technique was that it used the density value of each pixel to weight its local

contrast. In the first stage, the DWCE filter and a simple edge detector (Laplacian of Gaussian) were used to extract ROIs containing potential masses. In the second stage, the DWCE was reapplied to the ROI. Finally, to reduce the number of false positives, they used a set of texture features for classifying detected objects as masses or normal. They further improved the detection algorithm by adding an object-based region-growing algorithm [28].

Polakowski et al. [29] used a single difference of Gaussian (DoG) filter to detect masses. The DoG filter was designed to match masses that were approximately 1 cm in diameter. ROIs were selected from the filtered image. They used nine features based on size, contrast, circularity and Laws texture features to reduce the number of false positives and to then classify ROIs as malignant or normal.

The DoG filter, which is a band-pass filter, has been used by several researchers for the preliminary task of detection of potential masses in an image. The DoG filter must be matched to the size of the mass. Since the size of masses varies from a few millimeters to several centimeters [17], a number of DoG filters would be required, which would increase the computational complexity. Since the size of a potential mass is not known a priori, several researchers have used multiscale region-based methods for the detection of masses.

Brzakovic et al. [30] use a two-stage multiresolution approach for detection of masses. First they identified suspicious ROIs using Gaussian pyramids (Chapter 4.2) and a pyramid linking technique based on the intensity of edge links. Edges were linked across various levels of resolution. This was followed by a classification stage, where the ROIs were classified as malignant, benign, or normal on the basis of features like shape descriptors, edge descriptors, and area.

Qian et al. [31] developed a multiresolution and multi-orientation wavelet transform for the detection of masses and spiculation analysis. They observed that traditional wavelet transforms cannot extract directional information, which is crucial for a spiculation detection task and thus, they introduced a directional wavelet transform. Figure 7 shows the partitioning of the frequency domain with the directional wavelet transform. We note that in comparison, a conventional wavelet transform would produce a rectangular partitioning of the frequency domain. An input image was decomposed into two output images using the directional wavelet transform. One was a smoothed version of the original image and was used to segment the boundary of the mass. The second contained the high-frequency information and was used for directional feature extraction. The key ideas of the method were that at coarser resolutions, features such as the central mass region can be easily detected, whereas at finer resolutions, detailed directional features such as spicules can be localized.

As was the case for pixel-based methods, some researchers have developed region-based methods that are focused on the detection of masses with particular margin characteristics,

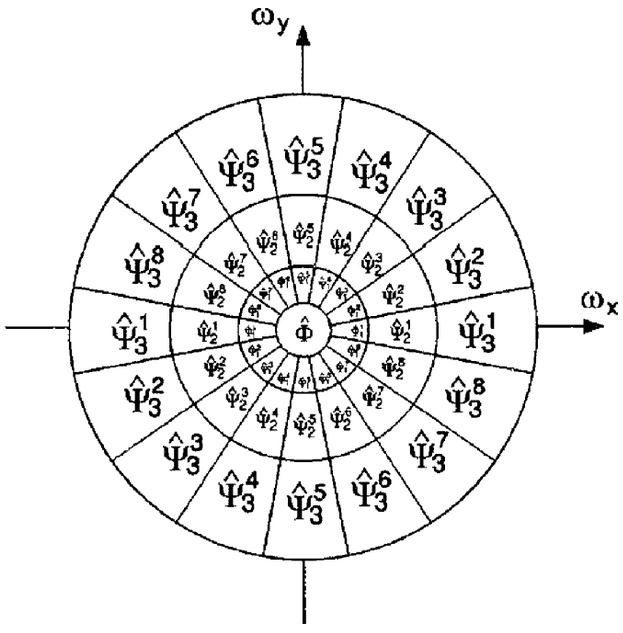


**FIGURE 7** Partitioning of the frequency domain achieved with the directional wavelet transform. The figure was obtained from [127] (©2004 IEEE).

such as circumscribed or spiculated masses. Lai et al. [32] developed a simple template matching algorithm to detect circumscribed masses only. They enhanced images using a modified median filtering technique to remove background noise. To cope with variations in the size of masses various templates with radii ranging from three to 14 pixels were used. To measure the similarity between a potential mass and the template, the authors chose the normalized cross-correlation as a similarity metric. This particular metric was chosen since it is invariant to the size of the template and the average brightness of the image. They also developed two features to reduce the number of false positives detected.

Groshong and Kegelmeyer [33] used the circular Hough transform for the detection of circumscribed lesions. The Hough domain, for circular objects consists of three parameters ($x$, $y$, and $r$) corresponding to the $x$ and $y$ centers and radius ($r$) of the object. Thus, a point in the three-dimensional Hough domain maps to a circle in the image domain. More details about the circular Hough transform can be found in [34]. They computed an edge image using a canny operator (Chapter 4.13) and selected a subset of the edges based on length and intensity. This subset of edges was the input to a circular Hough transform. The radius parameter search space ranged from 3 to 30 mm to account for masses of different sizes. Two features were extracted from the Hough domain for each pixel and ultimately these were classified as either belonging to a mass or normal tissue.

Zhang et al. [35] noted that the presence of spiculated lesions led to changes in the local mammographic texture.

They proposed that such a change could be detected in the Hough domain, which is computed using the Hough transform. They partitioned an image into overlapping ROIs and computed the Hough transform for each ROI. The Hough domain of each ROI was thresholded to detect local changes in mammographic texture and to determine the presence or absence of a spiculated mass.

Region-based methods have a number of advantages. In contrast to pixel-based methods, region-based detection takes into account the spatial information. Also, the features are directly correlated to important diagnostic information like the shape and margin of extracted regions. They are computationally less intensive than pixel-based methods. The main disadvantage is that if a classifier is used, there are fewer samples for training the classifier as compared to the pixel-based methods.

To conclude this section, we note that masses can have a range of sizes. Thus, a major limitation of both pixel-based and region-based methods is that the analysis is not done over a continuous range of scales. Cancerous lesions are stochastic biologic phenomena that manifest in images as having various structures occurring at different sizes and over ranges of spatial scales. For example, masses occupy definite regions; this region occupancy can be approached at a coarse scale of description or processing. However, the boundaries of masses require a more localized approach, although the sharpness, and hence the scales of interpretation of the lesion boundaries, can vary considerably. Moreover, the spiculations that are associated with many cancerous lesions occur with different widths, lengths, and densities, which suggests that their characterization will require analysis over scales.

### 2.1.2 Stage 2: Classification of Suspicious Regions as Mass or Normal Tissue

A number of researchers have focused solely on the second stage of detection in which suspicious regions are classified as mass or normal tissue. The purpose of the second stage is to reduce the number of false positives that were produced at the end of the first stage. A brief summary of stage two methods follows.

Researchers have used texture features to discriminate between mass and normal tissue. Sahiner et al. [36] proposed a convolution neural network for this task. They extracted texture features from the ROIs. Wei et al. [37] developed a classifier using texture features and linear discriminant analysis for this task. They computed multiresolution texture features from spatial gray-level dependence matrices. Wei et al. [38] also investigated the use of global and local multiresolution texture features for this task and for reducing the number of false-positive detections on a set of manually extracted ROI.

Radiologists use a number of image characteristics to discriminate between masses and normal tissue and researchers have attempted to emulate that process. Te Brake et al. [39] defined a number of features to discriminate between lesions and normal tissue that were designed to capture image characteristics like intensity, iso-density, location, and contrast. Kupinski and Giger [40] studied a regularized neural network for this task. Masses were detected using the bilateral subtraction scheme, which is described in Section 2.1.3. Features based on geometry intensity and the gradients of potential lesions were extracted. They also evaluated the effectiveness to minimize overtraining. Mutual information and a subregion hotelling observer have also been tested for this classification problem. Tourassi et al. [41] developed a template-matching technique for this problem. Each ROI in the database served as a template and mutual information was used a similarity metric to decide if a query ROI contained a mass. Baydush et al. [42] proposed a subregion hotelling observer for detecting whether a given ROI contained a mass or not.

### 2.1.3 Mass Detection Using Multiple Images

In the methods described in the previous section, the analysis was performed on the MLO, CC view images, or the images of the left or right breast only. That is, these methods use only a single image. Some researchers have also developed methods that use more than one image.

In one approach, the same view (CC or MLO) is used from both the left and right breasts. The intuition behind this method is that radiologists use the architectural symmetry between left- and right-breast images in the visual analysis of mammograms. Asymmetry between the left and right breast may correspond to potential abnormalities. One such mass detection method is based on the analysis of the symmetry between the corresponding mammograms of each breast. Yin et al. [43] computed the difference of the left- and right-breast images and binarized the difference image at various thresholds to detect deviations from architectural symmetry and thus possible masses. They described their method as a ''nonlinear bilateral subtraction'' scheme.

Radiologists also look at the current and prior mammograms (from previous exams) of a patient while performing a visual analysis of the mammograms. Thus, researchers have tried to use both the current and prior mammograms for the detection of potential abnormalities. One such method was developed by Zouras et al. [44], who investigated the potential of incorporating a temporal subtraction scheme to the bilateral subtraction technique.

A number of difficult preprocessing steps are required in both of the schemes mentioned above. For the bilateral subtraction scheme, a registration or alignment of the left and right breast must be performed before computing the difference. This is difficult because the left and right breasts are not exactly symmetric in shape and size on the

**TABLE 1** Summary of representative selection of mass detection algorithms

| Mass Type | Author | Method | No. Images | Stage One | | Stage Two | |
|---|---|---|---|---|---|---|---|
| | | | | TP | FPI | TP, % | FPI |
| All | Yin et al., 1991 | Pixel | 46 | — | — | 95 | 3.2 |
| All | Li et al., 1995 | Pixel | 95 | — | — | 90 | 2 |
| All | Zouras et al., 1996 | Pixel | 79 | — | — | 85 | 4 |
| All | Matsubara et al., 1996 | Pixel | 85 | — | — | 82 | 0.65 |
| All | Li et al., 2001 | Pixel | 200 | 97.3% | 14.81 | — | — |
| All | Petrick et al., 1996 | Region | 168 | 95.5% | 20 | 90 | 4.4 |
| All | Kobatake et al., 1999 | Region | 1214 | — | — | 90.4 | 1.3 |
| All | Brzakovic et al., 1990 | Region | 25 | — | — | 85 | — |
| All | Qian et al., 1999 | Region | 100 | — | — | 96 | 1.71 |
| Circumscribed | Lai et al., 1989 | Region | 17 | — | — | 100 | 1.7 |
| Circumscribed | Groshong et al., 1996 | Region | 44 | — | — | 80 | 1.34 |
| Spiculated | Kegelmeyer et al., 1994 | Pixel | 86 | — | — | 100 | 82% specificity |
| Spiculated | Karssemeijer et al., 1996 | Pixel | 50 | — | — | 90 | 1 |
| Spiculated | Liu et al., 2001 | Pixel | 38 | — | — | 84.2 | 1 |
| Spiculated | Zwiggelaar et al., 1998 | Pixel | 54 | — | — | 70 | 0.01 |
| All | Polakowski et al., 1997 | Region | 254 | 92% | 8.39 | 92 | 1.8 |

The input to these algorithms is the mammogram. Most detection algorithms consist of two stages. In the first stage the goal is to achieve high sensitivity. The second stage aims to reduce the number of false positives per image (FPI) without drastically reducing the sensitivity. It is not possible to make a comparison between these different algorithms since they have not been trained and tested on the same datasets. TP, true positive.

mammograms. In the temporal subtraction scheme, a similar alignment of the prior and current images must be performed.

Although the idea of searching for architectural asymmetries is appealing, current methods that try to use this concept are based on very simple subtraction-based techniques and compute a number of thresholds in an ad hoc manner. Choosing an appropriate threshold that would work across a large set of images is very difficult.

Table 1 gives a summary of various mass detection algorithms. Most authors do not report the performance for stage 1 of the detection algorithm. This is unfortunate, since comparisons between the stage-by-stage performance of different algorithms cannot be made. It also makes it more difficult to combine stages of different algorithms to achieve better performance. For example, given algorithms A and B, it may be that a third algorithm C, which is composed of the first stage of algorithm A and the second stage of algorithm B, could yield significantly better performance than either algorithm A or B. However, if the ''stagewise'' performance of the algorithms is not reported, then the combined algorithm C may not be conceived of or attempted.

## 2.2 Detection of Calcifications

Calcifications are small calcium deposits that form in the breast as a result of benign or malignant processes. Mammographically, they appear as bright white spots of various sizes and shapes. The important characteristics of calcifications are their size, shape or morphology, number, and distribution.

One of the main characteristics to consider in the detection of calcifications is that they are generally very small. Their size varies from 0.1 mm to 1 mm and the average diameter is 0.3 mm [17]. Small calcifications may be missed due to the overlapping breast parenchyma. Another issue is that in regions where the background tissue is dense, it is very difficult to localize the calcifications. Finally, calcifications sometimes have a low contrast to the background and can be mistaken as noise in the inhomogeneous background.

On the other hand, the high degree of localization of calcifications makes them somewhat easier to model (they are ''impulselike''), and indeed, a number of robust methods have been developed for the detection of calcifications, and a great deal of success has been achieved with these methods. The detection performance of current commercial systems is reported at 95% sensitivity at less than 1 FPI [86].

A number of different approaches have been applied for the detection of calcifications. Calcifications represent high spatial frequencies in the image. Thus, one approach to the calcification detection task is to localize the high spatial frequencies of the image. The wavelet transform is an optimal tool for such a task as compared with other transforms such as the Fourier transform, which only gives information on the frequency content and cannot spatially localize the frequencies. Thus, a number of authors have used wavelet transforms for the detection of microcalcifications [45–49]. In these

methods, the image is first processed by a subband decomposition filterbank. The coefficients in the subband images that correspond to high spatial frequencies are selectively weighted to enhance the calcifications. A new image with enhanced calcifications is created with the inverse wavelet transform. The calcifications are then detected using global and local thresholds. Finally, the individual calcifications are then grouped together to detect clusters. A summary of some methods that have used the wavelet transform is given below.

Strickland and Hahn [45] proposed a method using undecimated bi-orthogonal wavelet transforms and subband weighting to detect and segment clustered microcalcifications. Yoshida et al. [46] used undecimated wavelet transforms and supervised learning for calcification detection. Zhang et al. [47] developed a method to optimize the weights at individual scales of the wavelet decomposition. Qian et al. [48] used a tree-structured wavelet transform for multiresolution decomposition and selective reconstruction of subimages to segment microcalcifications. They used a nonlinear filter for suppressing image noise.

In most of the methods, the detection is carried out in the spatial domain. However, Gurcan et al. [49] performed the detection in the subband image domain. The key aspect of their method was that calcifications would produce outliers in the high-pass and band-pass subimages. Thus, the symmetry of the distribution of the band-pass and high-pass image coefficients is altered in regions containing the microcalcifications. The changes in the distribution were captured by computing the skewness and kurtosis of the distribution.

Another reason that wavelets have been so effective is that calcifications appear as small bright dots on the mammogram and can be viewed as point discontinuities. Recently, mathematicians have argued that wavelets have finite square supports and are ideal for capturing point discontinuities but not edges [50]. We believe that this fact intuitively explains the tremendous success of wavelet transform–based methods in the detection of calcifications and why they have been less successful for the detection of masses.

In addition to wavelets, other multiscale methods have been investigated. Netsch and Peitgen [51] proposed a multiscale detection method based on the Laplacian of Gaussian filter and a mathematic model. They used scale-space signatures obtained from Laplacian filtering for the detection of clustered microcalcifications.

Other non–wavelet-based methods have also been developed for the detection of calcifications. These methods generally try to make maximum use of the fact that calcifications have much higher intensity values than the surrounding tissue in a mammogram. These methods are more likely to fail when the calcifications are present in dense background tissue.

Chan et al. [52] used a difference–image processing technique to detect calcifications. In this methodology, they computed ''signal-enhanced'' and ''signal-suppressed'' images and subtracted these to obtain a difference image. Global and local level thresholding was then used to extract potential calcifications. In a later study [53], they incorporated an artificial neural network to reduce the number of FP clusters per image.

Davis and Dance [54] used local area thresholding to detect calcifications. Although they showed that this method was successful on a small test set, in general picking a threshold that will work successfully on a large set of images is extremely difficult.

Nishikawa et al. [55] combined the difference image technique with morphologic erosion filters and gray-level thresholding techniques to extract microcalcifications. To reduce the number of false positives, Zhang et al. [35] applied a shift-invariant artificial neural network. Zheng et al. [56] developed a multistage algorithm including Gaussian filtering, nonlinear global thresholding for calcification detection. They used a mixed feature-based neural network for detection.

Many of the methods described above had a false-positive reduction stage built in. A number of authors have focused on developing techniques to reduce the FPs. This is similar to the stage two of the mass detection methods. The main aim here is to classify ROIs as either containing calcifications (positive ROI) or normal tissue (negative ROI). Various schemes have been developed for this purpose.

Kim et al. [57] examined the performance of different statistical textures for this task. They compared the performance of a new texture analysis method called surrounding region-dependence method (SRDM) with gray-level co-occurrence matrix method (GLCM), gray-level run-length method (GLRLM), and gray level difference method (GLDM). They reported that the SRDM achieved the highest classification accuracy for this task. These texture analysis methods are co-occurrence-based methods, which attempt to characterize second-order properties of an image. In these methods, the general idea is to measure the joint probability of two image properties occurring under certain conditions. The image properties could be spatial gray-level intensities, intensity differences, and so forth. These properties are measured under specific conditions such as pixel spacing (magnitude and orientation), magnitude of intensity differences, and so forth. For example, the GLCM measures the probability of co-occurrence of image pixel intensities ''i'' and ''j'' under the condition of ''d'' pixels separation between the pixels. By comparison, the GLRLM estimates the probability of image pixels with intensity ''i'' occurring in a colinear sequence of length ''j.''

Nagel et al. [58] compared the performance of three methods for reducing false-positives. For the false-positive reduction task they examined a rule-based method, neural network, and a combined method that used both of these techniques. They reported that the combined method was more successful in eliminating false-positives because each of the two stages eliminated different types of false-positives.

An enhancement stage often precedes the detection step. This can be global or local fixed-neighborhood enhancement scheme. Some of the simple global enhancement techniques used are contrasting stretching and histogram equalization. In comparison, the local enhancement techniques use properties like the mean and standard deviation in a fixed neighborhood of the pixel to estimate the background. This method increases the local contrast by suppressing the background. One of the main limitations of these methods is that they enhance not only the calcifications but also the background structure and noise. Another major difficulty is the choice of metric to report the performance of the enhancement algorithm. Some of the metrics used to measure the performance of the enhancement algorithms are defined below:

$$\text{Contrast(C)} = \frac{f - b}{f + b} \tag{5}$$

where $f$ is the mean gray-level value of a particular object in the image called the foreground and $b$ is the mean gray-level value of a surrounding region called background.

$$\text{CII} = \frac{C_{\text{processed}}}{C_{\text{original}}} \tag{6}$$

Where $C_{\text{processed}}$ and $C_{\text{original}}$ are the contrasts in the processed and original images, respectively.

Li et al. [59] compared a fractal-based enhancement method with wavelet and morphologic enhancement methods. They used three metrics contrast improvement index (CII), peak signal-to-noise ratio (PSNR), and average signal-to-noise ratio (ASNR) [59]) to compare the performance of these three enhancement methods. However, they did not perform detection on the enhanced images. It is important to realize that defining an optimal metric to evaluate an enhancement algorithm is still an open research problem. Thus, it may be more appropriate to evaluate the performance of the enhancement algorithms by studying the effect of the enhancement on the accuracy of the detection algorithm that generally follows it.

## 2.3 Conclusion

We have attempted in Tables 1 through 4 to summarize the approaches that have been taken in CAD of masses and calcifications. However, we would like reiterate that it is

**TABLE 2** Summary of representative selection of mass detection algorithms applied on regions of interest (ROIs)

| Author | No. ROI | Detection Results | | |
|---|---|---|---|---|
| | | TPF (Sensitivity) | FPF (1-Specificity) | Az (Area under ROC Curve) |
| Wei et al. [37] | 678 | — | — | 0.86 |
| Wei et al. [38] | 168 | — | — | 0.92 |
| Kupinski et al. [40] | 495 | — | — | 0.93 |
| Tourassi et al. [41] | 1465 | — | — | 0.87 |
| Baydush et al. [42] | 1320 | 98% | 55.9% | 0.94 |
| Sahiner et al. [36] | 678 | 90% | 31% | 0.87 |

The inputs to these algorithms are ROIs which may be obtained automatically or manually. The aim is to determine if a given ROI contains a mass or normal tissue. It is not possible to make a comparison between these different algorithms since they have not been trained and tested on the same datasets. The TPF is sensitivity as defined in Eq. 1. The FPF is 1-specificity where specificity = number of true negative [normal] classifications/number of negative [normal] ROIs.

**TABLE 3** Summary of representative selection of calcification detection algorithms

| Author | No. Images | Stage One | | Stage Two | |
|---|---|---|---|---|---|
| | | TP | FPI | TP, % | FPI |
| Strickland and Hahn [45] | 40 | — | — | 0.9–0.99 | 3–7 |
| Yoshida et al. [46] | 39 | — | — | 95 | 1.5 |
| Qian et al. [48] | 100 | — | — | 94 | 1.6 |
| Gurcan et al. [49] | 40 (105 clusters) | — | — | 100 | 3.3 |
| Netsch and Peitgen [51 ] | 40 (105 clusters) | — | — | 0.84 | 1 |
| Chan et al. [52] | — | — | — | 80 | 1 |
| Chan et al. [53] | 52 | 100% (obvious cases) | 0.35 | 100 | 0.1 |
| | | 87% (subtle cases) | 4 | 90 | 1.5 |
| Davis and Dance [54] | 50 | | | 100 | 0.08 |
| Zhen et al. [56] | 55 | — | — | 100 | 0.18 |
| Nagel et al. [58] | 50 | — | — | 83 | 0.8 |

Most detection algorithms consist of two stages. In the first stage the goal is to achieve high sensitivity. The second stage aims to reduce the number of false-positives per image (FPI) without drastically reducing the sensitivity. It is not possible to make a comparison between these different algorithms since they have not been trained and tested on the same datasets. TP, true positive.

**TABLE 4** Summary of representative selection of calcification detection algorithms applied on regions of interest (ROIs)

| Author | No. ROI | Detection Results | | |
|---|---|---|---|---|
| | | TPF (Sensitivity) | FPF (Specificity) | Az (Area under ROC Curve) |
| Zhan et al. [47] | 158 | — | — | 0.92 |
| Zhang et al. [35] | 50 | — | — | 0.90 |
| Kim et al. [57] | 172 | — | — | 0.92 |

The inputs to these algorithms are ROIs, which may be obtained automatically or manually. The aim is to determine if a given ROI contains calcification or normal tissue. It is not possible to make a comparison between these different algorithms since they have not been trained and tested on the same datasets.

extremely difficult to make a fair comparison of the different algorithms as they are often evaluated with private or local databases. The performance of a detection method can very dramatically over different databases. The performance depends on factors such as the subtlety of cases and number of cases. The choice of the evaluation technique that is used to measure the performance is another crucial factor.

# 3 Computer-Aided Diagnosis of Mammographic Abnormalities

The ultimate aim of the CADx task is to help the radiologist in making recommendations for patient management. If a mass is suspected to be malignant, a biopsy must be performed. If not, the patient is either scheduled for a short-term follow-up or is returned to the normal screening population. If the information is insufficient for the radiologist to make a decision, special radiographic views are taken and/or complementary modalities like ultrasound or magnetic resonance imaging (MRI) are used to obtain additional information. A flowchart showing the steps involved in the diagnosis of abnormalities is shown in Fig. 3.

The diagnosis task is modeled as a two-class classification task. Features are extracted from ROIs containing the abnormality, and each ROI is classified using a classification algorithm such as a neural network. In most cases, the classification algorithm used is a supervised method that is first trained on a set of sample cases called the training set. The performance of the algorithm is then tested on a separate testing set. Here, care must be taken that there is no overlap between the training and testing sets. Special data sampling techniques like cross-validation are often used in the developmental stage. This topic is out of the scope of this chapter and the interested reader can refer to any of one of several excellent texts in machine learning or pattern recognition [25]. The challenges in designing robust classifiers in the context of medical imaging decision support systems have been clearly summarized by Dodd et al. [60].

The metrics used to report the accuracy of these algorithms are sensitivity and specificity. Sensitivity was previously defined in our discussion of the evaluation of detection systems (Equation 1). However, the use is slightly different here in that a true-positive classification is defined as a lesion for which the CAD predicts that it is cancerous and it is actually found to be malignant. Specificity is the fraction of benign lesions that are correctly identified by the CAD as being benign (Equation 6).

$$\text{Sensitivity} = \frac{\text{Number of True Positive Classifications}}{\text{Number of Malignant Lesions}} \quad (7)$$

$$\text{Specificity} = \frac{\text{Number of True Negative Classifications}}{\text{Number of Benign Lesions}} \quad (8)$$

A plot of sensitivity versus 1-specificity is called a receiver operating characteristic (ROC) curve and this is generally used to report the performance of the diagnosis algorithm [61, 62]. An example of an ROC curve is shown in Fig. 4. It is important to note that the ROC methodology can be correctly applied in classification tasks where localization of the abnormality is not an issue [63] like in the diagnosis task described above. However, for tasks where localization is an important issue, the ROC methodology has some inherent problems as it does not require correct localization of the abnormality. Also the ROC does not apply to situations where the radiologist has to detect and localize multiple lesions on the same image. For these situations the FROC curve should be used to report performance. An example of an FROC curve is shown in Fig. 4. The next two sections describe in detail the steps involved in the diagnosis of masses and calcifications.

## 3.1 Diagnosis of Masses

Radiologists characterize masses by their shape and margin properties. The mass shape can be round, oval, lobular, or irregular and the mass margin or boundary can be circumscribed, microlobulated, obscured, indistinct, or spiculated [4]. Figure 8 shows a schematic diagram of some mass shapes and boundary characteristics. We also note that masses with spiculated and indistinct boundaries have a greater probability of malignancy than circumscribed masses.

Most diagnosis algorithms begin with a region of interest (ROI) containing a suspicious mass. The ROI may have been selected manually by an experienced radiologist or may have been automatically selected. Most techniques developed for the diagnosis of masses consist of three stages: (a) segmentation of mass boundary in ROI, (b) feature extraction, and (c) classification. In the segmentation stage, the mass is segmented from the background normal tissue. Following this, features

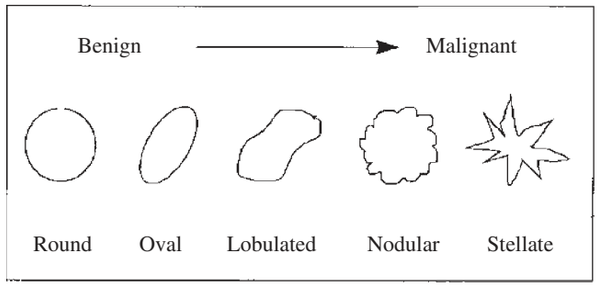


**FIGURE 8** Morphologic spectrum of masses. Figure obtained from [76] (©2004 IEEE).

that capture the differences between malignant and benign masses are extracted. Finally, masses are classified as malignant or benign. Each of these three stages is discussed in detail below.

### 3.1.1 Segmentation

The segmentation of the mass may be manual, semiautomated, or fully automated [64]. We will focus only on fully automated segmentation methods in this section. The segmentation of masses is extremely important as the success of the classification algorithm depends on this stage. The two major categories of segmentation methods are region growing and discrete contour models [65].

The area overlap measure is used to quantitatively report the performance of the segmentation algorithm. The area overlap measure is defined as the ratio of the area of the mass segmented automatically to the area of the mass segmented manually by an experienced radiologist.

However, there is interobserver variability in the manual segmentation of masses by radiologists. Thus, it has been suggested that it is more appropriate to observe if the variance between the boundaries marked by a radiologist and a computer falls within the variance between radiologists, rather than to measure the absolute difference of the computer boundary from any one of the radiologists.

Timp and Krassemeijer [65] proposed a new segmentation algorithm based on dynamic programming. They demonstrated that their algorithm performed better than two other segmentation methods (namely region growing and the discrete contour model). Guliato et al. [66] used fuzzy region growing methods for segmenting masses and classified the masses as malignant or benign based on the transition information around the segmented region. Kinoshita et al. [67] used shape and texture features based on gray-level co-occurrence matrices and a three-layer neural network for classification. Kupinski and Giger [68] proposed two extended region growing techniques for the segmentation of masses. One was based on the radial gradient and the other was based on simple probabilistic models.

### 3.1.2 Feature Extraction

Different shapes and margins have different likelihoods of malignancy. For example, masses with spiculated or indistinct margins have a higher probability of disease than masses with circumscribed margins. Similarly, irregular-shaped margins have a higher likelihood of malignancy than round margins [69]. Thus, for the diagnosis task, most features are designed to capture the shape and margin characteristics of masses. These features can be organized into (a) morphologic features and (b) texture features.

Morphologic features are directly inspired by characteristics for which a radiologist looks. On the other hand, texture features have been designed to capture important differences between malignant and benign masses that may not be evident to human eye. Thus, texture features have the potential to capture characteristics that are important diagnostically but are not easily extracted visually.

To extract textural information, a number of researchers have used the Haralick texture features. These features are computed from GLCMs. While it may be beneficial to extract texture-related features, it is important to note that this particular method of computing texture features has a number of limitations. Most mammograms are digitized at 12 bits per pixel (bpp). One cannot compute the texture features on these images directly because the co-occurrence matrices that these methods generate would be very large and sparse. Thus, the statistical features derived from them would not be reliable. Thus, most researchers quantize the image to 6 bpp before computing these features. However, a lot of information is lost during the quantization process and thus the features may not be reliable.

As Tuceryan and Jain [70] have discussed, the other major limitations are that there is no well-established method of selecting the displacement vector (d) and computing co-occurrence matrices for different values of the vector is not feasible. In addition, for a given vector, a large number of features can be computed from the co-occurrence matrix, and thus some form of feature selection method must be used to select the subset of the most discriminatory features.

### 3.1.3 Classification

Various classification techniques have been used for classifying masses as malignant or benign. Most of the techniques used are supervised methods. Artificial neural networks and linear discriminant analysis are some of the most popular techniques. A review of these methods can be obtained from machine learning textbooks [25, 71]. Brief summaries of methods for diagnosis of masses and their major themes follow.

Some authors have extracted texture and gradient features in a transform domain rather than in the spatial domain. The main intuition here is that masses can be better differentiated in the transform domain. That is, the features computed in

the transform domain would be more discriminatory than features computed in the spatial domain.

Sahiner et al. [72] proposed the rubber band straightening transformation (RBST) to transform a band of pixels surrounding the mass to a rectangular strip. They extracted texture features from the RBST image based on the SGLD matrices to classify masses as benign or malignant. They used a clustering algorithm and active contour models for segmentation. The main difficulty here is the accurate extraction of a band of pixels around the segmented mass. Claridge and Richter [73] used the Polar coordinate transform (PCT) to map lesions into polar coordinates. A spiculation measure was then computed from the PCT images to discriminate between circumscribed and spiculated masses. Hadjiiski et al. [74] classified masses as benign or malignant using texture features computed from the RBST image. They tested the performance of a hybrid classifier consisting of an adaptive resonance theory network cascaded with LDA. They used a set of manually segmented ROIs and reported a higher accuracy with the hybrid classifier than with a back propagation neural network or LDA. Pohlman et al. [75] used six morphologic features to classify masses as benign or malignant. To segment the lesions, they used an adaptive region growing technique, which required the selection of manual seed points.

Most studies focus on classifying masses as malignant or benign, however, some authors have also investigated the classification of masses into other categories. Bruce and Adhami [76] classified manually segmented masses as round, nodular, or stellate using the wavelet transform modulus maxima method. They used multiresolution shape features and LDA for classification. Kilday et al. [77] classified tumors as fibroadenoma, cyst, or cancer using a linear discriminant function. They used a set of seven shape features based on the radial distance measures (RDMs) from the centroid to the points on the boundary and tumor circularity. Rangayyan et al. [78] used morphologic features to characterize the roughness of tumor boundaries. The features used were moments of distances of boundary points from the center, Fourier descriptors, compactness, and chord-length statistics. They performed two experiments: classification of masses as circumscribed or spiculated and classification as benign or malignant. They achieved higher accuracy for the first task.

## 3.2 Diagnosis of Calcifications

Calcifications in mammograms may be observed individually or in clusters. Individual calcifications are less worrisome than clustered calcifications. A cluster is defined as a group of three or more calcifications within a 1 $cm^2$ area [17]. Radiologists characterize clusters of calcifications by distribution and the morphology of the calcifications [4]. The distribution can be diffuse, regional, segmental, linear, or clustered. Fourteen categories for describing the calcification morphology have been defined [17]. Some of the most common for biopsied lesions are dystrophic, punctate, indistinct, pleomorphic, or fine branching [17].

Most diagnosis algorithms begin with an ROI containing a cluster of calcifications. The ROI may have been selected manually by an experienced radiologist or may have been automatically selected. A flowchart showing the steps involved in the diagnosis of calcifications is shown in Fig. 3. The steps are (a) segmentation of individual calcifications, (b) feature extraction, and (c) classification.

### 3.2.1 Segmentation of Individual Calcifications

Segmentation is the most difficult step in the computer-aided diagnosis of calcifications. The extremely small size of calcifications makes this problem worse. The motivation for performing segmentation is that features are then extracted from the individual calcifications. As explained in the next section, there are two categories of features: individual calcification features and calcification cluster features. Radiologists do not look at every individual calcification to make a diagnosis but tend to focus on the global properties of a cluster to make a diagnosis. Thus, it maybe more appropriate to develop better cluster features than to develop new methods for individual calcification segmentation.

### 3.2.2 Feature Extraction

As was the case for masses, the features used for the diagnosis of calcification can be viewed as capturing morphologic or texture information. Researchers have reported that morphology is one of the most important clinical factors in calcifications diagnosis [79]. However, it is important to note that the accuracy of diagnosis algorithms which use morphologic features depends on the robustness of the segmentation algorithm. Features for calcification classification can also be organized in terms of whether they describe properties of the cluster as a whole or of the individual calcifications that make up the cluster. Some of the common cluster features include the number of microcalcifications, the mean microcalcification area, standard deviation of the microcalcification contrast, and the number of microcalcifications per unit area.

### 3.2.3 Classification

As for masses, a variety of classifiers have been used to automatically discriminate between benign and malignant clusters of microcalcifications. A summary of the important algorithms for calcification CADx follows.

Chan et al. [80] used morphologic and texture features for the classification of calcifications. The rationale for using texture features was that a malignancy would cause changes in the texture of the tissue surrounding it. Texture features were extracted from the SGLD matrices, and morphologic features consisted of both individual calcifications features and cluster features. They also used a genetic algorithm and LDA to select the best subset of features. Feature selection is an important

task, but neither of these techniques search the space of all possible subsets of features and hence do not necessarily give the most optimal feature subset.

Santo et al. [81] combined the output of multiple classifiers to classify calcifications as malignant or benign. Some of the classifiers classified individual microcalcifications of a cluster, whereas other classified the entire cluster. Thus, they used features to characterize single calcifications and clusters. The final output was a weighted combination of the outputs of the various classifiers used.

Tsujii et al. [82] used the Karhenen-Loeve (KL) transform or PCA to reduce the dimensionality of the feature set and applied a trend oriented radial basis function neural network to classify calcifications. It is important to note that while the KL transform is optimal in the mean square sense, it does not necessarily produce the most discriminatory features. KL would be optimal if the final goal was image or data compression. However, since in this case, the final goal is pattern classification, KL is not ideal and other techniques such as multiple discriminant analysis may be more appropriate for this task.

Morphology is regarded as the most important clinical factor for the diagnosis of calcifications and a number of authors have developed shape features for diagnosis. Kallergi [79] developed a classification method that used only morphologic features. These features were designed after an in-depth study of the clinical characteristics of calcifications and produced very impressive results [79]. Veldkamp et al. [83] used cluster shape features, cluster position features, and distribution features for the classification of calcifications. They used a sequential forward selection procedure for feature selection. Shen et al. [84] developed a set of shape features for classifying calcifications as malignant or benign. The features used were compactness, moments, and Fourier descriptors. The Fourier descriptors were defined as the Fourier coefficients of boundary pixels. One of the advantages of using these features is that they are translation, rotation, and scale invariant.

One limitation of using shape features is that the success of the shape features is very dependent on the accuracy of the segmentation algorithm. Sometimes, calcifications may have poor contrast and so the segmentation may not be very accurate. Thus, researchers have also developed algorithms that do not use shape features. Dhawan et al. [85] proposed image structure features for classification of calcifications. To represent the global texture they used second-order histogram based features and wavelet based features to represent the local texture. They also proposed cluster features based on the first order histograms of segmented clusters. Optimal features were selected using multivariate cluster analysis and a genetic algorithm based search method.

## 3.3 Conclusion

A number of mass CADx algorithms have been developed. Although, these methods have not been integrated into commercial systems yet, it is interesting to note that the reported performances of mass CADx algorithms are often been better than that of mass CAD algorithms. On the other hand, the performance of calcification CAD algorithms is much better than the performance of calcification CADx algorithms, probably due to the fact that segmentation is a very difficult task.

We have attempted in Tables 5 and 6 to summarize the approaches that have been taken in CADx of masses and calcifications. However, we would like reiterate that it is extremely difficult to make a fair comparison of the different algorithms as they are often evaluated with private or local databases. The performance of a CADx algorithm can very dramatically over different databases. The performance depends on factors such as the subtlety of cases and number

**TABLE 5** Summary of representative selection of mass diagnosis algorithms

| Author | No. Images | Diagnosis Results | | |
|---|---|---|---|---|
| | | TPF | FPF | Az (Area under ROC Curve) |
| Sahiner et al. [72] | 168 | — | — | 0.94 |
| Bruce and Adhami [79] | 60 | 80% | — | — |
| Kinoshita et al. [67] | 92 | 81% | — | — |
| Hadjiiski et al. [74] | 348 | — | — | 0.81 |
| Pohlman et al. [75] | 51 | — | — | 0.76–0.93 |
| Kilday et al. [77] | 82 | 69% | — | — |
| Rangayyan et al. [78] | 39 | 95% | — | — |

The inputs to these algorithms are ROIs which may be obtained manually or automatically. It is not possible to make a comparison between these different algorithms since they have not been trained and tested on the same datasets.

**TABLE 6** Summary of representative selection of calcification diagnosis algorithms

| Author | No. of Images | Diagnosis Results | | |
|---|---|---|---|---|
| | | TPF, % | FPF | Az (Area under ROC curve) |
| Chan et al. [80] | 145 | — | — | 0.89 |
| Santo et al. [81] | 102 | 75.7 | 73.5% | 0.79 |
| Tsujii et al. [82] | 128 | — | — | 0.76 |
| Kalleri [79] | 100 | 100 | 85% | 0.98 |
| Veldkamp et al. [83] | 280 | — | — | 0.83 |
| Shen et al. [84] | 143 | 100 | — | — |
| Dhawan et al. [85] | 191 | — | — | 0.96 |

The input to these algorithms are ROIs, which may be obtained manually or automatically. It is not possible to make a comparison between these different algorithms since they have not been trained and tested on the same datasets. TPF is sensitivity as defined in Eq. 7. FPF is 1-specificity as defined in Eq. 8.

of cases. The choice of the evaluation technique that is used to measure the performance is another crucial factor.

# 4 Commercial Computer-Aided Detection Systems

Three FDA-approved commercially available CAD systems have been developed to aid radiologists in detecting mammographic abnormalities. Currently, there are no FDA-approved systems for CAD.

## 4.1 R2 Technology, Inc.

R2 Technology's Image Checker® was the first commercial mammographic CAD system approved by the FDA [86]. This device is designed to search for signs that may be associated with breast cancer. Masses are marked with an asterisk, while microcalcification clusters are marked with a triangle. The detection accuracy of calcifications was reported as 98.5% sensitivity at 0.74 false positives per case (set of four images). The detection accuracy of masses was reported as 85.7% at 1.32 false-positive marks per case.

## 4.2 Intelligent Systems Software, Inc.

The FDA approved the Intelligent System Software Inc. (ISSI) CAD system MammoReader™ in 2002. MammoReader was designed to detect primary signs of breast cancer in mammogram images including microcalcification clusters, well- and ill-defined masses, spiculated lesions, architectural distortions, and asymmetric densities. Masses are marked with crosshairs and microcalcification clusters with outlines [87]. The reported overall sensitivity was 89.3% (91.0% in cases where microcalcifications were the only sign of cancer and 87.4% in the remaining cases where malignant masses were present). The system made 1.53 true-positive marks and 2.32 false-positive marks per case among cancer cases and 3.32 false-positive marks among cases without cancer.

## 4.3 CADx Medical Systems

CADx Medical Systems was the third company to receive approval for a mammographic CAD system called SecondLook™ [88]. SecondLook was designed to mark areas of a mammogram that are indicative of cancer. It marks masses with circles and microcalcification clusters with rectangles. The sensitivity of the system was reported to be 85% for screening-detected cancers (combination of masses and microcalcification clusters). Additionally, it identified cancer locations in 26.2% of mammograms acquired within 24 months before cancer diagnosis. CADx did not report SecondLook's false-positive rate.

## 4.4 Independent Studies of Commercial Computer-Aided Detection Systems

Several large-scale independent trials of the R2 Image Checker® system have been conducted to test the performance of this system in a clinical setting. In a study conducted by Vyborny et al. [89], it was shown that the R2 Image Checker system detected 86% of the spiculated masses at 0.24 FPI on a dataset of 375 images whereas it had a detection sensitivity of 53% for nonspiculated masses. All of these masses were given a subtlety rating of subtle, medium, or obvious by three radiologists.

It is important to note that of the 375 clearly spiculated masses, 271 were classified as "obvious," 73 had a subtlety rating of "medium," and only 31 had a subtlety rating of "subtle." While the R2 Image Checker system detected 94% of the obvious spiculated masses, it detected only 70% of the medium spiculated masses and 52% of the subtle spiculated masses. Though the overall results (86% at 0.24 FPI) for the detection of spiculated masses are impressive, it is important to note that the study used a large number of obvious masses and a much smaller number of subtle masses. Thus, there is still room for improvement, even in the detection of spiculated masses.

Freer and Ulissey [90] tested the performance of the R2 Image Checker system on more than 12,860 patients in a community breast center. For the first 20,624 radiographs, they observed that 14,214 computer cues or marks were made by the CAD system. Of these, 13,846 marks (97.4%) were dismissed by the radiologist as false positives. This corresponds to a false-positive rate of 0.671 FPI. The CAD system detected 67% (18 of 27) malignant masses and 100% of the clustered calcifications (22 of 22). The authors argue that dismissing the large number of false-positive marks was easy for a radiologist to do [90]. However, another study claims that dismissing false-positive cues can be difficult [91]. This study clearly showed that the R2 Image Checker system is better at the detection of calcifications than at the detection of masses.

Baker et al. [92] studied the performance of two CAD systems for the detection of architectural distortions on a set of 80 images. They observed that the R2 Image Checker system had a sensitivity of 38% at 0.7 FPI while the CADx SecondLook system had a sensitivity of 21% at 1.27 FPI. They concluded that the sensitivity of current systems for the detection of architectural distortions is very low and that considerable improvements are needed for this detection task. On the basis of these studies, radiologists tend to trust the calcification cues more than the mass cues. This is also documented in the literature and prominent radiologists like C. J. D'Orsi have published papers saying that "I would initially use only the calcification prompt and feel extremely comfortable that I have not missed any substantial calcifications when no cues for calcium are present" [128]. Thus,

there is room for improvement in the detection accuracy of architectural distortions.

All of the commercially available CAD systems perform much better at detecting calcifications than at detecting masses or architectural distortions. One also cannot make a direct comparison of these systems as there has been no clinical study that compares the performance of these systems on the same set of cases.

# 5 Recent Advances and Future Directions in Breast Cancer Computer-Aided Detection/Computer-Aided Diagnosis

In this section we highlight some of the most recent advances in breast cancer CAD/CADx and point to some ongoing challenges and areas of research that will require future work.

## 5.1 Computer-Aided Detection: Masses

Improvement is needed in the detection of all categories of masses. Improvements are required both in terms of increasing sensitivity and lowering the number of FPI. The sensitivity for detection of calcifications is much higher than that of masses and the number of FPI is lower for calcifications than for masses. This difference in performance has been exhibited in independent studies and manufacturer reports. For example, R2 Technologies reports that the Image Checker system operates with 98.5% sensitivity at 0.185 FPI for calcifications and 86% sensitivity at 0.24 FPI for spiculated masses [86]. An independent study reported that the Image Checker system performed with 100% sensitivity at 0.29 FPI for calcifications and 67% sensitivity at 0.39 FPI on masses (spiculated and nonspiculated together) [90]. However, R2 Technologies' assessment indicates that the system's performance is substantially lower for spiculated rated as ''subtle'' (sensitivity dropped to 52%) [86]. Thus, while such systems now provide radiologists with a powerful aid for calcification detection, improvements are needed in mass detection.

Masses can be of varying sizes and shapes and thus multiscale, multiorientation methods would be most appropriate for this task. With multiorientation methods, the aim is to extract the directional information present which may help improve the performance of the detection algorithms. To the best of our knowledge, there have been no studies on the statistical properties of masses such as the radius of the central mass region, number of spicules, length, or thickness of spicules. We believe that such a systematic study would help create better detection and diagnosis algorithms.

## 5.2 Computer-Aided Detection: Architectural Distortions

Architectural distortion (AD) refers to the mammographic presentation of a breast lesion in which the normal structure of the breast parenchyma is distorted as if being pulled into a central point, but without a radioopaque central density [4]. Although ADs are less prevalent than masses or calcifications, they are the third most common mammographic sign of cancer and are strongly suggestive of malignancy; approximately 48% to 60% of AD that are biopsied are found to be cancer [5, 92], and about 80% of those cancers are invasive [92]. It is estimated that 12% to 45% of cancers missed in mammographic screening are ADs [9, 93, 94] and thus it is important to detect architectural distortions accurately. However, detection of ADs is an extremely difficult task and current commercial systems perform very poorly on the detection of ADs [92]. A recent, small study suggests that current CAD systems do not detect AD with adequate sensitivity or specificity. Baker et al. [92] found that one commercial system achieved a per-image sensitivity of 30/80 (38%) at 0.70 FPI and another achieved a per-image sensitivity of 17/80 (21%) at 1.27 FPI. Computer aids that improve the detection of AD have the potential to reduce morbidity and save lives through earlier cancer diagnosis.

Current methods likely fail to detect ADs because they are typically designed to detect a radio-opaque circular density. New methods that focus on identifying radiating lines, regardless of the presence of a central mass region have the potential to be more sensitive for the detection of AD. We believe a multiscale and multiorientation approach would be most ideal for the detection of architectural distortions. Another important issue to contend with is that radiating lines may have variable widths, frequencies, and so forth Knowledge of degree of variation would help in the design of more sophisticated AD detection algorithms. However, to date there has been no systematic study of the natural range of appearances of ADs and spiculated masses.

## 5.3 Computer-Aided Diagnosis: All Lesion Types

A significant drawback to mammography is its poor positive-predictive value. Less than a third of mammographically suspicious breast lesions that are biopsied are found to be cancer [11]. Thus, it would be exceedingly valuable to produce CADx systems that could aid in the decision to recommend biopsy or short-term follow-up mammography. Avoiding benign biopsies would spare many women stress, anxiety, discomfort, and expense. Moreover, there is the possibility of increasing the sensitivity of mammography through CADx since it is estimated that about half of missed cancers are missed due to misinterpretation rather than oversight [9]. This is a challenging problem because it is difficult to reduce

the number of false positives while maintaining a high rate of cancer detection. While several studies have investigated CADx systems, no methods have progressed to commercial systems. A brief description of the status of current diagnosis methods and ideas to improve them are discussed in the following sections.

Currently, the performances reported in the literature are better for calcification detection algorithms than for calcification diagnosis algorithms. As mentioned earlier, the most challenging step in the diagnosis is the segmentation of calcifications. There are two possible ways in which the diagnostic performance could be improved. First, since segmentation cannot be perfect, features that are robust to segmentation errors are highly desirable. Second, it is known that radiologists do not look at every individual calcification to make a diagnosis but look at the global properties of the cluster. Thus, more emphasis could be given to designing sophisticated cluster features that capture the spatial arrangement of calcifications.

Mass diagnosis algorithms perform relatively better than the calcification diagnosis algorithms. This is possibly due to the fact that segmentation of masses is an easier task as compared to the segmentation of individual calcifications. However, they still must be improved before they can be clinically adopted and incorporated into commercial systems. There have been no similar studies conducted for the diagnosis of architectural distortions and this is an open area of research.

## 5.4 Computer-Aided Detection/ Computer-Aided Diagnosis: Multiview, Multimodality

Radiologists use information from both CC and MLO views to search for abnormalities. Most detection algorithms analyze only a single view at a time. Recently, some researchers have been working on using information from both the MLO and the CC image for the mass detection step [95]. This methodology faces a number of challenges. The registration of objects in the CC and MLO view and combination features from both views is a difficult task. These algorithms try to extract the correlation of lesions between the two views. However, this is difficult since the breast tissue is elastic and deformable and the tissue is compressed to different degrees when the two views are obtained and it is difficult to account for this compression.

Mammography is the modality used for the screening of breast cancer. However, in diagnostic mammography, complementary modalities like ultrasound or MRI are often used to obtain additional information. An open area of research is the task of how the information from different modalities could be combined by CAD/CADx systems to make a better diagnosis. The major challenge in this task would be the combination of features extracted from multiple modalities. A number of studies have investigated CAD/CADx using non–x-ray modalities, especially ultrasound, which have not been reviewed in this chapter (e.g., [96–112]).

Radiologists analyze current and prior mammograms (if available) simultaneously to detect signs of cancers. A few researchers have been trying to incorporate this principle into CAD and CADx algorithms. Hadjiiski et al. [113] proposed the use of texture and morphologic features for the diagnosis of malignant and benign masses from mammograms from multiple examinations. The main challenge is the registration of structures on the current and prior mammograms. This is a difficult task because the breast tissue is deformable. During mammography the breast tissue is compressed and this causes changes in the relative positions of the breast structures during multiple examinations.

## 5.5 Computer-Aided Detection/ Computer-Aided Diagnosis: Evaluation Methodologies

The standard methodology for reporting the performance of diagnosis algorithms is ROC analysis and its counterpart for detection algorithms is the FROC methodology. The ROC methodology has been studied in detail and a number of statistical measures for analyzing the ROC curves are available. In comparison, the FROC methodology is less developed and until recently, the statistical analysis of the FROC curves was quite weak (see [63, 114]). Traditional analyses treated responses on the same case as if they were made independently. For this reason, these methods have been rightly criticized [115, 116]. Recently, major progress has been made in this field [16], including a new method of analyzing the FROC data, called jackknife free-response receiver operating characteristic (JAFROC), which does not need to assume independence of ratings on the same image. This state-of-the-art FROC methodology, which is now rapidly evolving, will help to make fair comparisons between two systems and will help improve the evaluation of CAD systems

Most diagnostic algorithms perform a two-class classification. An ROI is either classified as malignant or benign and the performance of the algorithm is reported using an ROC curve. Since detection algorithms are not perfect, in a fully automated system, the ROIs to be classified would actually contain benign, malignant, or false-positive signals. Thus, a few groups [117] have started working on a three-class classification task for the diagnosis task. In this methodology, ROIs are classified as malignant, benign or false-positive.

Since CAD and CADx systems are expected to act as aids to radiologists, it is critical to evaluate whether radiologists change their management recommendations when the systems are used. The multiple-reader, multiple-case (MRMC) ROC method has been advocated as the "best practice" for evaluating competing imaging modalities, including CAD/CADx systems [118]. Pilot studies are needed to

obtain estimates of the components of variability that arise from patient cases and from readers as well as interaction terms [119–121]. While many aspects of good experimental design and analysis have been elucidated, there are still important questions to be addressed concerning how radiologists interact with decision aids and how the design of CAD/CADx systems might be modified to improve that interaction. For example, the effects of prevalence [122, 123] and cueing conditions are actively being studied [124].

## 5.6 Computer-Aided Detection/Computer-Aided Diagnosis: What Role in the Clinic?

Current commercial systems have been approved for use as "second readers" and much of the research has focused on this paradigm as well. Currently, we face a crisis in mammography in which women's access to breast cancer screening is being endangered by a shortage of breast imaging specialists. A recent report from the Institute of Medicine discusses this challenge and suggests that it would helpful to have nonspecialists, even technologists to prescreen mammograms [125]. We argue that a CAD system could potentially fill the same role. This would, of course, require improvements over the performance of current systems.

Another shift in the role of CAD/CADx systems that should be considered is tailoring the human–computer interaction. Systems have typically been investigated as one-size-fits-all tools that provide the same information in the same manner to all users. It is possible that future systems that interact in a flexible manner with radiologists may be more useful and useable. For example, perhaps a CAD system could be personalized on the basis of automatic classification of lesions into BIRADS$^{TM}$ categories (i.e., Dr. X wants to be prompted on lobulated or spiculated masses only) or a perhaps a CADx system could adjust its output based on the user's expectations about the disease prevalence [122].

## Acknowledgments

The authors would like to thank Kelly N. Forsythe and Nick Markey for bibliographic data entry. We also thank Prof. Nico Karssemeijer and W. Hagen (IEEE Publications) for assistance in reproducing certain figures for this chapter.

## References

[1] American Cancer Society, *Cancer Facts and Figures 2004*, (Atlanta, 2004).

[2] E. J. Feuer, L. Wun, C. C. Boring, et al., "The lifetime risk of developing breast cancer," *J. Nat. Cancer Inst.* 85, 892–897 (1993).

[3] L. Wun, R. M. Merrill, and E. J. Feuer, "Estimating lifetime and age-conditional probabilities of developing cancer," *Lifetime Data Analysis* 4, 169–186 (1998).

[4] American College of Radiology, *ACR BI-RADS—Mammography, Ultrasound & Magnetic Resonance Imaging*, 4th ed. American College of Radiology (Reston, VA, 2003).

[5] S. G. Orel, N. Kay, C. Reynolds, et al., "BI-RADS categorization as a predictor of malignancy," *Radiology* 211, 845–850 (1999).

[6] C. H. Lee, "Screening mammography: proven benefit, continued controversy," *Radiol. Clinics North Am.* 40, 395–407 (2002).

[7] K. Kerlikowske, P. A. Carney, B. Geller, et al., "Performance of screening mammography among women with and without a first-degree relative with breast cancer," *Annals Int. Med.* 133, 855–863 (2000).

[8] T. M. Kolb, J. Lichy, and J. H. Newhouse, "Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. [see comment]," *Radiology* 225, 165–175 (2002).

[9] R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* 184, 613–617 (1992).

[10] M. L. Giger, "Computer-aided diagnosis in radiology," *Acad. Radiol.* 9, 1–3 (2002).

[11] D. B. Kopans, "The positive predictive value of mammography," *AJR. Am. J. Roentgenol.* 158, 521–526 (1992).

[12] M. L. Giger, N. Karssemeijer, and S. G. Armato, III, "Computer-aided diagnosis in medical imaging," *IEEE Trans. Med. Imag.* 20, 1205–1208 (2001).

[13] M. L. Giger, "Computer-aided diagnosis of breast lesions in medical images, "*Comput. Science Engin.* 2, 39–45 (2000).

[14] K. Doi, H. MacMahon, S. Katsuragawa, et al., "Computer-aided diagnosis in radiology: potential and pitfalls," *Eur. J. Radiol.* 31, 97–109 (1999).

[15] C. J. Vyborny, M. L. Giger, and R. M. Nishikawa, "Computer-aided detection and diagnosis of breast cancer," *Radiol. Clin. North Am.* 38, 725–740 (2000).

[16] D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis and validation," *Medical Phys.* 31, 1–18 (2004).

[17] D. B. Kopans, *Breast Imaging* (Lippincott Williams, New York 1998).

[18] H. D. Li, M. Kallergi, L. P. Clarke, et al., "Markov random field for tumor detection in digital mammography," *Medical Imag., IEEE Trans.* 14, 565–576 (1995).

[19] W. P. Kegelmeyer, Jr., J. M. Pruneda, P. D. Bourland, et al., "Computer-aided mammographic screening for spiculated lesions," *Radiology* 191, 331–337 (1994).

[20] N. Karssemeijer and G. M. te Brake, "Detection of stellate distortions in mammograms," *IEEE Trans. Med. Imag.* 15, (1996).

[21] S. L. Liu, C. F. Babbs, and E. J. Delp, "Multiresolution detection of spiculated lesions in digital mammograms," *IEEE Trans. Image Process.* 10, 874–884 (2001).

[22] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, "Markov random field for tumor detection in digital mammography," *IEEE Trans. Med. Imag.* 14, 565–576 (1995).

[23] T. Matsubara, H. Fujita, T. Endo, et al.,"Development of mass detection algorithm based on adaptive thresholding technique

in digital mammograms,'' K. Doi, M. L. Giger et al. eds. 391–396 (Elsevier, Amsterdam, The Netherlands, 1996).
[24] H. Li, Y. Wang, K. J. Liu, et al., ''Computerized radiographic mass detection—part I: Lesion site selection by morphological enhancement and contextual segmentation,'' *IEEE Trans. Med. Imag.* 20, 289–301 (2001).
[25] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley-Interscience, New York, 2000).
[26] H. Kobatake, M. Murakami, H. Takeo, et al., ''Computerized detection of malignant Tumors on digital mammograms,'' *IEEE Trans. Med. Imag.* 18, 369–378 (1999).
[27] N. Petrick, H. P. Chan, B. Sahiner, et al., ''An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection,'' *IEEE Trans. Medical Imag.* 15, 59–67 (1996).
[28] N. Petrick, H. P. Chan, B. Sahiner, et al., ''Combined adaptive enhancement and regiongrowing segmentation of breast masses on digitized mammograms,'' *Medical Physics* 26, 1642–1654 (1999).
[29] W. E. Polakowski, D. A. Cournoyer, S. K. Rogers, et al., ''Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency,'' *IEEE Trans. Med. Imag.* 16, 811–819 (1997).
[30] D. Brzakovic, X. M. Luo, and P. Brzakovic, ''An approach to automated detection of tumors in mammograms,'' *IEEE Trans. Med. Imag.* 9, 233–241 (1990).
[31] W. Qian, L. Li, L. Clarke, et al., ''Comparison of adaptive and non adaptive cad methods for mass detection,'' *Academic Radiol.* 6, 471–480 (1999).
[32] S. M. Lai, X. Li, and W. F. Bischof, ''On techniques for detecting circumscribed masses in mammograms,'' *IEEE Trans. Med. Imag.* 8, 377–386 (1989).
[33] B. R. Groshong and W. P. Kegelmeyer, ''Evaluation of a Hough transform method for circumscribed lesion detection,'' K. Doi, M. L. Giger et al. eds. 361–366 (Elsevier, Amsterdam, The Netherlands, 1996).
[34] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. (Prentice Hall, Upper Saddle River, NJ, 2001).
[35] W. Zhang, K. Doi, M. L. Giger, et al., ''An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms,'' *Med. Physics* 23, 595–601 (1996).
[36] B. Sahiner, H.-P. Chan, N. Petrick, et al., ''Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images,'' *IEEE Trans. Med. Imag.* 15, 598–610 (1996).
[37] D. Wei, H. P. Chan, M. A. Helvie, et al., ''Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis,'' *Med. Physics* 22, 1501–1513 (1995).
[38] D. Wei, H. P. Chan, N. Petrick, et al., ''False positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis,'' *Med. Physics* 24, 903–914 (1997).
[39] G. M. te Brake, N. Karssemeijer, and J. H. Hendriks, ''An automatic method to discriminate malignant masses from normal tissue in digital mammograms,'' *Physics Med. Biol.* 45, 2843–2857 (2000).
[40] M. A. Kupinski and M. L. Giger, ''Investigation of regularized neural networks for the computerized detection of mass lesions in digital mammograms,'' *Proceedings of the 19th Annual International Conference of the IEEE* (1997).
[41] 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. Physics* 30, 2123–2130 (2003).
[42] A. H. Baydush, D. M. Catarious, C. K. Abbey, et al., ''Computer aided detection of masses in mammography using subregion Hotelling observers,'' *Med. Physics*, 30, 1781–1787 (2003).
[43] F. F. Yin, M. L. Giger, K. Doi, et al., ''Computerized detection of masses in digital mammograms: analysis of bilateral subtraction images,'' *Med. Physics* 18, 955–963 (1991).
[44] W. K. Zouras, M. L. Giger, P. Lu, et al., ''Investigation of a Temporal Subtraction Scheme for Computerized Detection of Breast Masses in Mammograms,'' presented at Digital Mammography, June 1996.
[45] R. N. Strickland and H. I. Hahn, ''Wavelet transforms for detecting microcalcifications in mammograms,'' *IEEE Trans. Med. Imag.* 15, 218–229 (1996).
[46] H. Yoshida, K. Doi, R. M. Nishikawa, et al., ''An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms,'' *Acad. Radiol.* 3, 621–627 (1996).
[47] W. Zhang, H. Yoshida, R. M. Nishikawa, et al., ''Optimally weighted wavelet transform based on supervised training for detection of microcalcifications in digital mammograms,'' *Med. Physics* 25, 949–956 (1998).
[48] W. Qian, M. Kallergi, L. P. Clarke, et al., ''Tree structured wavelet transform segmentation of microcalcifications in digital mammography,'' *Med. Physics* 22, 1247–1254 (1995).
[49] M. N. Gurcan, Y. Yardimci, A. E. Cetin, et al., ''Detection of microcalcifications in mammograms using higher order statistics,'' *IEEE Signal Process. Lett.* 4, 213–216 (1997).
[50] E. J. Candès and D. L. Donoho, ''Ridgelets: a key to higher-dimensional intermittency?'' *Phil. Trans. R. Soc. Lond. A.* 2495–2509 (1999).
[51] T. Netsch and H.-O. Peitgen, ''Scale-space signatures for the detection of clustered microcalcifications in digital mammograms,'' *IEEE Trans. Med. Imag.* 18, 774–786 (1999).
[52] H. P. Chan, K. Doi, S. Galhotra, et al., ''Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography,'' *Med. Physics* 14, 538–548 (1987).
[53] H. P. Chan, S. C. Lo, B. Sahiner, et al., ''Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network,'' *Med. Physics* 22, 1555–1567 (1995).
[54] D. H. Davies and D. R. Dance, ''Automatic computer detection of clustered calcifications in digital mammograms,'' *Physics Med. Biol.* 35, 1111–1118 (1990).
[55] R. M. Nishikawa, Y. Jiang, M. L. Giger, R. A. Schmidt et al. ''Performance of automated CAD schemes for the detection and classification of clustered microcalcifications,'' Digital Mammography, A. G. Gate et al. eds. 13–20 (Elsevier, Amsterdam, 1994).

[56] B. Zheng, Y. H. Chang, M. Staiger, et al., "Computer-aided detection of clustered microcalcifications in digitized mammograms," *Acad. Radiol.* 2, 655–662 (1995).

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

[58] R. H. Nagel, R. M. Nishikawa, J. Papaioannou, et al., "Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms," *Med. Physics* 25, 1502–1506 (1998).

[59] H. Li, K. J. Liu, and S. C. Lo, "Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms," *IEEE Trans. Med. Imag.* 16, 785–798 (1997).

[60] L. E. Dodd, R. F. Wagner, S. G. Armato, 3rd, et al., and Image Database Consortium Research, "Assessment methodologies and statistical issues for computer-aided diagnosis of lung nodules in computed tomography: contemporary research topics relevant to the lung image database consortium," *Acad. Radiol.* 11, 462–475 (2004).

[61] C. E. Metz, "Basic principles of ROC analysis," *Sem. Nucl. Med.* 8, 283–298 (1978).

[62] C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.* 21, 720–733 (1986).

[63] D. P. Chakraborty and L. H. Winter, "Free-response methodology: alternate analysis and a new observer-performance experiment," *Radiology* 174, 873–881 (1990).

[64] B. Sahiner, N. Petrick, H. P. Chan, et al., "Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization," *IEEE Trans. Med. Imag.* 20, 1275–1284 (2001).

[65] S. Timp and N. Karssemeijer, "A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography," *Med. Physics* 31, 958–971 (2004).

[66] D. Guliato, R. M. Rangayyan, J. A. Zuffo, et al., "Detection of breast tumor boundaries using iso-intensity contours and dynamic thresholding," *Proc. 4th Int. Workshop Digital Mammography* 253–260 (1998).

[67] S. K. Kinoshita, P. M. Azevedo Marques, A. F. F. Slaets, et al., "Detection and characterization of mammographic masses by artificial neural network," *Proc. 4th Int. Workshop Digital Mammography* (1998).

[68] M. A. Kupinski and M. L. Giger, "Automated Seeded Lesion Segmentation on Digital Mammograms," *IEEE Trans. Med. Imag.* 17, 510–517 (1998).

[69] L. Liberman, A. F. Abramson, F. B. Squires, et al., "The breast imaging reporting and data system: positive predictive value of mammographic features and final assessment categories," *AJR. Am. J. Roentgenol.* 171, 35–40 (1998).

[70] M. Tuceryan and A. K. Jain, "Texture analysis," in *The Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, eds. (World Scientific Publishing Co., 1998), 207–248.

[71] T. M. Mitchell, *Machine Learning*. (WCB/McGraw-Hill, Boston, 1997).

[72] B. Sahiner, H. P. Chan, N. Petrick, et al., "Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis," *Med. Physics* 25, 516–526 (1998).

[73] E. Claridge, Richter J. H., "Characterisation of mammographic lesions," presented at Proc. 2nd Int. Workshop Digital Mammography (1994).

[74] L. Hadjiiski, B. Sahiner, H. P. Chan, et al., "Classification of malignant and benign masses based on hybrid ART2LDA approach," *IEEE Trans. Med. Imag.* 18, 1178–1187 (1999).

[75] S. Pohlman, K. A. Powell, N. A. Obuchowski, et al., "Quantitative classification of breast tumors in digitized mammograms," *Med. Physics* 23, 1337–1345 (1996).

[76] L. M. Bruce and R. R. Adhami, "Classifying mammographic mass shapes using the wavelet transform modulus-maxima method," *IEEE Trans. Med. Imag.* 18, 1170–1177 (1999).

[77] J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computerized image analysis," *IEEE Trans. Med. Imag.* 12, 664–669 (1993).

[78] R. M. Rangayyan, N. M. El-Faramawy, J. E. L. Desautels, et al., "Measures of acutance and shape for classification of breast tumors," *IEEE Trans. Med. Imag.* 16, 799–810 (1997).

[79] M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters," *Med. Physics* 31, 314–326 (2004).

[80] H. P. Chan, B. Sahiner, K. L. Lam, et al., "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," *Med. Physics* 25, 2007–2019 (1998).

[81] M. De Santo, M. Molinara, F. Tortorella, et al., "Automatic classification of clustered microcalcifications by a multiple expert system," *Patt. Recog.* 36, 1467–1477 (2003).

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

[83] W. J. H. Veldkamp, N. Karssemeijer, J. D. M. Otten, et al., "Automated classification of clustered microcalcifications into malignant and benign types," *Med. Physics* 27, 2600–2608 (2000).

[84] L. Shen, R. M. Rangayyan, and J. E. L. Desautels, "Application of shape analysis to mammographic calcifications," *IEEE Trans. Med. Imag.* 13, 263–274 (1994).

[85] A. P. Dhawan, Y. Chitre, and C. Kaiser-Bonasso, "Analysis of mammographic microcalcifications using gray-level image structure features," *IEEE Trans. Med. Imag.* 15, 246–259 (1996).

[86] U. S. Food and Drug Administration, "Summary of Safety and Effectiveness Data: R2 Technologies," *P970058* (1998).

[87] U. S. Food and Drug Administration, "Summary and Safety of Effectiveness Data: ISSI," *P010038* (2002).

[88] U. S. Food and Drug Administration, "Summary of Safety and Effectiveness Data: CADx Medical Systems," *P010034* (2002).

[89] C. J. Vyborny, T. Doi, K. F. O'Shaughnessy, et al., "Breast cancer: importance of spiculation in computer-aided detection," *Radiology* 215, 703–707 (2000).

[90] 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, 781–786 (2001).

[91] B. Zheng, M. A. Ganott, C. A. Britton, et al., "Soft-Copy Mammographic Readings with Different Computer-assisted Detection Cuing Environments: Preliminary Findings," *Radiology* 221, 633–640 (2001).

[92] J. Baker, E. L. Rosen, J. Lo, et al., "Computer-aided diagnostics (CAD) in screening mammography: sensitivity of comercial CAD systems for detecting architectural distortion," *Am. J. Roentgenol.* 181 (2003).

[93] H. Burrell, A. Evans, A. Wilson, et al., "False-negative breast screening assessment. What lessons can we learn?," *Clin. Radiol.* 56, 385–388 (2001).

[94] H. Burrell, D. Sibbering, and A. Wilson, "Screening interval breast cancers: mammographic features of prognostic factors," *Radiology* 199, 811–817 (1996).

[95] S. Paquerault, N. Petrick, H. P. Chan, et al., "Improvement of computerized mass detection on mammograms: fusion of two–view information," *Med. Physics* 29, 238–247 (2002).

[96] K. G. Gilhuijs, M. L. Giger, and U. Bick, "Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging," *Med. Physics* 25, 1647–1654 (1998).

[97] B. K. Szabo, M. K. Wiberg, B. Bone, et al., "Application of artificial neural networks to the analysis of dynamic MR imaging features of the breast," *Eur. Radiol.* 14, 1217–1225 (2004).

[98] T. W. Vomweg, M. Buscema, H. U. Kauczor, et al., "Improved artificial neural networks in prediction of malignancy of lesions in contrast-enhanced MR-mammography," *Med. Physics* 30, 2350–2359 (2003).

[99] P. Gibbs and L. W. Turnbull, "Textural analysis of contrast-enhanced MR images of the breast," *Mag. Reson. Med.* 50, 92–98 (2003).

[100] G. Torheim, F. Godtliebsen, D. Axelson, et al., "Feature extraction and classification of dynamic contrast-enhanced $T2^*$-weighted breast image data," *IEEE Trans. Med. Imag.* 20, 1293–1301 (2001).

[101] Y. L. Huang and D. R. Chen, "Watershed segmentation for breast tumor in 2-D sonography," *Ultrasound Med. Biol.* 30, 625–632 (2004).

[102] S. F. Huang, R. F. Chang, D. R. Chen, et al., "Characterization of spiculation on ultrasound lesions," *IEEE Trans. Med. Imag.* 23, 111–121 (2004).

[103] B. Sahiner, H. P. Chan, M. A. Roubidoux, et al., "Computerized characterization of breast masses on three-dimensional ultrasound volumes," *Med. Physics* 31, 744–754 (2004).

[104] K. Drukker, M. L. Giger, and E. B. Mendelson, "Computerized analysis of shadowing on breast ultrasound for improved lesion detection," *Med. Physics* 30, 1833–1842 (2003).

[105] P. M. Shankar, V. A. Dumane, C. W. Piccoli, et al., "Computeraided classification of breast masses in ultrasonic B-scans using a multiparameter approach," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 50, 1002–1009 (2003).

[106] P. M. Shankar, V. A. Dumane, T. George, et al., "Classification of breast masses in ultrasonic B scans using Nakagami and K distributions," *Physics Med. Biol.* 48, 2229–2240 (2003).

[107] W. M. Chen, R. F. Chang, W. K. Moon, et al., "Breast cancer diagnosis using threedimensional ultrasound and pixel relation analysis," *Ultrasound Med. Biol.* 29, 1027–1035 (2003).

[108] R. F. Chang, W. J. Wu, W. K. Moon, et al., "Support vector machines for diagnosis of breast tumors on US images," *Acad. Radiol.* 10, 189–197 (2003).

[109] R. F. Chang, W. J. Wu, W. K. Moon, et al., "Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis," *Ultrasound Med. Biol.* 29, 679–686 (2003).

[110] C. M. Chen, Y. H. Chou, K. C. Han, et al., "Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks," *Radiology* 226, 504–514 (2003).

[111] V. A. Dumane, P. M. Shankar, C. W. Piccoli, et al., "Computer aided classification of masses in ultrasonic mammography," *Med. Physics* 29, 1968–1973 (2002).

[112] K. Horsch, M. L. Giger, L. A. Venta, et al., "Automatic segmentation of breast lesions on ultrasound," *Med. Physics* 28, 1652–1659 (2001).

[113] L. Hadjiiski, B. Sahiner, H. P. Chan, et al., "Analysis of temporal changes of mammographic features: computer-aided classification of malignant and benign breast masses," *Med. Physics* 28, 2309–2317 (2001).

[114] D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Med. Physics* 16, 561–568 (1989).

[115] C. E. Metz, "Evaluation of digital mammography by ROC analysis," in Digital Mammograph, K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt, eds. 61–68 (Elsevier Science, New York, 1996).

[116] R. G. Swensson, "Unified measurement of observer performance in detecting and localizing target objects on images," *Med. Physics* 23, 1709–1725 (1996).

[117] D. C. Edwards, L. Li, C. E. Metz, et al., "Estimating three-class ideal observer decision variables for computerized detection and classification of mammographic mass lesions," *Med. Physics* 31, 81–90 (2004).

[118] R. F. Wagner, S. V. Beiden, G. Campbell, et al., "Assessment of medical imaging and computer-assist systems: lessons from recent experience," *Acad. Radiol.* 9, 1264–1277 (2002).

[119] S. V. Beiden, R. F. Wagner, and G. Campbell, "Components-of-variance models and multiple-bootstrap experiments: an alternative method for random-effects, receiver operating characteristic analysis," *Acad. Radiol.* 7, 341–349 (2000).

[120] S. V. Beiden, R. F. Wagner, G. Campbell, et al., "Components-of-variance models for random-effects ROC analysis: the case of unequal variance structures across modalities," *Acad. Radiol.* 8, 605–615 (2001).

[121] S. V. Beiden, R. F. Wagner, G. Campbell, et al., "Analysis of uncertainties in estimates of components of variance in multivariate ROC analysis," *Acad. Radiol.* 8, 616–622 (2001).

[122] K. Horsch, M. L. Giger, C. E. Metz, et al., "Prevalence modified estimation of computer-determined probabilities of malignancy for CAD," presented at Radiological Society of North America Annual Meeting (Chicago, IL, 2003).

[123] D. Gur, H. E. Rockette, D. R. Armfield, et al., "Prevalence effect in a laboratory environment. [see comment]," *Radiology* 228, 10–14 (2003).

[124] B. Zheng, R. G. Swensson, S. Golla, et al., "Detection and classification performance levels of mammographic masses

under different computer-aided detection cueing environments,'' *Acad. Radiol.* 11, 398–406 (2004).

[125] E. E. Penhoet, J. E. Joy, and D. B. Petitti, *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis* (The National Academies Press, New York, 2004).

[126] M. Heath, K. W. Bowyer, D. Kopans, et al., ''The Digital Database for Screening Mammography,'' presented at 5th International Workshop on Digital Mammography (Toronto, Canada, 2000).

[127] L. Li, W. Qian, and L. P. Clarke, ''X-ray medical image processing using directional wavelet transform,'' presented at Acoustics, Speech, and Signal Processing (1996).

[128] C. J. D' Orsi, ''Computer-aided Detection: There is No Free Lunch,'' Radiology, 221: 585–586, 2001.