CADx of mammographic masses and clustered microcalcifications: A review

Matthias Elter[a)]
Fraunhofer Institute for Integrated Circuits (IIS), Am Wolfsmantel 33, 91058 Erlangen, Germany

Alexander Horsch[b)]
Institute for Medical Statistics and Epidemiology, TU München, Ismaninger Strasse 22, 81675 München, Germany and Department of Computer Science, University of Tromsø Breivika, N-9037 Tromsø, Norway

(Received 3 November 2008; revised 30 March 2009; accepted for publication 31 March 2009; published 5 May 2009)

Breast cancer is the most common type of cancer among women in the western world. While mammography is regarded as the most effective tool for the detection and diagnosis of breast cancer, the interpretation of mammograms is a difficult and error-prone task. Hence, computer aids have been developed that assist the radiologist in the interpretation of mammograms. Computer-aided detection (CADe) systems address the problem that radiologists often miss signs of cancers that are retrospectively visible in mammograms. Furthermore, computer-aided diagnosis (CADx) systems have been proposed that assist the radiologist in the classification of mammographic lesions as benign or malignant. While a broad variety of approaches to both CADe and CADx systems have been published in the past two decades, an extensive survey of the state of the art is only available for CADe approaches. Therefore, a comprehensive review of the state of the art of CADx approaches is presented in this work. Besides providing a summary, the goals for this article are to identify relations, contradictions, and gaps in literature, and to suggest directions for future research. Because of the vast amount of publications on the topic, this survey is restricted to the two most important types of mammographic lesions: masses and clustered microcalcifications. Furthermore, it focuses on articles published in international journals. © *2009 American Association of Physicists in Medicine*. [DOI: 10.1118/1.3121511]

Key words: mammography, computer-aided diagnosis, CADx, review

I. INTRODUCTION

According to the American Cancer Society,[1] breast cancer is the most common type of cancer among women in the United States and accounts for more than 25% of all cancers diagnosed in US women. Mammography is regarded as the most effective tool for breast cancer detection and diagnosis available today. However, mammogram interpretation is a repetitive and thus an error-prone task. This leads to 10%–30% of all cancers to be missed by radiologists. One possible solution to this problem are computer-aided detection (CADe) systems that automatically detect suspicious lesions in mammograms, which otherwise might have been missed by the radiologist and serve as a reminder by pointing out their location. For a survey of the state of the art of CADe systems, the reader is referred to a recent review article by Nishikawa.[2]

Based on the inspection of mammograms and often supplemental ultrasound and magnetic resonance images, radiologists give a recommendation for the subsequent patient management. Based on the level of suspicion of malignancy of the lesions found in the mammograms, usually a recommendation is made for a short- or long-term follow-up examination or (in the case of higher suspicion of malignancy) for a breast biopsy (invasive removal and pathological testing of a suspicious area of the breast). However, the characterization of lesions as benign or malignant based on their appearance in mammograms is a difficult task even for expert radiologists. Because a mammogram is a two-dimensional image of a three-dimensional breast, superposition of breast tissue often produces patterns that appear like suspicious masses to a radiologist or alters the appearance of real mammographic lesions. Furthermore, lesions with typically malignant characteristics sometimes may represent benign lesions and vice versa. It is reported that usually less than 30% of all breast biopsies actually show a malignant pathology.[3–7] The high number of unnecessary breast biopsies causes major mental and physical discomfort for the patients as well as unnecessary expenses spent for examinations. Therefore, computer-aided diagnosis (CADx) systems have been proposed in the past years with the aim to support radiologists in the discrimination of benign and malignant mammographic lesions and to increase the positive predictive value (PPV) of mammogram interpretation. The PPV measures the percentage of breast biopsies that are tested positive for cancer. Furthermore, CADx systems can potentially improve the sensitivity of mammography because approximately half of all missed cancers seem to be missed due to misclassification rather than due to oversight.[8] While a large amount of CADx approaches can be found in literature, no comprehensive review of the state of the art seems to exist. Therefore, this work provides a review of the state of the art of the computer-aided discrimination of benign and

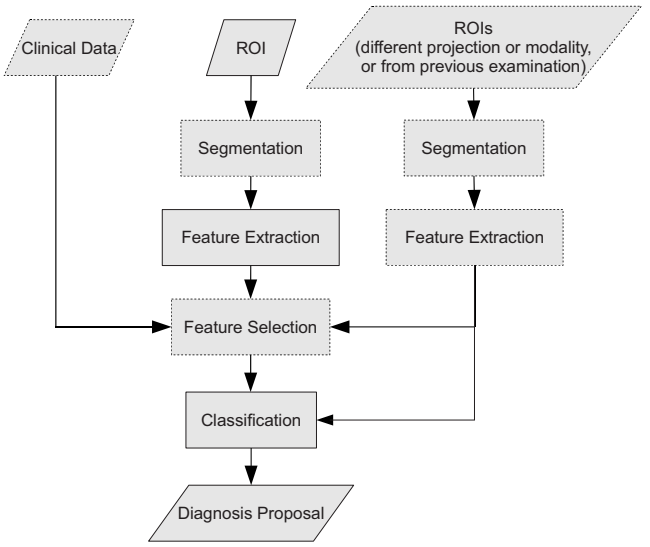


FIG. 1. Illustration of the design of a typical CADx system. Borders of optional inputs and optional processing steps are dotted.

malignant mammographic lesions. More precisely it concentrates on the most important types of mammographic lesions, masses, and clustered microcalcifications, and does not cover less common types like architectural distortions. Care has been taken to make this review as complete as possible. However, because of the vast amount of publications on the topic, the focus is on articles published in international journals.

The task of discriminating benign and malignant lesions is usually modeled as a two-class classification problem. Most approaches start with a region of interest (ROI) depicting the lesion that shall be classified. The ROI may have been delineated manually by a radiologist or automatically by a computer-aided detection system. It usually is a rectangular subimage cut from a mammogram. Most CADx systems have a four-stage process: lesion segmentation, feature extraction, feature selection, and finally classification. Figure 1 shows a flowchart of a typical CADx system.

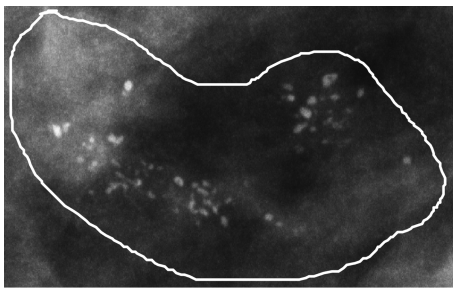
Often one or more stages of the process illustrated in Fig. 1 are omitted by CADx systems. For example, in many systems, no automatic selection of features is done, while other approaches do not require an explicit segmentation of the lesion from the background tissue in the ROI. Furthermore, some approaches use multiple ROIs containing the lesion cut from different mammographic projections [e.g., craniocraudal (CC) and mediolateral oblique (MLO)], from additional modalities (like ultrasound), or from previous examinations (temporal change analysis). In these approaches, lesion segmentation and feature extraction are usually done independently in all ROIs containing the lesion. Finally, many CADx approaches use clinical data like the patient's age as additional features. These features are not extracted from the ROI containing the lesion but instead from, usually textual, annotations containing information about the whole case.

The organization of the rest of this work closely follows the design of typical CADx systems, as illustrated in Fig. 1. Starting with an overview of approaches for the segmentation of clustered microcalcifications and mammographic masses in Secs. II and III, respectively, we proceed with a review of the various methods that are employed for the automatic extraction of features representing attributes of clustered microcalcifications and mammographic masses in Secs. IV and V, respectively. While the methods employed for lesion segmentation and feature extraction differ for clustered microcalcifications and mammographic masses, approaches for the remaining steps of feature selection and classification usually can be applied to both lesion types. Therefore, no differentiation by lesion type is made in the reviews of methods for these steps. A survey of state of the art approaches is given in Secs. VI and VII, respectively.

In Secs. VIII and IX, special topics, CADx using temporal change analysis and multiview/multimodal CADx, are discussed. Furthermore, besides CADx approaches that employ features that are automatically extracted by a computer, alternative CADx approaches that instead use features that are manually extracted by the radiologist have been proposed in the past years. Hence, this review also covers the discrimination of mammographic lesions using human-extracted features. CADx systems of this type usually have a similar design as systems using computer-extracted features as described above. However, because of using human-extracted features, the segmentation step is not necessary. Furthermore the extraction of features is done manually by the radiologist instead of automatically by the CADx system. Because of these differences, an outline of the state of the art of CADx approaches using human-extracted features is provided separately in Sec. X. This work closes with an overview of CADx system evaluation in Sec. XI and a summary and conclusion in Sec. XII.

II. SEGMENTATION OF MICROCALCIFICATIONS

The morphology and the distribution of individual microcalcifications in a cluster are the most important attributes considered by radiologists for the discrimination of benign and malignant types of this kind of lesion. The segmentation of individual microcalcifications in a ROI depicting a microcalcification cluster is a crucial prerequisite for the automatic extraction of features representing these attributes. A large variety of approaches to the segmentation of microcalcifications can be found in literature and thus only a brief overview of the most important approaches is given here. While the focus of this work is on the diagnosis and not on the detection of mammographic lesions, several calcification segmentation approaches that were originally proposed for the detection of calcification clusters are discussed. The reasoning is that most computer-aided detection systems for clustered microcalcifications have two steps: (i) Segmentation of individual calcification particles in a mammogram and (ii) detection of clusters by an analysis of the spacial distribution of the particles. The first step of these CADe approaches usually is feasible for the segmentation of calcification particles in a ROI instead of a full mammogram too. CADx systems usually work on ROIs containing a single


(a)

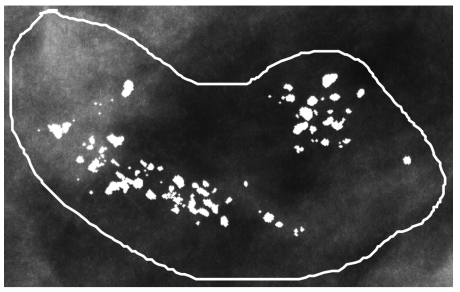

(b)

FIG. 2. An example of a ROI depicting a microcalcification cluster (a) and the result of an automatic segmentation of the individual calcification particles from the background tissue and each other (b).

cluster of calcifications. The segmentation task consists of automatically finding all individual calcifications and of the pixel-exact delineation of their shape. Figure 2(b) illustrates the results of a segmentation of the individual particles in the microcalcification cluster ROI displayed in Fig. 2(a).

II.A. Semiautomatic methods

Early approaches to the problem often have been semiautomatic and included a mandatory manual step. For example, Shen *et al.*[9] proposed a technique based on region growing that requires the radiologist to manually select a seed pixel for each microcalcification particle. A similar approach was proposed by Jiang *et al.*[10] who fitted a third-degree polynomial surface to reduce background tissue and applied a gray-level based region-growing algorithm afterward. A more recent semiautomatic approach, proposed by Paquerault *et al.* in 2004,[11] also requires manual selection of calcification seed points and is based on a transformation into polar coordinates, followed by an analysis of the radial gradient map.

II.B. Low-level features

Other early approaches determine whether a pixel belongs to a microcalcification or the background based on simple criteria like high absolute gray levels or high local contrast.[12,13] However, these approaches tend to fail when microcalcifications are embedded in dense background tissue. Schmidt *et al.*[14] tried to avoid this problem by combining the enhancement in high-local-contrast areas with a reduction in background tissue using a fitted polynomial.

II.C. Wavelet transform

Because of their small size and their high degree of localization, microcalcifications represent high-spatial frequencies in the image. The wavelet transform is an attractive option for the detection of high-spatial-frequency components of an image because it can spatially localize high-frequency components. Hence, it was used by many authors for the segmentation of microcalcifications.[15–27] The general idea of these approaches is to decompose a ROI into its subbands using the wavelet transformation and to weight the coefficients of the subbands so that microcalcifications are enhanced, and background tissue, as well as noise, is suppressed once the inverse wavelet transform is applied to the data.

II.D. Laplacian of Gaussian

Microcalcification particles appear as bright spots in mammograms. A common approach for the detection of bright spots in images, also referred to as "blob detection" in the computer-vision literature, is based on the Laplacian of Gaussian (LoG) filter. Bright spots correspond to local maxima in an image if it is convolved with a LoG filter kernel of an appropriate size. Netsch and Peitgen[28] exploited this observation and proposed to identify the position and size of individual calcification particles using a LoG scale-space representation of a mammogram. The LoG operator can be computed in an alternative, more efficient way as the limit case of the difference between two Gaussian smoothed images. This approach is usually referred to as the difference of Gaussians (DoG) approach and is used by Salfity *et al.* for microcalcification segmentation.[29]

II.E. Morphological operators

Another class of approaches for microcalcification segmentation is based on mathematical morphology, more precisely on gray-level morphological operators. Nishikawa *et al.*[30] combined morphological erosion operators with a difference image technique to segment calcification particles. Betal *et al.*[31] as well as Fu *et al.*[32] used the so-called top-hat operator (which is defined as a subtraction of a morphologically opened image from the original image), followed by edge detection and flood filling for calcification segmentation. Finally, as early as 1993 Dengler *et al.*[33] proposed a combination of a DoG operator and a top-hat operator for the localization and exact delineation of microcalcifications, respectively.

II.F. Miscellaneous methods

There are a variety of methods using miscellaneous techniques that do not fit into the categories discussed so far. Examples are the approaches by Ibrahim *et al.*[34] who used a triple ring filter, Linguraru *et al.*[35] who used a biological model of contrast detection and Nakayama *et al.*[36,37] who employed filter banks. A few approaches to calcification segmentation[38–40] are based on modeling the background tissue and the ductal patterns of a mammogram using a fractal model. Because the fractal model fails to adjust to the high-

TABLE I. Summary of strengths and weakness of the classes of microcalcification segmentation methods mentioned in this section.

Method class	Interaction	Strengths	Weaknesses
Semiautomatic	Some to intense	Accurate number of particles	Can be time consuming, in some cases almost impossible, high interobserver variability
Threshold, contrast	None to some	Very simple and fast	Often fail in case of dense background tissue, high inter observer variability
Wavelets	None	Exact shape of particles	Moderately accurate at acceptably low FP rates
LoG, DoG	None	Simple model, exact locations	Model fits Gauss-shaped particles only
Morphology	None	Exact shape of particles	···

frequency image components that represent calcification particles, a calcification image can be obtained by subtracting the model from the original image. Some researches employ fuzzy-logic techniques for the segmentation of microcalcifications. Cheng *et al.*[41] proposed to employ fuzzy set theory and geometrical statistics to increase the contrast of, and finally segment, microcalcifications. Verma and Zakos[42] proposed a semiautomatic calcification segmentation approach based on simple fuzzy-logic rules.

II.G. Preprocessing

The noise levels in mammograms strongly depend on the brightness in different image regions. Bright regions that represent dense (e.g., glandular) tissue usually have higher noise levels than regions representing less dense (e.g., fatty) tissue. Hence, microcalcification segmentation often involves thresholds that are adaptive to the local noise level. As an alternative to adaptive thresholds, several researchers proposed preprocessing steps that equalize the noise level in mammograms. Karssemeijer was the first to propose a noise equalization approach for screen-film mammograms in 1993,[43] which was later extended by Veldkamp and Karssemeijer[44] as well as Netsch and Peitgen.[28] Finally, in 2004 McLoughlin *et al.*[45] proposed an approach for noise equalization especially targeted toward modern full-field-digital mammograms.

II.H. Discussion

Despite the multitude of approaches and their considerable success, the problem of segmenting individual microcalcification particles in a ROI is not completely solved. The automatic delineation of the contours of particles with low contrast to the background tissue remains a major challenge. High noise levels in some mammograms and curvy-linear structures (like ducts and vessels) also complicate the calcification segmentation task. Furthermore, many approaches have only been evaluated on small and/or proprietary data sets and direct comparisons of the performance of competing methods on the same set of data are rarely provided. Moreover, lesion segmentation approaches are usually evaluated indirectly by comparing the classification performance of a full CADx system using different segmentation modules. This is due to the fact that acquiring pixel-exact ground truth for the segmentation task is very difficult because of the number, small size, and low contrast of the calcification particles in a cluster. However, a standalone evaluation and comparison of segmentation approaches based on pixel-exact ground truth would allow to compare the performance of different approaches independent of the other modules (e.g., feature extraction and selection) of a CADx system.

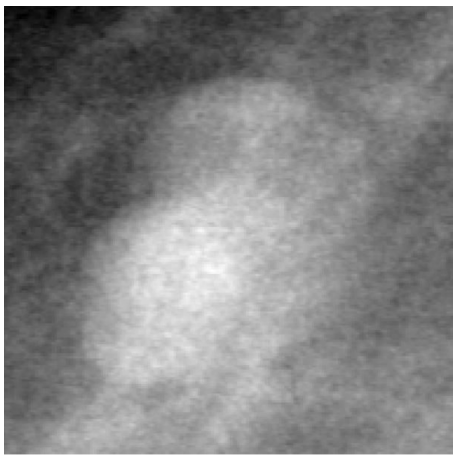
Both the LoG and DoG approaches return the position and size but not the exact shape of individual particles. Hence, often a second segmentation step that returns the shape of a particle based on its position and size is employed. In contrast, approaches based on morphological operations or the wavelet transform are more suited for returning the exact shape of individual particles. Furthermore, noise equalization or adaptive thresholds seem to be crucial aspects for successful segmentation of microcalcifications due to the heterogeneous noise levels present in mammograms. Table I provides a summary of the qualitative strengths and weaknesses of the classes of microcalcification segmentation methods mentioned in this section.

III. SEGMENTATION OF MAMMOGRAPHIC MASSES

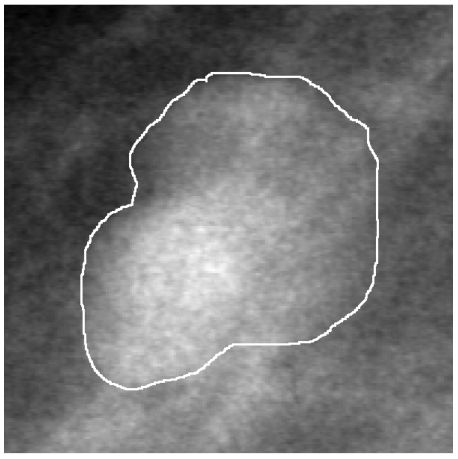
The shape and the margin characteristics of mammographic masses are crucial features for the discrimination of benign and malignant forms of this lesion type. Given a ROI containing a mammographic mass, most techniques for the extraction of features that represent characteristics of a mass' shape or margin require a segmentation of the mass from the background tissue. However, compared to the segmentation of microcalcifications, the segmentation of masses is more difficult because of their often fuzzy and highly irregular contours and their low contrast. Similar to the state of the art of microcalcification segmentation, a large variety of approaches to the problem of mass segmentation have been proposed in the past two decades. Hence, an overview and taxonomy of a representative selection of methods is given in this section. Figure 3(a) is an example of a ROI depicting a mammographic mass and Fig. 3(b) shows the results of an automatic segmentation of the mass from the background tissue. Part of the lesion's margin is obscured by tissue superpositions; a major challenge for automatic mass segmentation approaches.

III.A. Semiautomatic methods

Especially in early CADx approaches,[46,47] the segmentation problem was solved totally or partially by hand: Mass borders are outlined by hand or using semiautomatic pro-



(a)



(b)

FIG. 3. An example of a ROI depicting a mammographic mass (a) and the results of an automatic segmentation of the mass from the background tissue (b). Part of the lesion's margin is obscured by tissue-superpositions; a major challenge for automatic mass segmentation approaches.

cesses that require mandatory manual steps. An example for semiautomatic approaches is the work by Kilday *et al.*[48] who manually applied gray-level thresholds.

III.B. Region growing

Region growing is one of the methods that was most often applied to the segmentation of masses. It is a bottom-up segmentation method in which, starting from a seed pixel, neighboring pixels are iteratively added to the foreground region if they fulfill a similarity criterion. Many variations in the basic region growing approach have been proposed for the segmentation of masses. Giger *et al.*[49] proposed a semiautomatic region-growing approach. Huo *et al.*[50] later automated this approach by automatic background correction and seed point definition. Kupinski and Giger[51] proposed two region-growing approaches based on the radial gradient index (RGI) and a probabilistic model, respectively. Petrick *et al.*[52] combined a contrast enhancement filter and region growing for the segmentation of masses. Sahiner *et al.*[53] applied region growing as a postprocessing step for a segmentation approach based on pixel-by-pixel K-means clustering. Mendez *et al.*[54] as well as Pohlman *et al.*[55] also applied region growing to the segmentation of masses. Finally, Guliato *et al.*[56] combined fuzzy set theory and region growing in a method that takes the uncertainty, present at the boundaries of tumors, into account.

III.C. Active contours

In contrast to region-based approaches, which divide the ROI into mass and background pixels, contour-based approaches try to detect the mass' boundary. A variety of contour-based approaches for mass segmentation are based on active contour models (also called dynamic contours). The general idea of dynamic contour models is to approximate the boundary of a mass by minimizing the energy function of a closed contour consisting of connected line segments. The energy function usually has internal and external energy terms. The internal energy represents the curvature of the contour, while the external energy represents image features like the presence of edges. Active contour approaches strongly depend on the choice of the initial contour. te Brake and Karssemeijer[57] proposed a discrete dynamic contour model that was initialized with a circular contour of a fixed size. Sahiner *et al.*[58,59] initialized an active contour model with an initial contour obtained using K-means clustering.

III.D. Level sets

Level set approaches are a class of segmentation methods that are related to active contour models. In fact these approaches are sometimes referred to as implicit active contours in literature. Rather than evolving a contour itself, in level set approaches, a contour is represented as the zero level of a higher-dimensional scalar function. A two-dimensional contour for example can be represented by the zero level of a three-dimensional cone-shaped surface. The surface is designed such that its intersection with the *xy*-plane matches the represented contour. The main advantage of level set approaches over traditional active contour models is that they can easily handle topological changes like splitting or merging of parts of the contour. Shi *et al.*[60] recently proposed a level set approach for mass segmentation with an initial contour obtained by K-means clustering. Yuan *et al.*[61] initialized a level set approach using a contour obtained using the radial gradient index approach earlier proposed by Kupinski and Giger.[51] Their approach furthermore employs background trend correction and a dynamic stopping criterion.

III.E. Dynamic programming

Timp and Karssemeijer[62] proposed a segmentation approach based on boundary tracing using dynamic programming that guarantees resulting contours to be closed. A key component of dynamic programming approaches for the delineation of a contour is the cost function, which is employed to obtain the path that most efficiently represents the contour of the object that is to be delineated. Timp and Karsse-

TABLE II. A summary of strengths and weaknesses of the classes of mass segmentation methods considered in this section.

Method class	Interaction	Strengths	Weaknesses
Drawing manually	Intense	Can be very accurate	Cumbersome, high interobserver variability
Adjusting manually	Low to medium	Less cumbersome, can be very accurate	High interobserver variability
Region growing	None to some	Simple implementation, easy to comprehend	Fuzzy borders can easily cause leakage
Active contours	None	Closed contours	Initial contour critical
Level sets	None	Closed contours, flexible to topological changes	Initial contour critical, computationally expensive
Dyn. programming	None	Very flexible via (combination of) cost functions	Often no closed contours, computationally expensive

meijer's approach was later applied by Varela *et al.*[63] and recently extended by Dominguez and Nandi,[64] who proposed two modified versions of the method that achieve superior performance by modifications to the cost function.

III.F. Discussion

While encouraging progress was made in the segmentation of mammographic masses in the past years, the task cannot be considered to be solved. In fact this is even less the case for mammographic masses than for microcalcifications. An inherent property of x-ray mammography—three-dimensional anatomical structures are projected onto a two-dimensional image plane—probably is the main cause for the complexity of mass segmentation in mammograms. As previously discussed in the section on calcification segmentation, more comparative studies of the performance of competing segmentation approaches are needed. Such studies require reference databases with pixel-exact ground truth and standardized performance metrics. While Timp and Karssemeijer,[62] for example, used a criterion, which simply measures the overlap of the automatically segmented area of a mass and the ground truth area defined by a radiologist, Dominguez and Nandi[64] employed a metric that combines an overlap criterion with measures of under- and oversegmentation. Fortunately, there are already some studies that directly compare different mass segmentation approaches. te Brake and Karssemeijer[57] compared their discrete dynamic contour approach with the region-growing approaches proposed by Kupinski and Giger.[51] Timp and Karssemeijer[62] compared their dynamic programming approach with both an active contour and a region-growing method. The mean overlap percentage for the proposed dynamic programming was 0.69, for the other two methods 0.60 and 0.59, respectively. Shi *et al.*[60] compared a level set and an active contour approach and found the level set approach to significantly outperform the active contour method. Yuan *et al.*[61] compared the performance of their level set approach with the RGI approach proposed by Kupinski and Giger.[51] They employed the percentage of correctly segmented ROIs for a given overlap threshold and found the level set approach to outperform RGI at a threshold of 0.4 with 85% compared to 73%. However, a direct comparison of the performance of novel approaches is still hard to achieve, given the fact that rarely source code for proposed methods is made publicly available for research purposes. Table II provides a summary of the qualitative strengths and weaknesses of the classes of mass segmentation methods mentioned in this section. Given the inherent difficulties of mass segmentation, it is questionable that a fully automatic and at the same time robust approach to the problem can be found. This implies that minimally interactive solutions, which provide simple but effective ways for the radiologist to adjust the segmentation results, might be a reasonable direction for future research.

IV. FEATURE EXTRACTION FOR CLUSTERED MICROCALCIFICATIONS

In this section a survey of feature extraction methods, which have been proposed for the characterization of clustered microcalcifications, is provided. Radiologists usually characterize microcalcification clusters based on the morphology and location of the cluster, the morphologies of the individual calcification particles, and the distribution of the particles within the cluster. Most approaches to the automatic characterization of calcifications mimic the radiologist's strategy, and hence, four major classes of features for the discrimination of microcalcification clusters can be identified: Features that represent the morphology and location of the cluster, features that characterize the morphology and optical density of the individual calcification particles, features that describe the spatial distribution of the individual particles within the cluster, and finally features that represent the texture of the background tissue the calcifications are embedded in. The rest of this section is structured based on this taxonomy.

IV.A. Morphology of the cluster

Several researchers proposed to calculate the convex hull of the centroids or the contour pixels of the particles in a cluster as a representation of the cluster's shape.[14,31,65] Based on this shape representation, the area and the perimeter as well as the circularity, rectangularity, orientation, and eccentricity of the cluster can be obtained.[31,65] Moreover, normalized central moments are often obtained from a shape representation of the cluster.[66] A basic feature that describes the morphology of a microcalcification cluster is the number of individual calcification particles in the cluster which can trivially be obtained from a segmentation of the individual calcifications; it is employed by almost all approaches to the characterization of calcification clusters (e.g., Refs. 10, 12, 31, 65, 67, and 68). Besides the number of calcifications, often the calcification coverage, which is defined as the ratio

of the sum of the individual calcification areas and the cluster area, is used to describe how densely packed a cluster is with calcifications.

IV.B. Location of the cluster

The probability of malignancy of lesions like calcification clusters also depends on their location in the breast. For example, it is a well established fact that malignant lesions are more often located in the upper outer quadrant than other quadrants of the breast.[69] Thus, some approaches obtain features that are based on the location of a microcalcification cluster in the mammogram. Veldkamp *et al.*[67] for example, proposed to use the relative distance of a cluster to the pectoral muscle and the breast edge as discriminative features. Note that location-based features usually require a robust segmentation of landmarks (like the mamilla, pectoral muscle, and breast boundary) in the mammograms, which is a nontrivial problem of its own.

IV.C. Morphology of individual calcifications

Instead of features describing the morphologies of individual calcification particles, usually the statistics (like the mean, standard deviation, minimum, maximum, or median) of features of the individual particles in a cluster are obtained. For example, often the means and standard deviations of the areas, perimeters, circularities, rectangularities, orientations, and eccentricities of the particles are employed.[10,12,13,32,36,65–67] Furthermore, the means and standard deviations of the individual normalized central moments as well as of the moments of the border pixels are used by some researchers.[9,32,66,70,71] Moreover, Betal *et al.*[31] proposed to use mathematical morphology to analyze the shape of calcification particles and extracted the features representing the number of in-foldings, the elongation as well as narrow and wide irregularities of particles using this approach. Note that the accuracy of features describing the morphology of individual particles requires a robust segmentation and due to the discrete nature of images generally decreases with decreasing size of the particles. Shen *et al.*[9] proposed to use features based on normalized Fourier descriptors of the contours of calcifications as descriptions of the shape of individual particles. This approach was later employed by others.[66] Yet another method for the description of the shape of calcification particles was proposed by Bocchi and Nori,[71] who derived transformation and rotation invariant shape features using the Radon transform.

IV.D. Optical density of individual calcifications

The optical density of calcification particles is often characterized by statistics of the mean and variance of the gray values as well as the contrast of individual particles.[12,14,31,65,67,70] Furthermore, Jiang *et al.*[10] proposed to measure the effective thickness and volume of individual particles based on a mathematical model of image formation. They found the standard deviations of the effective thicknesses and volumes to be useful discriminants.

IV.E. Distribution of individual calcifications

Basic features describing the spatial distribution of the particles within a cluster are statistics (again usually the mean and standard deviation) of the distances between individual particles[12,14] as well as the distances of the particle centroids to the cluster centroid.[31,65] Also sometimes employed are the eccentricity and the normalized central moments of the particle centroids. A more complex approach was proposed by Leichter *et al.*,[13] who obtained the mean number of nearest neighbors of the particles in a cluster using a two-dimensional Delaunay triangulation of the particle centroids.

IV.F. Texture of the background tissue

Several approaches to the discrimination of benign and malignant microcalcification clusters assume that the presence of calcifications alters the texture of the tissue surrounding the calcifications. Furthermore, some approaches presume that the pixel pattern of calcification particles in a cluster itself can be represented by textural features. Therefore, the use of texture features has been proposed for the classification of calcification clusters. A brief overview of the most commonly employed texture feature extraction methods is given here. Note that for the extraction of texture features, a segmentation of the individual calcification particles is not necessary. Hence, this group of features, in contrast to the features described so far, does not depend on the robustness of the employed segmentation approach.

Probably the most popular class of texture features are those derived from gray-level co-occurrence matrices, which represent second-order statistics of the gray levels in a ROI, as described by Haralick *et al.*[72] They are employed to characterize microcalcifications by a broad range of researchers.[12,32,70,73–75] Texture analysis based on wavelet packets was first proposed by Laine and Fan[76] and is employed for microcalcification characterization by some groups.[12,70] Furthermore, Soltanian *et al.*[70] compared the discriminative power of standard wavelet packet features with multiwavelet packet features and found the latter, which use multiple scaling functions and mother wavelets, to be superior.

IV.G. Discussion

A broad variety of features for the characterization of microcalcifications have been developed in the past years. The number of particles, their shape, and their distribution within the cluster seem to have high discriminative power. However, due to the lack of direct comparisons of the features developed by different research groups on the same data, exact quantitative comparisons are hardly possible. Hence, we see a demand for a study that compares a broad range of features from different feature groups (morphology, distribution, optical density, and texture of the background) on a large, public dataset. Such a study would allow researches to directly compare the discriminative power of newly developed features with the state of the art. Furthermore, ap-

proaches to microcalcification segmentation often have limited success in exactly delineating the shape of individual calcification particles; hence more researches seem to be necessary to measure the robustness of shape features against segmentation errors.

V. FEATURE EXTRACTION FOR MAMMOGRAPHIC MASSES

This section contains a survey of feature extraction methods for mammographic masses. Similar to the situation regarding clustered microcalcifications, most approaches to the feature extraction for mammographic masses are based on the lesion attributes that are used for lesion characterization by radiologists. Radiologists characterize masses based on their shape, the characteristics of their margin, and their optical density.[77] A broad range of techniques for the extraction of features that resemble lesion attributes used by radiologists has been proposed. However, higher-order features that do not directly resemble attributes used by radiologists are also employed.

V.A. Shape

Based on a segmentation of the mass contour, several groups employed basic morphological features to represent the shape of a mass.[46–48,52,78] Similar to the approaches used to represent the morphology of individual calcification particles, these include the area and the perimeter as well as the circularity, rectangularity, orientation, and eccentricity of the mass. Furthermore, normalized central moments and moments of the border pixels of a mass are sometimes used to represent the shape.[46] Kilday *et al.*[48] introduced a set of features based on the normalized radial length (NRL) to represent the mass shape. The NRL is defined as the Euclidean distance of each pixel on the object contour to the object's centroid. The NRL feature set was later evaluated and applied by several other groups.[47,52,58,79] Another class of shape features, which is often applied,[46,59] is based on Fourier descriptors of the mass' contour.

V.B. Margin characteristics

Huo and co-workers[50,80,81] proposed two features that measure the amount of spiculation of a mass based on an analysis of the radial gradient of its contour. Furthermore, two features that describe the margin characteristics of a mass have been proposed by Zheng *et al.*,[79] who simply obtained the standard deviation as well as the skew of the gradient strength of the pixels on the contour of a mass. Spiculated margins are a major characteristic of malignant masses; hence, several groups proposed to apply texture analysis on bands of pixels that are close to the margin of a segmented mass. Extending this idea, Sahiner *et al.*[53] developed a transformation called the rubber-band-straightening transform (RBST) that transforms a band of margin pixels onto the Cartesian plane. They showed that texture features extracted from this transformed image, in which spiculations approximately resemble vertical lines, are superior to texture features extracted from the untransformed band of pixels. Texture features based on the RBST have later been applied in some other approaches.[82,83] The sharpness of a mass' margin is recognized as an important mammographic lesion attribute. Shi *et al.*[60] developed a margin abruptness feature that measures the margin sharpness using line detection in RBST images. Rangayyan *et al.*[46] also designed a feature called acutance that measures the sharpness or abruptness of the mass margin. Varela *et al.*[63] proposed mass margin features that measure the sharpness of the margin as well as the presence of microlobulations. They found the performance of the microlobutation features to be lower than the performance of the sharpness features, which was expected as the presence of microlobulations is known to be a less specific attribute for the discrimination of benign and malignant masses. Mudigonda *et al.*[84] also proposed two gradient-based features that measure the sharpness of a mass' margin by analyzing the image gradient in a ribbon of pixels surrounding the mass' contour.

V.C. Optical density

The optical density of a mass is often represented by simple features like the mean gray level of the mass region, its local contrast, or by texture analysis features.[63,79,80,85]

V.D. Texture

Examples for higher-order features that do not represent lesion attributes used by radiologists are features obtained by texture analysis. Similar to feature extraction for clustered calcifications, as described in the previous section, features based on the gray-level co-occurrence matrix, gray-level run-length metrics and wavelet decompositions have been popular choices for the characterization of masses.[53,63,83,84,86–88] However, in contrast to the diagnosis of calcifications, in mass CADx approaches, texture analysis is not always performed on the full ROI but often only on special regions within the ROI. Based on a segmentation of the mass from the background, texture analysis is often restricted to the mass region, excluding the background tissue region or on bands of pixels close the mass' margin.

V.E. Embedded calcifications

Mammographic masses, especially malignant ones, often contain calcifications. For many types of mammographic masses, the presence and the characteristics of embedded calcifications are highly discriminative attributes. Still, only few approaches to CADx of masses incorporate calcification features. One of the few examples is the work of Shi *et al.*,[60] who proposed to use an attribute that represents the number of microcalcifications that are present in a mass ROI as a feature.

V.F. Discussion

Features that measure the amount of spiculation and the sharpness of the border seem to have the highest discriminative power for mammographic masses. The appearance of

TABLE III. A summary of strengths and weaknesses of three feature selection approaches.

Method class	Strengths	Weaknesses
Genetic algorithms	Can escape from local minima	More parameters, needs careful adjustment
Forward/backward selection	Efficient	Often stuck in local minima
Stepwise LDA	Efficient	Often stuck in local minima

spiculations seems to be very variable, as the number, thickness, and frequency of the spiculi vary considerably. A systematic study of the range of these parameters in spiculated masses might help in the design and implementation of appropriate lesion features. Again, direct comparisons of the features developed by different research groups on the same data are rare, and a large study that compares the discriminative power of a broad variety of features on a large, public dataset [like the DDSM (Ref. 89)] is still missing. Because of the fact that the robustness of the segmentation results for masses seems to be limited, the development of features that do not require explicit segmentations might be an interesting field for future research. Furthermore, the location of both masses and microcalcification clusters in relation to mammographic landmarks (like the breast border or the pectoral muscle) is a characteristic that is often neglected by CADx approaches but seems to have considerable discriminative power.

VI. FEATURE SELECTION

A well known machine learning problem is the fact that classification or regression performance often decreases with an increase in the feature space's dimension. This effect, which is called the "curse of dimensionality," is due to the fact that with increasing dimension of the feature space, the distribution of instances becomes increasingly sparse. The discriminative power of features employed in CADx systems varies: While some are highly significant for the discrimination of mammographic lesions, others are redundant or even irrelevant. Hence, automatic selection of a subset of features from a higher-dimensional feature vector is a common module in mammography CADx approaches, as well as in pattern recognition systems in general.

Selecting the optimal feature subset for supervised learning problems requires an exhaustive search of all possible subsets of features of the chosen cardinality, which is not practical in most situations because the number of possible subsets given N features is 2^N-1 (the empty set is excluded). Hence, in practical machine learning applications, usually a satisfactory instead of the optimal feature subset is searched.

There are two classes of practical approaches to feature selection. Approaches of the first class are called *filter approaches* and are generally independent of the employed classifier. In filter approaches, features are selected based on a metric, which usually measures their discriminative power. The second class of approaches, called *wrapper approaches*, wrap the target classifier to find a good feature subset using a nonexhaustive search strategy. Two wrapper approaches and one filter approach are most commonly used in mammography CADx systems. Several groups employed genetic algorithms (GAs) for the feature selection task in CADx systems.[12,70,75,79,82] Besides GAs, sequential forward selection and the related sequential backward elimination approaches, which both are sequential feature selection methods, are popular feature selection strategies in the CADx community.[32,67,78] A filter approach, selection using stepwise linear discriminant analysis (LDA) is also a popular method[58–60,74,75,83] to decrease the dimension of the feature vector.

The described feature selection techniques are heuristics that retrieve a satisfactory and not a globally optimal feature subset. The genetic algorithm approach is known to get stuck in local minima less often than the other two approaches. This is mainly due to the mutation operator, which adds a stochastic element to the search strategy. In contrast the search strategies of both sequential feature selection and stepwise LDA are purely deterministic. However, the GA approach needs careful adjustments of its parameters (the mutation and crossover probabilities, the number of generations and individuals, etc.) and tends to evaluate more candidate feature sets which can be computationally expensive. Table III summarizes the strengths and weaknesses of the discussed feature selection approaches.

VII. CLASSIFICATION

The discrimination of benign and malignant mammographic lesions is a supervised learning problem, which is defined as the prediction of the value of a function for any valid input after training a learner using examples of input and target output pairs. For the problem at hand, the function has only two discrete values: Benign or malignant. Hence the problem of discriminating benign and malignant lesions can be modeled as a two-class classification problem.[90] A variety of classifiers have been applied in the state of the art CADx approaches to solve this problem. The k nearest neighbors classifier is not only one of the most commonly employed classifiers for the discrimination of mammographic lesions[9,12,67,79,91] but also one of the simplest and most popular classifiers in general. Artificial neural networks (ANNs) seem to be the most commonly used type of classifiers in mammography CADx systems.[12,14,32,50,63,65,66,71,74,80,88,92–95] Support vector machines[32,65,88] and LDA[12,47,53,58–60,73,75,78,82,95,96] are also popular in the CADx community. Examples of classifiers that are less frequently applied in CADx systems include Bayes classification,[36] generalized dynamic fuzzy neural networks,[86] and rule-based expert systems.[65] Furthermore, Wei *et al.*[68] applied two state of the art kernel-based classifiers, kernel Fisher discriminant,

and relevance vector machines for the discrimination of mammographic lesions. Moreover, they have investigated several committee machines, like boosting and bagging,[90] for CADx approaches. Hadjiiski *et al.*[83] investigated a hybrid approach for the classification of mammographic lesions that combines LDA with an adaptive resonance theory network (an unsupervised, self-organizing neural network).

While the mammographic lesion discrimination performance of some classification approaches has been directly compared,[12,32,65,68] the results of these comparisons (which are of course based on systems using different segmentation, feature extraction steps and evaluated on different data) often contradict. Hence, no distinctive conclusion about which classifier is best suited for this task can be drawn. This finding correlates with the observation that for most classification problems the selection of features usually has more influence on the classification performance than the choice of classifier.

VIII. TEMPORAL CHANGE ANALYSIS

Interval change analysis, the comparison of serial mammograms for the detection and diagnosis of mammographic lesions, is routinely used by radiologists in clinical practice. The comparison of current and prior mammograms is reported to significantly improve the performance of radiologists both in the detection[97] as well as in the diagnosis of mammographic lesions.[98] Therefore, several groups investigated the use of prior mammograms in CADe (Ref. 99) as well as in CADx systems.[100,101] Hadjiiski *et al.*[100] used interval change information for the classification of benign and malignant masses. They extracted texture features from mass ROIs cut from the prior and the current mammograms, as well as difference features by subtracting the texture features of the prior from the features of the current mammograms. A comparison of this approach with a system that only used features from the current mammogram resulted in a significant increase in ROC performance (A_z increased from 0.82 to 0.88). Timp *et al.*[101] also discussed how the inclusion of temporal change information improves the performance of a mass CADx system. Their approach includes two kinds of temporal features: Difference and similarity features. While difference features indicate the change in feature values determined on prior and current views, similarity features measure to what extend two regions are comparable in appearance and hence may be useful for lesions that are visible on the prior view as well as for newly developing lesions. They also found a significant increase in A_z from 0.74 to 0.77 when temporal features were included in the classification. Interestingly, they found that the proposed similarity features particularly contributed to this increase in performance. This led to comparable improvements both for masses that were visible and those that were not visible on the prior view. Finding corresponding masses in prior and current mammograms and the registration (in terms of geometry and gray scale) of prior [often screen-film mammography (SFM)] and current [usually full-field-digital mammography (FFDM)] mammograms are important subproblems that need to be solved in CADx approaches that incorporate temporal change information. However, this is a difficult task, as the relative position and the appearance of a breast lesion is usually different during multiple examinations, due to the variable compression and deformation of the breast tissue during mammography. Sanjay-Gopal *et al.*,[102] Filev *et al.*[103] as well as Timp *et al.*[104] published work on the problem of finding corresponding lesions. Snoeren and Karssemeijer[105,106] as well as Engeland *et al.*[107] proposed solutions to the problem of gray scale and geometric registration of serial mammograms.

IX. MULTIVIEW AND MULTIMODAL CADx

Besides the inclusion of temporal change information into CADx systems as discussed in Sec. VIII, the incorporation of information from multiple mammographic views as well as from complementary modalities (e.g., breast sonography, breast MRI, and breast elastography) are important topics of current and future CADx research. In breast cancer screening, usually the two standard mammography views CC and MLO are acquired. Often additional views, like the mediolateral (ML) view as well as special view mammograms (e.g., spot compression or spot compression magnification views) are also available. Radiologists of course consider the appearance of a lesion in all available mammographic views, as well as in the images acquired using complementary modalities, in their diagnosis. Hence, CADx approaches should probably do so as well. While several CADx systems published so far consider information from the standard CC and MLO views (e.g. Refs. 67, 68, 108, and 109), systems that include lesion features from additional views or even from additional modalities are rare. Huo *et al.*[110] investigated the use of special view mammograms in the CADx of mammographic masses. Their results (A_z=0.95 using the special views, A_z=0.78 and A_z=0.75 using the CC and MLO views, respectively) indicate that the CADx of special view mammograms significantly improves the classification of masses. However, the ROC performance that was achieved when all three views were used (A_z=0.95) was not higher than when only the special view mammograms were used (A_z=0.95). This indicates that the CC and MLO views might not add significant diagnostic information to a system that already includes spot compression or spot compression magnification views. Drukker *et al.*[111,112] proposed a multimodal CADx approach that incorporates information from mammograms and breast sonography. They found that the system's performance significantly improved when lesion features from both modalities were combined. However, classification performance depended on specific methods for combining features from multiple images per lesion (mean, minimum, or maximum). They achieved a maximum area under the ROC curve of A_z=0.95 for the multimodal system.

Fully automatic merging of information from multiple mammographic views and multiple modalities of course requires automatic methods for finding corresponding lesion ROIs. Besides the work on finding corresponding lesions in mammographic views already mentioned in Sec. VIII, addi-

tional work has been published by Engeland *et al.*[113] which specifically considers the case of finding corresponding ROIs in MLO and CC mammograms.

Finding corresponding ROIs in different modalities is an even harder problem, still open for future research. Another question that is not fully solved is how features extracted from different views and modalities should be merged. Several approaches have been proposed so far (e.g., Ref. 111), but a definitive guide still seems to be unavailable. Another interesting topic is the effect of correlation when features from multiple sources are combined. Theoretical investigations on this topic can be found in the work of Liu *et al.*[114]

X. CADx BASED ON HUMAN-EXTRACTED FEATURES

Radiologists usually describe the attributes of a mammographic lesion using standardized lesion descriptions. In the past years computer-aided diagnosis systems have been proposed that use these kinds of lesion descriptions as human-extracted input features for CADx systems. In this section an overview of the CADx approaches based on human-extracted features that have been proposed so far is given. Early work dates back to 1993, when Wu *et al.*[92] proposed to use a set of 14 lesion descriptions extracted by radiologists as input features for an artificial neural network for the computer-aided diagnosis of mammographic masses and microcalcifications. They demonstrated that a simple three-layer feed-forward ANN using human-extracted input features can perform at a higher level than the average performance of a group of radiologists.

Later CADx approaches using human-extracted input features have usually been based on a standard set of lesion descriptions which have been published in the BI-RADS™ (Ref. 69) atlas, a quality assurance guide designed to standardize breast imaging reporting, by the American College of Radiology.

Baker *et al.*[115,116] proposed an ANN approach to deduce diagnosis proposals from BI-RADS™ lesion descriptions. Their approach was later extended and evaluated by others.[93,117–120] Alternative approaches based on case-based reasoning (CBR), constraint satisfaction neural network, and Bayesian networks were later proposed by Floyd *et al.*,[121] Bilska-Wolak *et al.*,[122,123] Tourassi *et al.*,[124] and Fischer *et al.*,[125] respectively. The prime advantage of the CBR approaches over the earlier proposed approaches is the transparent reasoning process that leads to the system's diagnosis suggestion. Elter *et al.*[126] proposed to use an entropic similarity metric to solve the problem of handling numerical and categorical features in a CBR system for breast cancer diagnosis. They have furthermore demonstrated that decision tree learners are a good alternative to the CBR approaches as the induced models are also very easy to understand for the radiologists. Recently, Gupta *et al.*[96] investigated using BI-RADS™ attributes from two mammographic views (MLO and CC) as input features for a CADx system based on a linear discriminant analysis classifier. Furthermore, Burnside *et al.*[127] investigated how well the lesion descriptors for microcalcifications in the BI-RADS™ atlas are suited to predict the risk of malignancy.

While mammography is the most important technique for breast cancer diagnosis, supplemental modalities like sonography or MRI are often applied in addition to mammography for the diagnosis of some types of mammographic masses. Hence it is no surprise that recent investigations by Jesneck *et al.*[95] suggest that the performance of CADx systems based on human-extracted features can significantly be improved if lesion descriptions from multiple modalities, like mammography and sonography, are combined.

A common problem that all CADx approaches based on human-extracted features face is the apparent inter- and intraobserver variability in lesion descriptions as described for example by Baker *et al.*[128] Recent work[126] shows that the performance of a CADx system based on BI-RADS attributes that was trained on a mammography database compiled in a European institution decreased when it was applied to data from the DDSM database which was compiled at US institutions and vice versa. This finding suggests that the performance of CADx systems based on human-extracted features might be limited by the apparent interobserver variability in lesion descriptions.

XI. PERFORMANCE EVALUATION

Receiver operating characteristic (ROC) curve analysis[129–131] is the standard methodology for the evaluation of the classification performance of CADx approaches. Most CADx systems perform a two-class classification (a ROI containing a lesion is either benign or malignant). The performance of such systems can be evaluated using classic ROC curve analysis. However, the output of CADe systems (automatically detected lesion ROIs) can also be used as input for CADx systems. In this case, which has been rarely considered so far,[132] three classes have to be distinguished (benign lesion, malignant lesion, and normal breast tissue) and three-class extensions of ROC analysis have to be applied. Tables IV and V provide summaries of the ROC performance of a representative selection of mass and calcification CADx systems, respectively. However, it is not possible to directly compare the classification performance of the listed approaches because they have not been trained and tested on the same data. Also note that besides the advancements in CADx approaches over the past decade, no clear trend of improving classification performance over time can be identified in Tables IV and V. This is an indication that ROC performance of CADx approaches strongly depends on the evaluation dataset.

CADx systems are designed to support the radiologist in the diagnosis of lesions by providing a diagnosis suggestion. Therefore, it is important to evaluate if and how the diagnoses of radiologists improve when CADx systems are used. Observer studies using multireader, multicase ROC analysis[138] are the de facto standard approach to this problem. In the following we provide an overview of the CADx observer studies published so far. Early observer studies have

TABLE IV. Overview of the ROC performance of a representative selection of mass CADx approaches. For each study, the year, the number of lesions *#l*, the setup (single view, multiview, and multimodal) and the (image-based) area under the ROC curve A_z are listed. Note that direct comparisons of the A_z values are not reasonable as the CADx systems have been evaluated on different databases.

	Year	# *l*	Setup	A_z
Shi *et al.* (Ref. 60)	2008	427	Multiview	0.85
Guliato *et al.* (Ref. 133)	2008	111	Singleview	0.94
Delogu *et al.* (Ref. 134)	2007	226	Singleview	0.78
Varela *et al.* (Ref. 63)	2006	1076	Singleview	0.81
Drukker *et al.* (Ref. 111)	2005	100	Multimodal	0.92
Timp and Karssemeijer (Ref. 62)	2004	1210	Multiview	0.74
Lim and Er (Ref. 86)	2004	343	Singleview	0.87
Sahiner *et al.* (Ref. 59)	2001	249	Single/multiview	0.87/0.91
Mudigonda *et al.* (Ref. 84)	2000	39	Singleview	0.85
Huo *et al.* (Ref. 81)	2000	110	Singleview	0.82
Hadjiski *et al.* (Ref. 83)	1999	348	Singleview	0.81
Sahiner *et al.* (Ref. 53)	1998	168	Singleview	0.94
Huo *et al.* (Ref. 80)	1998	95	Singleview	0.94
Huo *et al.* (Ref. 50)	1995	95	Singleview	0.83

TABLE V. Overview of the ROC performance of a representative selection of microcalcification CADx approaches. For each study, the year, the number of lesions *#l*, the setup (single view, multiview, and multimodal), and the (image-based) area under the ROC curve A_z are listed. Note that direct comparisons of the A_z values are not reasonable as the CADx systems have been evaluated on different databases.

	Year	# *l*	Setup	A_z
Karahaliou *et al.* (Ref. 135)	2007	100	Singleview	0.96
Weit *et al.* (Ref. 68)	2005	386	Multiview	0.85
Papadopoulos *et al.* (Ref. 65)	2005	105/25	Singleview	0.79/0.81
Soltanian-Zadeh *et al.* (Ref. 70)	2004	103	Singleview	0.89
Leichter *et al.* (Ref. 136)	2004	324	Singleview	0.87
Kallergi *et al.* (Ref. 66)	2004	100	Singleview	0.98
Salfity *et al.* (Ref. 29)	2003	131	Multiview	0.93
Markopoulos *et al.* (Ref. 94)	2001	240	Singleview	0.94
Veldkamp *et al.* (Ref. 67)	2000	280	Singleview	0.83
Chan *et al.* (Ref. 75)	1998	145	Singleview	0.89
Buchbinder *et al.* (Ref. 137)	1998	161	Singleview	0.88
Chan *et al.* (Ref. 74)	1997	86	Singleview	0.88
Betal *et al.* (Ref. 31)	1997	38	Multiview	0.84
Jiang *et al.* (Ref. 10)	1996	107	Singleview	0.83
Dhawan *et al.* (Ref. 12)	1996	191	Singleview	0.86

been published in 1999 by Chan *et al.*[108] and Jiang *et al.*[139] While the former evaluate the effects of a CADx system, developed at the University of Michigan, on radiologists' classification of masses, the later performed a similar evaluation for a microcalcification CADx approach developed at the University of Chicago. Both found statistically significant improvements in ROC performance when CADx was used as a diagnostic aid and concluded that their systems have the potential to assist radiologists in the classification of lesions and thus to potentially help to reduce unnecessary breast biopsies. In a follow-up study, Jiang *et al.*[140] demonstrated that their CADx approach, moreover, has the potential to reduce the variability among radiologists in the interpretation of mammograms, another problem of mammogram interpretation. More recently, they compared the use of their CADx approach with independent double reading by a second radiologist[141] and found significant improvements in ROC performance when their CADx system was used but only insignificant improvements when independent double reading was performed. Moreover, they even found that CADx improved diagnostic performance to an extent approaching the maximum possible performance. In another follow-up study, Rana *et al.*[142] applied the microcalcification CADx system, developed by Jiang *et al.*[139,140] on screen-film mammograms (SFMs), to full-field-digital mammograms. They concluded that the CADx system maintained consistently high performance in classifying calcifications in FFDMs mammograms without requiring substantial modifications from its initial development on SFMs. Huo *et al.*[109] and Leichter *et al.*[143] later published observer studies for mass CADx systems, developed at the University of Chicago and the Jerusalem College of Technology, respectively. In agreement with the work mentioned so far, they found that CADx can improve the diagnostic performance of radiologists. Observer studies have also been performed to evaluate a multimodal CADx approach as well as CADx approaches based on serial mammograms (temporal change analysis). Horsch *et al.*[112] evaluated the effect of a multimodal CADx workstation, which analyses mammograms and breast sonograms, on the diagnostic performance of radiologists and found significant improvements in ROC performance. Hadjiiski *et al.*[144,145] performed similar studies to evaluate a CADx system based on serial mammograms and interval change analysis, developed at the University of Michigan. Again, significant improvements of the radiologists' diagnosis accuracy are reported. Table VI provides an overview on the CADx observer studies mentioned above. Studies with special topics [e.g., interobserver variability,[140] comparison with double reading,[141] and SFM versus FFDM (Ref. 142)] are omitted.

While considerable evidence has been collected that CADx system have the potential to improve the diagnostic performance of radiologists, still no commercial CADx system is available today and open issues remain. The studies published so far are retrospective studies performed on relatively few cases (cp. Table VI). Hence, large prospective studies are still needed to further evaluate the effect of CADx systems on radiologists. Moreover, additional research on how exactly radiologists interact with CADx systems and how this interaction can be improved (e.g., by including semiautomatic modules, displaying reference lesions, including relevance feedback in lesion retrieval) would be interesting.

XII. CONCLUSION

The CADx community has made considerable progress on all aspects of mammography CADx systems in the past two decades. However, several open issues remain. Most no-

TABLE VI. Overview of CADx observer studies that investigate whether CADx improves the ROC performance of radiologists. For each study, the lesion type *lt* (masses or calcifications), the number of observers *#o*, the number of lesions *#l*, the setup (single view, multiview, temporal change analysis, multimodal) and the average of area under the ROC curve A_z of the observers without and with the CADx aid are listed. For the first study, instead of the average, the range of the observer's A_z values is listed.

	Year	*lt*	# *o*	# *l*	Setup	A_z without CAD	A_z with CAD
Chan *et al.* (Ref. 108)	1999	m	6	103	Singleview	[0.79,0.92]	[0.87,0.96]
Jiang *et al.* (Ref. 139)	1999	c	10	104	Singleview	0.61	0.75
Leichter *et al.* (Ref. 143)	2000	m	1	40	Singleview	0.66	0.81
Huo *et al.* (Ref. 109)	2002	m	12	110	Multiview	0.93	0.96
Hadjiski *et al.* (Ref. 144)	2004	m	10	97	Multiview, temporal	0.79	0.84
Hadjiski *et al.* (Ref. 145)	2006	m	10	90	Multiview, temporal	0.83	0.87
Horsch *et al.* (Ref. 112)	2006	m	10	97	Multimodal	0.87	0.92

tably, the progress that is made by newly proposed algorithms often is hard to measure due to the lack of direct comparisons of the performance of competing approaches on consistent and publicly available data sets. Such comparisons of course would not only require more publicly available reference databases but also that CADx researchers evaluate their algorithms on these data sets or on clearly defined subsets. While the DDSM database[89] provides a large public set of screen film mammograms, unfortunately no comparable reference database of modern full-field-digital mammograms is publicly available today. Because the performance of a CADx approach can vary dramatically over different databases, as it depends on factors such as the number and subtlety of the cases, standardized evaluation databases and evaluation techniques are crucial factors for the comparison of competing CADx approaches.

The CADx of mammographic lesions remains a challenging task. Especially the robust automatic segmentation of mammographic lesions is far from trivial. Particularly challenging are the accurate segmentation of mammographic masses with ill-defined or obscured borders and the segmentation of microcalcification particles with low contrast to the background tissue.

Furthermore, besides the classification performance, the transparency of the decision process of a CADx system to the radiologist, which is believed to strongly influence the probability that a given CADx system is accepted by radiologists in clinical practice, is another important factor in the design of CADx systems. Systems that are built on case-based reasoning, for example, seem to be more intelligible to the radiologist than systems based on neural networks or support vector machines. Improved human-computer interaction of CADx systems, in general, is an interesting field for future research, as most CADx systems proposed so far are one-size-fits-all tools.

FFDM is currently replacing analog SFM in clinical practice. FFDM is known to have higher contrast but lower resolution than SFM. While initial results suggest that CAD approaches developed on SFMs can be applied to FFDM with minor modifications,[142,146] more researches on the implications of the adoption of FFDM on both CADe and CADx approaches are required.

Radiologists usually consider information from multiple mammographic views, prior mammograms, complementary modalities (e.g., ultrasound and MRI) and clinical patient data in their diagnosis of a breast lesion. Hence, CADx systems should probably do so as well. While considerable progress has been made on integrating prior mammograms (temporal change analysis), features from additional modalities and patient data into CADx approaches, the CADx system of the future should incorporate all available sources of information into the decision making process. This not only requires the integration of the various CADx approaches that have been developed independently so far but also seamless integration of CADx systems into the clinical IT infrastructure (e.g., PACS, RIS, and HIS) for access to all required data. Furthermore, robustly finding corresponding ROIs depicting a lesion in different views and modalities is an unsolved problem.

Table VII summarizes some open research topics, urgent demands for further research and major challenges.

Even though CADx approaches have matured consider-

TABLE VII. A summary of demands for future CADx research and major challenges.

Research topic	Urgent demands	Major challenges
Validation	Freely available data and tools	Consensus, costs
Segmentation	Robust automatic segmentation	Accurate segmentation of masses with ill-defined or obscured borders
Feature extraction	Large comparative study	Development of segmentation-independent features
System	Integrate information from multiple views, additional modalities and prior examinations	Robust registration of lesions in different views and modalities
User interaction	Intuitive interfaces	Seamless integration in clinical workflows

ably in the past years, still no FDA approved system is on the market and in clinical use today. One reason is that the use of CADx systems as diagnostic aids has very serious implications on the patient management and thus almost perfect performance is required. While the positive effect of CADx systems on the diagnostic performance of radiologists has been demonstrated in several observer studies, in contrast to CADe systems, large prospective clinical studies are still missing.

[a]Telephone: +49 9131 776 7327; Fax: +49 9131 776 7399; Electronic mail: matthias.elter@iis.fraunhofer.de
[b]Telephone: +49 89 4140-4330.

[1]American Cancer Society, *Breast Cancer Facts & Figures 2007–2008* (American Cancer Society, Atlanta, 2007).
[2]R. M. Nishikawa, "Current status and future directions of computer-aided diagnosis in mammography," Comput. Med. Imaging Graph. **31**(4–5), 224–235 (2007).
[3]G. Hermann, C. Janus, I. S. Schwartz, B. Krivisky, S. Bier, and J. G. Rabinowitz, "Nonpalpable breast lesions: Accuracy of prebiopsy mammographic diagnosis," Radiology **165**(2), 323–326 (1987).
[4]D. B. Kopans, "The positive predictive value of mammography," AJR, Am. J. Roentgenol. **158**(3), 521–526 (1992).
[5]D. D. Adler and M. A. Helvie, "Mammographic biopsy recommendations," Curr. Opin. Radiol. **4**(5), 123–129 (1992).
[6]J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," J. Natl. Cancer Inst. **95**(18), 1384–1393 (2003).
[7]W. E. Barlow, C. Chi, P. A. Carney, S. H. Taplin, C. D'Orsi, G. Cutter, R. E. Hendrick, and J. G. Elmore, "Accuracy of screening mammography interpretation by characteristics of radiologists," J. Natl. Cancer Inst. **96**(24), 1840–1850 (2004).
[8]R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," Radiology **184**(3), 613–617 (1992).
[9]L. Shen, R. M. Rangayyan, and J. L. Desautels, "Application of shape analysis to mammographic calcifications," IEEE Trans. Med. Imaging **13**(2), 263–274 (1994).
[10]Y. Jiang, R. M. Nishikawa, D. E. Wolverton, C. E. Metz, M. L. Giger, R. A. Schmidt, C. J. Vyborny, and K. Doi, "Malignant and benign clustered microcalcifications: Automated feature analysis and classification," Radiology **198**(3), 671–678 (1996).
[11]S. Paquerault, L. M. Yarusso, J. Papaioannou, Y. Jiang, and R. M. Nishikawa, "Radial gradient-based segmentation of mammographic microcalcifications: Observer evaluation and effect on cad performance," Med. Phys. **31**(9), 2648–2657 (2004).
[12]A. P. Dhawan, Y. Chitre, and C. Kaiser-Bonasso, "Analysis of mammographic microcalcifications using gray-level image structure features," IEEE Trans. Med. Imaging **15**(3), 246–259 (1996).
[13]I. Leichter, R. Lederman, P. Bamberger, B. Novak, S. Fields, and S. S. Buchbinder, "The use of an interactive software program for quantitative characterization of microcalcifications on digitized film-screen mammograms," Invest. Radiol. **34**(6), 394–400 (1999).
[14]F. Schmidt, E. Sorantin, C. Szepesvàri, E. Graif, M. Becker, H. Mayer, and K. Hartwagner, "An automatic method for the identification and interpretation of clustered microcalcifications in mammograms," Phys. Med. Biol. **44**(5), 1231–1243 (1999).
[15]L. P. Clarke, M. Kallergi, W. Qian, H. D. Li, R. A. Clark, and M. L. Silbiger, "Tree-structured non-linear filter and wavelet transform for microcalcification segmentation in digital mammography," Cancer Lett. **77**(2–3), 173–181 (1994).
[16]W. Qian, M. Kallergi, L. P. Clarke, H. D. Li, P. Venugopal, D. Song, and R. A. Clark, "Tree structured wavelet transform segmentation of microcalcifications in digital mammography," Med. Phys. **22**(8), 1247–1254 (1995).
[17]R. N. Strickland and H. I. Hahn, "Wavelet tranforms for detecting microcalcifications in mammograms," IEEE Trans. Med. Imaging **15**(2), 218–229 (1996).
[18]H. Yoshida, K. Doi, R. M. Nishikawa, M. L. Giger, and R. A. Schmidt, "An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms," Acad. Radiol. **3**(8), 621–627 (1996).
[19]J. J. Heine, S. R. Deans, D. K. Cullers, R. Stauduhar, and L. P. Clarke, "Multiresolution statistical analysis of high-resolution digital mammograms," IEEE Trans. Med. Imaging **16**(5), 503–515 (1997).
[20]W. Zhang, H. Yoshida, R. M. Nishikawa, and K. Doi, "Optimally weighted wavelet transform based on supervised training for detection of microcalcifications in digital mammograms," Med. Phys. **25**(6), 949–956 (1998).
[21]W. Qian, L. P. Clarke, D. Song, and R. A. Clark, "Digital mammography: Hybrid four-channel wavelet transform for microcalcification segmentation," Acad. Radiol. **5**(5), 354–364 (1998).
[22]T. C. Wang and N. B. Karayiannis, "Detection of microcalcifications in digital mammograms using wavelets," IEEE Trans. Med. Imaging **17**(4), 498–509 (1998).
[23]M. J. Lado, P. G. Tahoces, A. J. Méndez, M. Souto, and J. J. Vidal, "A wavelet-based algorithm for detecting clustered microcalcifications in digital mammograms," Med. Phys. **26**(7), 1294–1305 (1999).
[24]S. Yu and L. Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," IEEE Trans. Med. Imaging **19**(2), 115–126 (2000).
[25]P. Heinlein, J. Drexl, and W. Schneider, "Integrated wavelets for enhancement of microcalcifications in digital mammography," IEEE Trans. Med. Imaging **22**(3), 402–413 (2003).
[26]G. Lemaur, K. Drouiche, and J. DeConinck, "Highly regular wavelets for the detection of clustered microcalcifications in mammograms," IEEE Trans. Med. Imaging **22**(3), 393–401 (2003).
[27]E. Regentova, L. Zhang, J. Zheng, and G. Veni, "Microcalcification detection based on wavelet domain hidden markov tree model: study for inclusion to computer aided diagnostic prompting system," Med. Phys. **34**(6), 2206–2219 (2007).
[28]T. Netsch and H. O. Peitgen, "Scale-space signatures for the detection of clustered microcalculations in digital mammograms," IEEE Trans. Med. Imaging **18**(9), 774–786 (1999).
[29]M. F. Salfity, R. M. Nishikawa, Y. Jiang, and J. Papaioannou, "The use of a priori information in the detection of mammographic microcalcifications to improve their classification," Med. Phys. **30**(5), 823–831 (2003).
[30]R. M. Nishikawa, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computer-aided detection of clustered microcalcifications on digital mammograms," Med. Biol. Eng. Comput. **33**(2), 174–178 (1995).
[31]D. Betal, N. Roberts, and G. H. Whitehouse, "Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology," Br. J. Radiol. **70**(837), 903–917 (1997).
[32]J. C. Fu, S. K. Lee, S. T. C. Wong, J. Y. Yeh, A. H. Wang, and H. K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," Comput. Med. Imaging Graph. **29**(6), 419–429 (2005).
[33]J. Dengler, S. Behrens, and J. F. Desaga, "Segmentation of microcalcifications in mammograms," IEEE Trans. Med. Imaging **12**(4), 634–642 (1993).
[34]N. Ibrahim, H. Fujita, T. Hara, and T. Endo, "Automated detection of clustered microcalcifications on mammograms: Cad system application to mias database," Phys. Med. Biol. **42**(12), 2577–2589 (1997).
[35]M. G. Linguraru, K. Marias, R. English, and M. Brady, "A biologically inspired algorithm for microcalcification cluster detection," Med. Image Anal. **10**(6), 850–862 (2006).
[36]R. Nakayama, Y. Uchiyama, R. Watanabe, S. Katsuragawa, K. Namba, and K. Doi, "Computer-aided diagnosis scheme for histological classification of clustered microcalcifications on magnification mammograms," Med. Phys. **31**(4), 789–799 (2004).
[37]R. Nakayama, R. Watanabe, K. Namba, K. Takeda, K. Yamamoto, S. Katsuragawa, and K. Doi, "Computer-aided diagnosis scheme for identifying histological classification of clustered microcalcifications by use of follow-up magnification mammograms," Acad. Radiol. **13**(10), 1219–1228 (2006).
[38]F. Lefebvre, H. Benali, R. Gilles, E. Kahn, and R. Di Paola, "A fractal approach to the segmentation of microcalcifications in digital mammograms," Med. Phys. **22**(4), 381–390 (1995).
[39]H. Li, K. J. Liu, and S. C. Lo, "Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms," IEEE Trans. Med. Imaging **16**(6), 785–798 (1997).

[40]L. Bocchi, G. Coppini, J. Nori, and G. Valli, "Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks," Med. Eng. Phys. **26**(4), 303–312 (2004).
[41]H. D. Cheng, Y. M. Lui, and R. I. Freimanis, "A novel approach to microcalcification detection using fuzzy logic technique," IEEE Trans. Med. Imaging **17**(3), 442–450 (1998).
[42]B. Verma and J. Zakos, "A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques," IEEE Trans. Inf. Technol. Biomed. **5**(1), 46–54 (2001).
[43]N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms," Int. J. Pattern Recognit. Artif. Intell. **7**(6), 1357–1376 (1993).
[44]W. J. Veldkamp and N. Karssemeijer, "Normalization of local contrast in mammograms," IEEE Trans. Med. Imaging **19**(7), 731–738 (2000).
[45]K. J. McLoughlin, P. J. Bones, and N. Karssemeijer, "Noise equalization for detection of microcalcification clusters in direct digital mammogram images," IEEE Trans. Med. Imaging **23**(3), 313–320 (2004).
[46]R. M. Rangayyan, N. M. El-Faramawy, J. E. Desautels, and O. A. Alim, "Measures of acutance and shape for classification of breast tumors," IEEE Trans. Med. Imaging **16**(6), 799–810 (1997).
[47]L. M. Bruce and R. R. Adhami, "Classifying mammographic mass shapes using the wavelet transform modulus-maxima method," IEEE Trans. Med. Imaging **18**(12), 1170–1177 (1999).
[48]J. Kilday, J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computerized image analysis," IEEE Trans. Med. Imaging **12**(4), 664–669 (1993).
[49]M. L. Giger, C. J. Vyborny, and R. A. Schmidt, "Computerized characterization of mammographic masses: Analysis of spiculation," Cancer Lett. **77**(2–3), 201–211 (1994).
[50]Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, P. Lu, D. E. Wolverton, and R. A. Schmidt, "Analysis of spiculation in the computerized classification of mammographic masses," Med. Phys. **22**(10), 1569–1579 (1995).
[51]M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," IEEE Trans. Med. Imaging **17**(4), 510–517 (1998).
[52]N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," Med. Phys. **26**(8), 1642–1654 (1999).
[53]B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," Med. Phys. **25**(4), 516–526 (1998).
[54]A. J. Méndez, P. G. Tahoces, M. J. Lado, M. Souto, and J. J. Vidal, "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms," Med. Phys. **25**(6), 957–964 (1998).
[55]S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote, and S. Grundfest-Broniatowski, "Quantitative classification of breast tumors in digitized mammograms," Med. Phys. **23**(8), 1337–1345 (1996).
[56]D. Guliato, R. M. Rangayyan, W. A. Carnielli, J. A. Zuffo, and J. E. Leo Desautels, "Segmentation of breast tumors in mammograms using fuzzy sets," J. Electron. Imaging **12**, 369–378 (2003).
[57]G. M. te Brake and N. Karssemeijer, "Segmentation of suspicious densities in digital mammograms," Med. Phys. **28**(2), 259–266 (2001).
[58]B. Sahiner, N. Petrick, H. P. Chan, L. M. Hadjiiski, C. Paramagul, M. A. Helvie, and M. N. Gurcan, "Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization," IEEE Trans. Med. Imaging **20**(12), 1275–1284 (2001).
[59]B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," Med. Phys. **28**(7), 1455–1465 (2001).
[60]J. Shi, B. Sahiner, H.-P. Chan, J. Ge, L. Hadjiiski, M. A. Helvie, A. Nees, Y.-T. Wu, J. Wei, C. Zhou, Y. Zhang, and J. Cui, "Characterization of mammographic masses based on level set segmentation with new image features and patient information," Med. Phys. **35**(1), 280–290 (2008).
[61]Y. Yuan, M. L. Giger, H. Li, K. Suzuki, and C. Sennett, "A dual-stage method for lesion segmentation on digital mammograms," Med. Phys. **34**(11), 4180–4193 (2007).
[62]S. Timp and N. Karssemeijer, "A new 2d segmentation method based on dynamic programming applied to computer aided detection in mammography," Med. Phys. **31**(5), 958–971 (2004).
[63]C. Varela, S. Timp, and N. Karssemeijer, "Use of border information in the classification of mammographic masses," Phys. Med. Biol. **51**(2), 425–441 (2006).
[64]A. Rojas Domínguez and A. K. Nandi, "Improved dynamic-programming-based algorithms for segmentation of masses in mammograms," Med. Phys. **34**(11), 4256–4269 (2007).
[65]A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," Artif. Intell. Med. **34**(2), 141–150 (2005).
[66]M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters," Med. Phys. **31**(2), 314–326 (2004).
[67]W. J. Veldkamp, N. Karssemeijer, J. D. Otten, and J. H. Hendriks, "Automated classification of clustered microcalcifications into malignant and benign types," Med. Phys. **27**(11), 2600–2608 (2000).
[68]L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," IEEE Trans. Med. Imaging **24**(3), 371–380 (2005).
[69]American College of Radiology, Breast Imaging Reporting and Data System BI-RADS Atlas, 2006.
[70]H. Soltanian-Zadeh, F. Rafiee-Rad, and S. Pourabdollah-Nejad, "Comparison of multiwavelet, wavelet, haralick and shape features for microcalcification classification in mammograms," Pattern Recogn. **37**, 1973–1986 (2004).
[71]L. Bocchi and J. Nori, "Shape analysis of microcalcifications using radon transform," Med. Eng. Phys. **29**(6), 691–698 (2007).
[72]R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst. Man Cybern. **3**(6), 610–621 (1973).
[73]D. L. Thiele, C. Kimme-Smith, T. D. Johnson, M. McCombs, and L. W. Bassett, "Using tissue texture surrounding calcification clusters to predict benign vs malignant outcomes," Med. Phys. **23**(4), 549–555 (1996).
[74]H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, K. L. Lam, D. D. Adler, and M. M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network," Phys. Med. Biol. **42**(3), 549–567 (1997).
[75]H. P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," Med. Phys. **25**(10), 2007–2019 (1998).
[76]A. Laine and J. Fan, "Texture classification by wavelet packet signatures," IEEE Trans. Pattern Anal. Mach. Intell. **15**(11), 1186–1191 (1993).
[77]L. Tabar, P. B. Dean, and T. Tot, *Teaching Atlas of Mammography*, 3rd ed. (Thieme, New York, 2001).
[78]D. M. Catarious, A. H. Baydush, and C. E. Floyd, "Incorporation of an iterative, linear segmentation routine into a mammographic mass cad system," Med. Phys. **31**(6), 1512–1520 (2004).
[79]B. Zheng, A. Lu, L. A. Hardesty, J. H. Sumkin, C. M. Hakim, M. A. Ganott, and D. Gur, "A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment," Med. Phys. **33**(1), 111–117 (2006).
[80]Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," Acad. Radiol. **5**(3), 155–168 (1998).
[81]Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, and C. E. Metz, "Computerized classification of benign and malignant masses on digitized mammograms: A study of robustness," Acad. Radiol. **7**(12), 1077–1084 (2000).
[82]B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Design of a high-sensitivity classifier based on a genetic algorithm: Application to computer-aided diagnosis," Phys. Med. Biol. **43**(10), 2853–2871 (1998).
[83]L. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, and M. Helvie, "Classification of malignant and benign masses based on hybrid art2lda approach," IEEE Trans. Med. Imaging **18**(12), 1178–1187 (1999).
[84]N. R. Mudigonda, R. M. Rangayyan, and J. E. Desautels, "Gradient and texture analysis for the classification of mammographic masses," IEEE Trans. Med. Imaging **19**(10), 1032–1043 (2000).
[85]G. M. te Brake, N. Karssemeijer, and J. H. Hendriks, "An automatic method to discriminate malignant masses from normal tissue in digital mammograms," Phys. Med. Biol. **45**(10), 2843–2857 (2000).

[86]W. K. Lim and M. J. Er, "Classification of mammographic masses using generalized dynamic fuzzy neural networks," Med. Phys. **31**(5), 1288–1295 (2004).
[87]S. Gupta and M. K. Markey, "Correspondence in texture features between two mammographic views," Med. Phys. **32**(6), 1598–1606 (2005).
[88]M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, D. Cavouras, and S. Theodoridis, "Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers," Artif. Intell. Med. **37**(2), 145–162 (2006).
[89]M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, in *Proceedings of the Fifth International Workshop on Digital Mammography*, edited by M. J. Yaffe (Medical Physics Publishing, 2001), pp. 212–218.
[90]R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* (Wiley-Interscience, New York, 2000).
[91]R. Nakayama, R. Watanabe, K. Namba, K. Takeda, K. Yamamoto, S. Katsuragawa, and K. Doi, "An improved computer-aided diagnosis scheme using the nearest neighbor criterion for determining histological classification of clustered microcalcifications," Methods Inf. Med. **46**(6), 716–722 (2007).
[92]Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer," Radiology **187**(1), 81–87 (1993).
[93]J. Y. Lo, J. A. Baker, P. J. Kornguth, J. D. Iglehart, and C. E. Floyd, "Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features," Radiology **203**(1), 159–163 (1997).
[94]C. Markopoulos, E. Kouskos, K. Koufopoulos, V. Kyriakou, and J. Gogas, "Use of artificial neural networks (computer analysis) in the diagnosis of microcalcifications on mammography," Eur. J. Radiol. **39**(1), 60–65 (2001).
[95]J. L. Jesneck, J. Y. Lo, and J. A. Baker, "Breast mass lesions: Computer-aided diagnosis models with mammographic and sonographic descriptors," Radiology **244**(2), 390–398 (2007).
[96]S. Gupta, P. F. Chyn, and M. K. Markey, "Breast cancer cadx based on bi-rads descriptors from two mammographic views," Med. Phys. **33**(6), 1810–1817 (2006).
[97]A. A. J. Roelofs, N. Karssemeijer, N. Wedekind, C. Beck, S. van Woudenberg, P. R. Snoeren, J. H. C. L. Hendriks, M. Rosselli del Turco, N. Bjurstam, H. Junkermann, D. Beijerinck, B. Séradour, and C. J. G. Evertsz, "Importance of comparison of current and prior mammograms in breast cancer screening," Radiology **242**(1), 70–77 (2007).
[98]C. Varela, N. Karssemeijer, J. H. C. L. Hendriks, and R. Holland, "Use of prior mammograms in the classification of benign and malignant masses," Eur. J. Radiol. **56**(2), 248–255 (2005).
[99]S. Timp and N. Karssemeijer, "Interval change analysis to improve computer aided detection in mammography," Med. Image Anal. **10**(1), 82–95 (2006).
[100]L. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. Gurcan, "Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses," Med. Phys. **28**(11), 2309–2317 (2001).
[101]S. Timp, C. Varela, and N. Karssemeijer, "Temporal change analysis for characterization of mass lesions in mammography," IEEE Trans. Med. Imaging **26**(7), 945–953 (2007).
[102]S. Sanjay-Gopal, H. P. Chan, T. Wilson, M. Helvie, N. Petrick, and B. Sahiner, "A regional registration technique for automated interval change analysis of breast lesions on mammograms," Med. Phys. **26**(12), 2669–2679 (1999).
[103]P. Filev, L. Hadjiiski, B. Sahiner, H.-P. Chan, and M. A. Helvie, "Comparison of similarity measures for the task of template matching of masses on serial mammograms," Med. Phys. **32**(2), 515–529 (2005).
[104]S. Timp, S. van Engeland, and N. Karssemeijer, "A regional registration method to find corresponding mass lesions in temporal mammogram pairs," Med. Phys. **32**(8), 2629–2638 (2005).
[105]P. R. Snoeren and N. Karssemeijer, "Gray scale registration of mammograms using a model of image acquisition," Lecture Notes in Computer Science **2732**, 401–412 (2003).
[106]P. R. Snoeren and N. Karssemeijer, "Gray-scale and geometric registration of full-field digital and film-screen mammograms," Med. Image Anal. **11**(2), 146–156 (2007).
[107]S. van Engeland, P. Snoeren, J. Hendriks, and N. Karssemeijer, "A comparison of methods for mammogram registration," IEEE Trans. Med. Imaging **22**(11), 1436–1444 (2003).
[108]H. P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. Sanjay-Gopal, "Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: An roc study," Radiology **212**(3), 817–827 (1999).
[109]Z. Huo, M. I. Giger, C. J. Vyborny, and C. E. Metz, "Breast cancer: Effectiveness of computer-aided diagnosis observer study with independent database of mammograms," Radiology **224**(2), 560–568 (2002).
[110]Z. Huo, M. L. Giger, and C. J. Vyborny, "Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis," IEEE Trans. Med. Imaging **20**(12), 1285–1292 (2001).
[111]K. Drukker, K. Horsch, and M. L. Giger, "Multimodality computerized diagnosis of breast lesions using mammography and sonography," Acad. Radiol. **12**(8), 970–979 (2005).
[112]K. Horsch, M. L. Giger, C. J. Vyborny, L. Lan, E. B. Mendelson, and R. E. Hendrick, "Classification of breast lesions with multimodality computer-aided diagnosis: Observer study results on an independent clinical data set," Radiology **240**(2), 357–368 (2006).
[113]S. van Engeland, S. Timp, and N. Karssemeijer, "Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views," Med. Phys. **33**(9), 3203–3212 (2006).
[114]B. Liu, C. E. Metz, and Y. Jiang, "Effect of correlation on combining diagnostic information from two images of the same patient," Med. Phys. **32**(11), 3329–3338 (2005).
[115]J. A. Baker, P. J. Kornguth, J. Y. Lo, M. E. Williford, and C. E. Floyd, "Breast cancer: Prediction with artificial neural network based on bi-rads standardized lexicon," Radiology **196**(3), 817–822 (1995).
[116]J. A. Baker, P. J. Kornguth, J. Y. Lo, and C. E. Floyd, "Artificial neural network: Improving the quality of breast biopsy recommendations," Radiology **198**(1), 131–135 (1996).
[117]J. Y. Lo, J. A. Baker, P. J. Kornguth, and C. E. Floyd, "Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks," Acad. Radiol. **6**(1), 10–15 (1999).
[118]M. K. Markey, J. Y. Lo, R. Vargas-Voracek, G. D. Tourassi, and C. E. Floyd, "Perceptron error surface analysis: A case study in breast cancer diagnosis," Comput. Biol. Med. **32**(2), 99–109 (2002).
[119]J. Y. Lo, M. K. Markey, J. A. Baker, and C. F. Floyd, "Cross-institutional evaluation of bi-rads predictive model for mammographic diagnosis of breast cancer," AJR, Am. J. Roentgenol. **178**(2), 457–463 (2002).
[120]M. K. Markey, G. D. Tourassi, M. Margolis, and D. M. DeLong, "Impact of missing data in evaluating artificial neural networks trained on complete data," Comput. Biol. Med. **36**(5), 516–525 (2006).
[121]C. E. Floyd, J. Y. Lo, and G. D. Tourassi, "Case-based reasoning computer algorithm that uses mammographic findings for breast biopsy decisions," AJR, Am. J. Roentgenol. **175**(5), 1347–1352 (2000).
[122]A. O. Bilska-Wolak and C. E. Floyd, "Development and evaluation of a case-based reasoning classifier for prediction of breast biopsy outcome with bi-rads lexicon," Med. Phys. **29**(9), 2090–2100 (2002).
[123]A. O. Bilska-Wolak, C. E. Floyd, J. Y. Lo, and J. A. Baker, "Computer aid for decision to biopsy breast masses on mammography Validation on new cases," Acad. Radiol. **12**(6), 671–680 (2005).
[124]G. D. Tourassi, M. K. Markey, J. Y. Lo, and C. E. Floyd, "A neural network approach to breast cancer diagnosis as a constraint satisfaction problem," Med. Phys. **28**(5), 804–811 (2001).
[125]E. A. Fischer, J. Y. Lo, and M. K. Markey, "Bayesian networks of BI-RADS™ mark descriptors for breast lesion classification," in *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society* (IEEE Computer Society, Los Alamitos, 2004), Vol. 4, pp. 3031–3034.
[126]M. Elter, R. Schulz-Wendtland, and T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two cad approaches that both emphasize an intelligible decision process," Med. Phys. **34**(11), 4164–4172 (2007).
[127]E. S. Burnside, J. E. Ochsner, K. J. Fowler, J. P. Fine, L. R. Salkowski, D. L. Rubin, and G. A. Sisney, "Use of microcalcification descriptors in bi-rads 4th edition to stratify risk of malignancy," Radiology **242**(2), 388–395 (2007).

[128]J. A. Baker, P. J. Kornguth, and C. E. Floyd, "Breast imaging reporting and data system standardized mammography lexicon: Observer variability in lesion description," AJR, Am. J. Roentgenol. **166**(4), 773–778 (1996).
[129]C. E. Metz, "Basic principles of roc analysis," Semin Nucl. Med. **8**(4), 283–298 (1978).
[130]C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum likelihood estimation of receiver operating characteristic (roc) curves from continuously-distributed data," Stat. Med. **17**(9), 1033–1053 (1998).
[131]E. E. Metz and X. Pan, ""Proper" binormal roc curves: Theory and maximum-likelihood estimation," J. Math. Psychol. **43**(1), 1–33 (1999).
[132]D. C. Edwards, L. Lan, C. E. Metz, M. L. Giger, and R. M. Nishikawa, "Estimating three-class ideal observer decision variables for computerized detection and classification of mammographic mass lesions," Med. Phys. **31**(1), 81–90 (2004).
[133]D. Guliato, R. M. Rangayyan, J. D. Carvalho, and S. A. Santiago, "Polygonal modeling of contours of breast tumors with the preservation of spicules," IEEE Trans. Biomed. Eng. **55**(1), 14–20 (2008).
[134]P. Delogu, M. E. Fantacci, P. Kasae, and A. Retico, "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier," Comput. Biol. Med. **37**(10), 1479–1491 (2007).
[135]A. Karahaliou, S. Skiadopoulos, I. Boniatis, P. Sakellaropoulos, E. Likaki, G. Panayiotakis, and L. Costaridou, "Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis," Br. J. Radiol. **80**(956), 648–656 (2007).
[136]I. Leichter, R. Lederman, S. S. Buchbinder, P. Bamberger, B. Novak, and S. Fields, "Computerized evaluation of mammographic lesions: What diagnostic role does the shape of the individual microcalcifications play compared with the geometry of the cluster?," AJR, Am. J. Roentgenol. **182**(3), 705–712 (2004).
[137]S. S. Buchbinder, I. S. Leichter, P. N. Bamberger, B. Novak, R. Lederman, S. Fields, and D. J. Behar, "Analysis of clustered microcalcifications by using a single numeric classifier extracted from mammographic digital images," Acad. Radiol. **5**(11), 779–784 (1998).
[138]D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "Roc rating analysis: generalization to the population of readers and cases with the jackknife method," Invest. Radiol. **27**, 723–731 (1992).
[139]Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," Acad. Radiol. **6**(1), 22–33 (1999).
[140]Y. Jiang, R. M. Nishikawa, R. A. Schmidt, A. Y. Toledano, and K. Doi, "Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications," Radiology **220**(3), 787–794 (2001).
[141]Y. Jiang, R. M. Nishikawa, R. A. Schmidt, and C. E. Metz, "Comparison of independent double readings and computer-aided diagnosis (cad) for the diagnosis of breast calcifications," Acad. Radiol. **13**(1), 84–94 (2006).
[142]R. S. Rana, Y. Jiang, R. A. Schmidt, R. M. Nishikawa, and B. Liu, "Independent evaluation of computer classification of malignant and benign calcifications in full-field digital mammograms," Acad. Radiol. **14**(3), 363–370 (2007).
[143]I. Leichter, S. Fields, R. Nirel, P. Bamberger, B. Novak, R. Lederman, and S. Buchbinder, "Improved mammographic interpretation of masses using computer-aided diagnosis," Eur. Radiol. **10**(2), 377–383 (2000).
[144]L. Hadjiiski, H.-P. Chan, B. Sahiner, M. A. Helvie, M. A. Roubidoux, C. Blane, C. Paramagul, N. Petrick, J. Bailey, K. Klein, M. Foster, S. Patterson, D. Adler, A. Nees, and J. Shen, "Improvement in radiologists' characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis: An roc study," Radiology **233**(1), 255–265 (2004).
[145]L. Hadjiiski, B. Sahiner, M. A. Helvie, H.-P. Chan, M. A. Roubidoux, C. Paramagul, C. Blane, N. Petrick, J. Bailey, K. Klein, M. Foster, S. K. Patterson, D. Adler, A. V. Nees, and J. Shen, "Breast masses: Computer-aided diagnosis with serial mammograms," Radiology **240**(2), 343–356 (2006).
[146]J. Wei, L. M. Hadjiiski, B. Sahiner, H.-P. Chan, J. Ge, M. A. Roubidoux, M. A. Helvie, C. Zhou, Y.-T. Wu, C. Paramagul, and Y. Zhang, "Computer-aided detection systems for breast masses: Comparison of performances on full-field digital mammograms and digitized screen-film mammograms," Acad. Radiol. **14**(6), 659–669 (2007).