Pattern Recognition 43 (2010) 299–317

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Automated breast cancer detection and classification using ultrasound images: A survey

H.D. Cheng[a,*], Juan Shan[a], Wen Ju[a], Yanhui Guo[a], Ling Zhang[b]

[a]Department of Computer Science, Utah State University, Logan, UT 84322, USA
[b]School of Mathematics and System Science, Shandong University, China

ARTICLE INFO

Article history:
Received 9 September 2008
Received in revised form 16 April 2009
Accepted 14 May 2009

Keywords:
CAD (computer-aided diagnosis)
Automated breast cancer detection and classification
Ultrasound (US) imaging
Feature extraction and selection
Classifiers

ABSTRACT

Breast cancer is the second leading cause of death for women all over the world. Since the cause of the disease remains unknown, early detection and diagnosis is the key for breast cancer control, and it can increase the success of treatment, save lives and reduce cost. Ultrasound imaging is one of the most frequently used diagnosis tools to detect and classify abnormalities of the breast. In order to eliminate the operator dependency and improve the diagnostic accuracy, computer-aided diagnosis (CAD) system is a valuable and beneficial means for breast cancer detection and classification. Generally, a CAD system consists of four stages: preprocessing, segmentation, feature extraction and selection, and classification. In this paper, the approaches used in these stages are summarized and their advantages and disadvantages are discussed. The performance evaluation of CAD system is investigated as well.

1. Introduction

Breast cancer is the second leading cause of death for women all over the world and more than 8% women will suffer this disease during their lifetime. In 2008, there were reported approximately 182,460 newly diagnosed cases and 40,480 deaths in the United States [4]. Since the causes of breast cancer still remain unknown, early detection is the key to reduce the death rate (40% or more) [2]. The earlier the cancers are detected, the better treatment can be provided. However, early detection requires an accurate and reliable diagnosis which should also be able to distinguish benign and malignant tumors. A good detection approach should produce both low false positive (FP) rate and false negative (FN) rate.

Previously, the most effective modality for detecting and diagnosing is mammography [1,2]. However, there are limitations of mammography in breast cancer detection. Many unnecessary (65–85%) biopsy operations are due to the low specificity of mammography [5]. The unnecessary biopsies not only increase the cost, but also make the patients suffer from emotional pressure. Mammography can hardly detect breast cancer in adolescent women with dense breasts. In addition, the ionizing radiation of mammography can increase the health risk for the patients and radiologists.

Currently, an important alternative to mammography is ultrasound (US) imaging, and it shows an increasing interest in the use of ultrasound images for breast cancer detection [6–8]. Statistics showed that more than one out of every four researches is using ultrasound images, and the proportion increases more and more quickly [3]. Studies have demonstrated that using US images can discriminate benign and malignant masses with a high accuracy [9,10]. Use of ultrasound can increase overall cancer detection by 17% [11] and reduce the number of unnecessary biopsies by 40% which can save as much as $1 billion per year in the United Sates [12]. Breast ultrasound (BUS) imaging is superior to the mammography in the facts: (1) Since having no radiation, ultrasound examination is more convenient and safer than mammography for patients and radiologists in daily clinical practice [11,13,16]. It is also cheaper and faster than mammography. Thus, ultrasound is especially fit for the low-resource countries in different continents [153]. (2) Ultrasound is more sensitive than mammography for detecting abnormalities in dense breasts, hence, it is more valuable for women younger than 35 years of age [11,14]. (3) There is a high rate of false positives in mammography which causes a lot of unnecessary biopsies [10]. In contrast, the accuracy rate of BUS imaging in the diagnosis of simple cysts can reach 96–100% [9]. US imaging becomes one of the most important diagnostic tools for breast cancer detection. However, sonography is much more operator-dependent than mammography, reading ultrasound image requires well-trained and experienced radiologists. Even well-trained experts may have a high inter-observer variation rate, therefore, computer-aided diagnosis (CAD) is needed to help radiologists in breast cancer detection and classification [13]. Recently, several CAD approaches have been studied to minimize the effect of the operator-dependent nature inherent in US imaging [15],

* Corresponding author.
E-mail address: hengda.cheng@usu.edu (H.D. Cheng).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.05.012

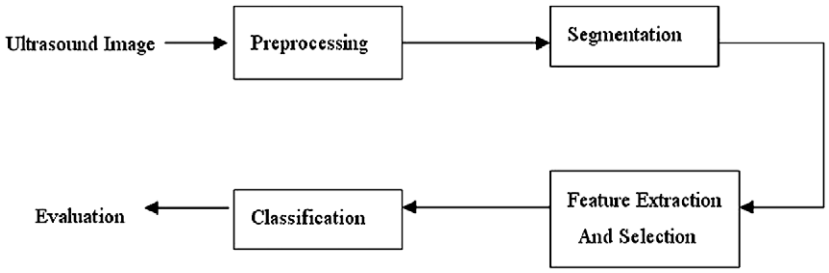

Fig. 1. CAD system for breast cancer detection and classification.

and to increase the diagnostic sensitivity and specificity [13,16]. As much as 65–90% of the biopsies turned out to be benign, therefore, a crucial goal of breast cancer CAD systems is to distinguish benign and malignant lesions to reduce FPs. Many techniques such as linear discriminant analysis (LDA), support vector machine (SVM) and artificial neural network (ANN) [5,10,17,18,20] have been studied for mass detection and classification. Most of the CAD systems need a large number of samples to construct the models or rules, but [22] proposed a novel diagnosis system requiring very few samples.

This survey focuses on summarizing the approaches for breast cancer detection and classification utilizing BUS images. Generally, the ultrasound CAD systems for breast cancer detection involve four stages as shown in Fig. 1.

(1) *Image preprocessing*: The major limitations of BUS imaging are the low contrast and interference with speckle [3]. The task of image preprocessing is to enhance the image and to reduce speckle without destroying the important features of BUS images for diagnosis.
(2) *Image segmentation*: Image segmentation divides the image into non-overlapping regions, and it will separate the objects from the background. The regions of interest (ROIs) will be allocated for feature extraction.
(3) *Feature extraction and selection*: This step is to find a feature set of breast cancer lesions that can accurately distinguish lesion/non-lesion or benign/malignant. The feature space could be very large and complex, so extracting and selecting the most effective features is very important. Most of the reported effective features are listed in Table 4.
(4) *Classification*: Based on the selected features, the suspicious regions will be classified as lesion/non-lesion or benign/malignant by various classification methods. The commonly used classifiers are discussed in Section 5.

Some CAD systems do not have image preprocessing and image segmentation components. In such a framework, only some texture features obtained directly from the images or ROIs are used as inputs of classifiers [13,16,20,22]. The advantage of such CAD system is its simple structure and fast processing speed, and disadvantage is that the features extracted directly from ROIs may not provide robust and accurate performance.

At last, we need to measure the performance of CAD systems. There is no a benchmark database of US images for comparing the performance of the algorithms/CAD systems, and it makes the evaluation of different CAD systems very difficult or even impossible. This indicates the necessity to build a benchmark BUS image base accessible to the public.

2. Preprocessing

The preprocessing of BUS images consists of speckle reduction and image enhancement. Speckle is a form of multiplicative noise generated by a number of scatterers with random phase within the resolution cell of ultrasound beam [33,34]. Ref. [29] has demonstrated that the *k*-distribution is a good model for the amplitude distribution of the received signal. A more generalized statistical model, the homodyned *k*-distribution, has been analyzed in [30]. It combined the features of the *k*-distribution and Rice distribution to better account for the statistics of the signal. To detect speckles, the parameters for the speckles should be estimated first. The speckle parameters of the *k*-distribution model can be estimated based on the moments [31]. An iterative method using the statistics of ultrasound signal is proposed to find the parameters of the homodyned *k*-distribution model [32]. Speckle makes the visual observation and interpretation difficult. Therefore, removing speckle without destroying important features for diagnosis is critical. Some speckle reduction techniques only work well on additive noise, and logarithmic compression is often employed to convert multiplicative noise into additive noise [33]. Image enhancement is used to improve the quality of low contrast images. We will review speckle reduction and image enhancement separately, however, many techniques can achieve both goals at the same time.

2.1. Speckle reduction

Speckle reduction techniques are classified into three groups: (1) filtering techniques [34–59]; (2) wavelet domain techniques [60–79]; and (3) compounding approaches [80–83].

2.1.1. Filtering techniques

Most filters are traditional techniques in spatial domain and can be categorized as linear and nonlinear filters.

2.1.1.1. Linear filters.

2.1.1.1.1. Mean filter. The mean filter [41,42] replaces each pixel by the average value of the intensities in its neighborhood. It can locally reduce the variance and is easy to implement. It has the effect of smoothing and blurring the image, and is optimal for additive Gaussian noise in the sense of mean square error. Speckled image is a multiplicative model with non-Gaussian noise, and therefore, the simple mean filter is not effective in this case.

2.1.1.1.2. Adaptive mean filter. In order to alleviate the blurring effect, the adaptive mean filters [35–40] have been proposed to achieve a balance between straightforward averaging (in homogeneous regions) and all-pass filtering (where edges exist). They adapt to the properties of the image locally and selectively remove speckles from different parts of the image. They use local image statistics such as mean, variance and spatial correlation to effectively detect and preserve edges and features. The speckle noise is removed by replacing it with a local mean value. The adaptive mean filters

outperform mean filters, and generally reduce speckles while preserving the edges.

The Lee [36], Kuan [35] and Frost [37] filters are well-known examples of adaptive mean filters. They are based on the multiplicative speckle model which can be represented as

$$I(t) = R(t)u(t)$$

where $t=(x, y)$ is the coordinates of the current pixel, $R(t)$ denotes the intensity of the ideal image without speckle, $I(t)$ is the observed image intensity and $u(t)$ is the speckle with mean $\bar{u}$ and variance σ_u^2.

The Lee and Kuan filters have the general form:

$$\hat{R}(t) = \bar{I}(t) + [I(t) - \bar{I}(t)] \times W(t)$$

where $\hat{R}(t)$ is the output of the filter, $\bar{I}(t)$ and $\sigma_I^2(t)$ are the local mean and variance of the intensities within the filter window, respectively. $W(t)$ denotes the coefficient function of the adaptive filter. The Lee filter has a coefficient function defined as [38]

$$W(t) = 1 - \frac{C_u^2}{C_I^2(t)} \quad \text{where } C_u^2 = \frac{\sigma_z^2}{\bar{z}^2}, \quad C_I^2(t) = \frac{\sigma_I^2(t)}{\bar{I}^2(t)}$$

σ_z^2 and $\bar{z}$ are the intensity variance and mean over a homogeneous area of the image, respectively. Thus C_u^2 can be considered as a constant for a given image.

The coefficient function of Kuan filter is defined as [38]

$$W(t) = \frac{1 - C_u^2/C_I^2(t)}{1 + C_u^2}$$

From the formulations of the Lee and Kuan filters, the difference between them is only a constant factor $1 + C_u^2$. So we could discuss them together. In the homogeneous regions, $C_I^2(t) \to C_u^2$. Thus the value of $W(t)$ approaches 0, which makes the filters act like a mean filter. On the other hand, in the areas where edges exist, $C_I^2(t) \to \infty$, and the value of $W(t)$ approaches 1, which tends to preserve the originally observed image and makes the filters behave like an all-pass filter.

The Frost filter can be represented as

$$\hat{R}(x, y) = \sum_i \sum_j m(x+i, y+j) \times I(x+i, y+j)$$

where i and j are the indices of the filter window and m is the weighting function [38]:

$$m(x+i, y+j) = K_0 \exp[-KC_I^2(t)\sqrt{i^2+j^2}] \quad \text{where } t=(x, y)$$

where K_0 is a normalizing constant, and K is a damping factor.

Frost filter has similar attributes as Lee and Kuan filters. The damping factor K is chosen such that when in homogeneous regions, $KC_I^2(t)$ approaches 0. Thus the value of $m(x+i, y+j)$ approaches 1, which makes the filter act like a mean filter; in the areas where edges exist, $KC_I^2(t)$ becomes so large that the value of $m(x+i, y+j)$ approaches 0 for the pixels surrounding (x, y), and remains 1 for the pixel (x, y). This makes the filter behave like an all-pass filter preserving the originally observed image.

The classical Lee, Kuan and Frost filters are only reliable in a bounded field. Ref. [38] enhanced the Lee and Frost filters by dividing the image into three classes according to the local coefficient of variation $C_I(t)$. If $C_I(t)$ is below a lower threshold, pure averaging is used. All-pass filter is performed when $C_I(t)$ is above a higher threshold. When $C_I(t)$ exists in between the two thresholds, standard Lee and Frost filters are applied. The enhanced filters adequately average the homogeneous areas and preserve the edges better than the standard filters. Ref. [45] proposed a directional adaptive mean filter based on 2D texture homogeneity histogram to suppress speckles in ultrasound images.

2.1.1.2. Nonlinear filters.

2.1.1.2.1. Order-statistic filter. Order-statistic filters are particularly effective in reducing noise whose probability density function has significant tail. The median filter [34,41,42] is a special case of order-statistic filters. It preserves the edge sharpness and produces less blurring than mean filter. Specially, it is effective when image is affected by impulsive noise. Several researchers have experimented with adaptive median filters which outperform the median filters [43,46]. An adaptive weighted median filter (AWMF) was developed to achieve maximum speckle reduction in uniform areas and to preserve the edges and features [47]. The weighted median of a sequence $\{X_i\}$ is defined as the pure median of the extended sequence which is generated by repeating each term X_i by w_i times. Here, $\{w_i\}$ are the corresponding weight coefficients. The weight coefficients are adjusted according to the local statistics as

$$w(i, j) = [w(K+1, K+1) - bd\sigma^2/\bar{u}]$$

where b is a scaling constant, $\bar{u}$ and σ^2 are the local mean and variance of the $(2K+1)*(2K+1)$ window, d is the distance of the point (i, j) from the center of the window at $(K+1, K+1)$, and $[x]$ denotes the nearest integer to x if x is positive, or zero if x is negative. However, this algorithm uses an operator which can cause difficulties in enhancing image features such as line segments. To overcome this drawback, [48] applied a bank of oriented one-dimensional median filters and retained at each point the largest value among all the filter bank outputs. The directional median filter suppresses speckle noise while retaining the structure of the image, particularly, the thin bright streaks.

2.1.1.2.2. MAP filter. Maximum a posteriori (MAP) filter [41,49,50] estimates an unobserved signal x by maximizing Bayes theorem:

$$f(x|z) = \frac{f(z|x)f(x)}{f(z)}$$

where $f(x|z)$ is the a posteriori probability density function, $f(x)$ is the denoised original signal model, $f(z|x)$ is the maximum likelihood term and $f(z)$ is the model of the observed data. To utilize it, the priori knowledge of the probability density function (PDF) of the image is needed. The PDF is assumed to be Gaussian distributed in [49]. Ref. [51] modified the MAP filter in [49] by assuming a gamma and symmetric Beta distribution.

Comparisons of standard de-speckle filters (Fig. 2) with the adaptive MAP filter for ultrasound images are presented in [41]. MAP Gauss denotes the MAP filter with Gaussian distribution assigned to the original image, and MAP Pearlman Gauss denotes the MAP Gauss filter using the adaptive windowing proposed in [154]. MDb1 is the filter using Daubechies wavelets with Db1 basis. Contrast to speckle ratio (CRS) is used to evaluate the performance and the results show that the MAP Pearlman Gauss is the best among the filters being compared.

2.1.1.2.3. Nonlinear diffusion. Nonlinear diffusion is actually an adaptive filter, where the direction and strength of the diffusion are controlled by an edge detection function. It can remove speckles and enhance edges at the same time. It removes speckles by modifying the image via solving a partial differential equation (PDE).

Ref. [52] proposed the nonlinear PDE for smoothing image in a continuous domain:

$$\begin{cases} \dfrac{\partial I}{\partial t} = div[c(|\nabla I|) \cdot \nabla I] \\ I(t=0) = I_0 \end{cases}$$

where ∇ is the gradient operator, div is the divergence operator, $||$ denotes the magnitude, $c(x)$ is the diffusion coefficient and I_0 is the

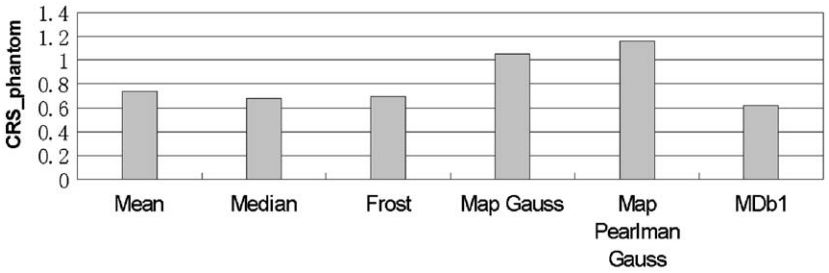

Fig. 2. CRS values of phantom image using various filters [41].

original image. The diffusion coefficient function $c(x)$ should monotonically decrease, the diffusion decreases as the gradient strength increases, and the diffusion is stopped across edges.

Anisotropic diffusion (AD) performs well with additive Gaussian noise. However, edge estimation using gradient operator makes it difficult to handle multiplicative noisy image. In order to eliminate such disadvantage, Speckle Reducing Anisotropic Diffusion (SRAD) is proposed particularly for envelope US images without logarithmic compression [53]. In SRAD, the instantaneous coefficient of variation servers as the edge detector in speckled images. The function exhibits high values at edges and produces low values in homogeneous regions. Thus, it ensures the mean-preserving behavior in the homogeneous regions, and edge-preserving and edge-enhancing at the edges. Ref. [44] extended 2D SRAD to 3D SRAD to process 3D ultrasound images more efficiently. Nonlinear coherence enhancement diffusion (NCD) is another method for handling speckle noise [33]. Unlike SRAD, NCD works with the US images after logarithmic compression. It combines three different models. According to speckle extent and image anisotropy, the NCD model changes progressively from isotropic diffusion through anisotropic coherent diffusion to, finally, mean curvature motion.

Diffusion stick method for speckle suppression is proposed in [54]. It divides the traditional rectangular filter kernel into a set of asymmetric sticks of variable orientations. The weighted sum of averages along each stick is used to produce the filtered images. Ref. [55] developed a multi-resolution median-anisotropic diffusion interactive method. It used a two resolution level process to convert speckle to quasi-impulsive noise. Then the low-resolution images are processed by the median-anisotropic diffusion interactive algorithm. The computational cost is lower than that of the conventional AD schemes. In [56] a hybrid method is designed based on median filtering, improved AD filtering and isotropic diffusion filtering. The gradient matrix is analyzed, and the thresholds are chosen by experiments. The hybrid method combines the three filtering methods for three different grayscale gradient ranges, respectively.

The advantage of nonlinear diffusion is that the speckle reduction is carried out directionally by the edge function and the edges are enhanced. The disadvantage is that it relies on the diffusion flux to iteratively eliminate the small variations caused by noise, and to preserve the large variations caused by edges. For the multiplicative noisy image, however, the general signal/noise relationship no longer exists, since the variations caused by noise may be larger than those caused by signal. Hence, it is not suitable for this case.

2.1.1.2.4. Other nonlinear filters. Geometric filter (GF) is a nonlinear, iterative algorithm which changes the pixel values in a neighborhood based on their relative values [34]. The geometric concepts (convex, 8-Hull) and the algorithm were described in [57]. GF effectively removes speckle noise while preserves important details. An adaptive algorithm called aggressive region growing filtering (ARGF) was proposed in [58]. It selected a filtering region in the image using an appropriately estimated homogeneity threshold for region growing. Homogeneous regions are smoothed by applying an arithmetic mean filter and edge pixels are filtered using a nonlinear median filter. A directional line-matched filtering scheme was proposed in [59]. It could detect and enhance the image features while suppressing speckle noise.

The filter techniques are simple and fast, however, they have certain limitations as they are sensitive to the size and shape of the filter window. If the window size is too large, over-smoothing will occur. If the window size is too small, the smoothing capability of the filter will decrease and the speckle noise cannot be reduced effectively. Considering window shape, the square window, which is mostly adopted, will lead to corner rounding of rectangular features. Some despeckling filters require thresholds which have to be estimated empirically.

2.1.2. Wavelet domain techniques

The discrete wavelet transform (DWT) translates the image into an approximation sub-band consisting of the scale coefficients and a set of detail sub-bands at different orientations and resolution scales composed of the wavelet coefficients [72]. DWT provides an appropriate basis for separating the noise from an image. As the wavelet transform is good at energy compaction, the small coefficients more likely represent noise, and large coefficients represent important image features. The coefficients representing features tend to persist across the scales and form spatially connected clusters within each sub-band. These properties make DWT attractive for denoising. A number of wavelet-based despeckling techniques have been developed. The general procedure is: (1) calculate the discrete wavelet transform; (2) remove noise by changing the wavelet coefficients; and (3) apply the inverse wavelet transform (IDWT) to construct the despeckled image. The techniques are grouped as: (1) wavelet shrinkage; (2) wavelet despeckling under Bayesian framework; and (3) wavelet filtering and diffusion.

2.1.2.1. Wavelet shrinkage. The wavelet shrinkage is based on thresholding the wavelet coefficients. It suppresses the coefficients representing noise while retains the coefficients that more likely representing image features. It is usually performed using one of the two dominant thresholding schemes: hard thresholding and soft thresholding.

Suppose the image in wavelet domain is represented as

$$o = s + n$$

where o is the observed wavelet coefficients, s is the noise-free component and n is the additive noise. The wavelet shrinkage estimator can be represented as

$$\hat{s} = Ho$$

where H denotes shrinkage factor. For the classical wavelet thresholding rules, a threshold value T is defined and H is specified as follows. For hard thresholding

$$H = \begin{cases} 0 & \text{if } |o| < T \\ 1 & \text{if } |o| \geq T \end{cases}$$

For soft thresholding

$$H = \begin{cases} 0 & \text{if } |o| < T \\ 1 - T/|o| & \text{if } |o| \geq T \end{cases}$$

where $|o|$ denotes the absolute value of o. Soft thresholding provides smoothness when applied to an image while hard thresholding preserves the features of an image. Applications of hard and soft thresholding can be found in [61,62,73–75]. Most of them have focused on developing the best uniform threshold. Adaptive thresholding which makes threshold values adaptive to the spatially changing statistics of the images has attracted more attention [76–77]. Adaptive thresholding improves the performance by incorporating additional local information such as the identification of edge into the despeckling algorithm.

The drawback of thresholding methods is that the choice of the threshold is usually done in an *ad hoc* manner.

2.1.2.2. Wavelet despeckling under Bayesian framework. An alternate approach to the standard thresholding technique is employing Bayesian rules [60,63–71]. It relies on the knowledge of the wavelet coefficient statistics. This approach assumes that p is a random variable with a given *prior* probability density function. Given a set of wavelet coefficients q, the goal is to find the Bayesian risk estimator $\hat{p}$ minimizing the conditional risk, which is the cost averaged over the conditional distribution of p (denoted as $P_{p|q}(p|q)$):

$$\hat{p}(q) = \arg\min_{\hat{p}} \int L[p, \hat{p}(q)] P_{p|q}(p|q)\, dp$$

where L is a cost function to be specified.

In [67], the two-sided generalized Nakagami distribution (GND) is used to model the speckle wavelet coefficients, and the wavelet coefficients are modeled by the generalized Gaussian distribution (GGD). Combining these statistics priors with the Bayesian MAP (*maximum a posteriori*) criterion, the algorithm can deal with either envelope speckle image or log-compressed image. Ref. [69] designed both the *minimum absolute error* (MAE) and the MAP estimators for alpha-stable signal mixed in Gaussian noise. Ref. [64] extended the approach [69] in two aspects: the use of bivariate alpha-stable distributions to model the signal wavelet coefficients and the use of oriented 2D dual-tree complex wavelet transform in the multi-scale decomposition step. Ref. [70] employed a preliminary detection of the wavelet coefficients representing the features of interest to empirically estimate the conditional PDFs of the useful feature coefficients and background noise. It has to be applied to the original speckled image before log-compression. A speckle reduction algorithm is developed by integrating the wavelet Bayesian despeckling technique with Markov random filed based image regularization [71]. The wavelet coefficients are modeled by a two-state Gaussian mixture model and their spatial dependence is characterized by a Markov random field imposed on the hidden state of Gaussian mixtures.

Most of the thresholding methods do not take into account the specific properties of the image. Wavelet despeckling under Bayesian Framework outperforms the thresholding methods by exploiting the statistics of wavelet coefficients. The disadvantage of wavelet despeckling under Bayesian Framework is that it relies on prior distributions of the noise-free image, however, in the real world, there is no speckle-free US images since speckle is inherent in US images.

Table 1
Edge map FOM of various filters [155].

	Noisy	Lee	Kuan	Gamma	Frost	Geometric	Oddy	Wavelet
FOM(%)@ $L = 1.9$	0.9	6.1	6.4	6.2	28.2	18.3	10.9	38.1
FOM(%)@ $L = 9.4$	3.7	49.5	49.6	55.0	58.6	58.5	44.4	64.1

2.1.2.3. Wavelet filtering and diffusion. Besides thresholding, we can use filtering or diffusion method in wavelet domain to reduce speckle [78,79]. Wiener filtering is applied in the wavelet domain [78]. The experimental results show that the approach performs better than wavelet thresholding visually and quantitatively. Normalized modulus-based nonlinear multi-scale wavelet diffusion (NMWD) is proposed for speckle suppression and edge enhancement [79]. The approach has more favorable despeckling properties than that of nonlinear diffusion because the multi-scale representation gives more efficient signal/noise separation. It also outperforms wavelet-based despeckling methods by taking the advantage of edge enhancement inherited from nonlinear diffusion. Both the envelop speckle image and log-compressed image can be directly processed using this technique.

A study that compares different speckle filters in the image domain and wavelet domain is presented in [155]. It compared wavelet coefficient shrinkage (WCS) filter and several standard speckle filters (Lee, Kuan, Frost, Geometric, Kalman, Gamma, etc.) It calculates the figure of merit (FOM) of the edge map to get a quantitative evaluation of edge preservation and the results show that wavelet domain filters preserve image details better (Table 1).

The disadvantage of wavelet-based despeckling methods is that the time complexity is increased due to the DWT and IDWT operations.

2.1.3. Compounding approaches

In compounding approaches, the image acquisition procedure has been modified to produce several images of the same region that are partially correlated or non-correlated, and averages them to form a single image. There are two general methods for de-correlation among the individual images. Spatial compounding is obtained by generating each original image while the transducer is located at different spatial locations [80,81]. 3D spatial compounding is adopted to reduce speckle in 3D ultrasound images [83]. Frequency compounding is generated when the transducer operates at different frequencies [82]. The compounding technique reduces speckle at the expense of increasing the complexity of image registration and reconstruction.

Some speckle reduction methods are listed in Table 2.

2.2. Image enhancement

As stated at the beginning of the preprocessing section, many methods enhance the image and remove speckle at the same time. Nonlinear diffusion is such an example. It not only preserves edges but also enhances edges by inhibiting diffusion across edges and allowing diffusion on either side of the edges. Since we already reviewed those techniques in the previous section, now we will focus on the algorithms merely for image enhancement.

Histogram equalization is used to enhance the contrast [13]. The multi-peak generalized histogram equalization was proposed in [85]. It combined multi-peak histogram equalization with local information to enhance the contrast. Ref. [86] proposed stick technique for image enhancement. Sticks (line segments) in different orientations are used as the templates and the orientation which is most likely to represent a line is selected to improve edge information. This

Table 2
Speckle reduction methods.

Method	Description	Advantage	Disadvantage
Filtering techniques [34–59]	Use moving window to convolve the filter with the image to reduce speckles	Simple and fast	1. Single scale representation is difficult to discriminate signal from noise 2. Sensitive to the size and shape of the filter window
Wavelet approaches [60–79]	Transform image to wavelet domain and remove noise by modifying wavelet coefficients	1. In wavelet domain, the statistics of the signals are simplified 2. Noise and signal are processed at different scales and orientations	DWT and IDWT computations increase time complexity
Compounding approaches [80–83]	Average images obtained by varying scanning frequency or view angle	Simple	Requires hardware support. Increases time complexity by registration and reconstruction

algorithm only enhances edges and the non-line features are not affected. A contrast-enhancement algorithm based on fuzzy logic and the characteristics of breast ultrasound images was proposed in [19]. It used the maximum fuzzy entropy principle to map the original image into fuzzy domain and then the edge and textural information were extracted to describe the lesion features. The contrast ratio measuring the degree of enhancement is calculated and modified. The defuzzification process is finally applied to obtain the enhanced image. Experimental results show that the method could effectively enhance the image details without over- or under-enhancement.

3. Segmentation

Image segmentation is a critical and essential component and is one of the most difficult tasks in image processing and pattern recognition, and determines the quality of the final analysis.

Segmentation [87] is a partition of the image I into non-overlapping regions

$$\cup I_i = I \quad \text{and} \quad I_i \cap I_j = \varnothing \quad i \neq j$$

Computer-aided diagnosis system will help radiologists in reading and interpreting sonography. The goal for the segmentation is to locate the suspicious areas to assist radiologists in diagnoses.

3.1. Histogram thresholding

Histogram thresholding is one of the widely used techniques for monochrome image segmentation [87,88]. Histogram thresholding was proposed for segmenting breast ultrasound images [89–92]. The algorithms [90,91] proposed for segmenting masses in US images involved the following steps: (1) preprocessing using cropping and median filtering, (2) multiplying the preprocessed image with a Gaussian constrain function, (3) determining the potential lesion margins through gray-value thresholding, and (4) maximizing a utility function for potential lesion margins. However, the center, width and height of the lesions needed to be selected manually or semi-manually.

Another thresholding algorithm [89,92] had four steps: First, the regions of interest (ROIs) were preprocessed with a 4×4 median filter to reduce the speckle noise and to enhance the features. Second, a 3×3 unsharp filter was constructed using the negative of a two-dimensional Laplacian filter to emphasize the elements with meaningful signal level and to enhance the contrast between object and background. Third, the ROIs were converted to a binary image by thresholding. The threshold was determined by the histogram of ROIs. If a valley of histogram between 33% and 66% of the pixel population could be found, this intensity value was selected as the threshold. If there was no such valley in that range, the intensity of 50% pixel population was selected as the threshold value. Finally, the selected nodule's boundary pixels were obtained using morphologic operations.

Ref. [93] adopted an automatic threshold method [94] to obtain the initial image for the gradient vector flow (GVF) snake to locate the tumor contour. The thresholding method is too simple and primitive, and does not perform well for the images with histograms that are unimodal.

3.2. Active contour model

The active contour model, more widely known as snake [95], is a framework for delineating an object outline from a possibly noisy 2D image, and has been massively used as an edge-based segmentation method. This approach attempts to minimize the energy associated with the current contour as the sum of the internal and external energies. The snake model modifies its shape actively and approximates the desired contour. During the deformation process, the force is calculated from the internal energy and external energy. The external energy derived from image feature energy is used to extract the contour of the desired object boundary. The internal energy derived from the contour model is used to control the shape and regularity of the contour [95].

The snake model has been extensively used for US images [93,96–100]. The active contour model was applied to a 3D ultrasonic data file for segmenting breast tumor [93,96], and a snake technique was used to obtain the tumor contour for pre- and post-operative malignant breast excision [93].

Combining intensity and texture with empirical domain-specific knowledge and directional gradient, a deformable shape-based model [97] was studied to find lesion margins automatically. A formulation of the empirical rules used by radiologists in detecting ultrasonic breast lesions was employed to automatically determine a seed point in the image. Followed by region growing to obtain an initial segmentation of the lesion, image pixels were classified according to the intensity and texture. Boundary points were found using the directional gradient of the image. These boundary points were supplied as the initial estimate to a deformable model. No manual initialization of the contour was required. The directional gradient of the image was used as the stopping criterion.

Level set method is employed to improve the active contour segmentation for ultrasound images. Ref. [98] discussed a level set maximum likelihood method to achieve a maximum likelihood segmentation of the target. The Rayleigh probability distribution was utilized to model gray level behavior of ultrasound images. A partial differential equation-based flow was derived as the steepest descent of an energy function taking into account the density probability distribution of the gray levels as well as smoothness constraints. A level set formulation for the associated flow was derived to search the minimal value of the model. Finally, the image was segmented according to the minimum energy.

The methods based on snake-deformation model were used to handle only the ROIs, not the entire image. Automatically generating

a suitable initial contour is very difficult, and the snake-deformation procedure is very time-consuming.

3.3. Markov random field

Ultrasound image segmentation can be considered as a labeling problem where the solution is to assign a set of labels to pixels, which is a natural representation for Markov random fields. Markov random field model has been used for US image segmentation [101–107]. The algorithm alternatively approximates the maximization of the posterior (MPM) estimation of the class labels, and estimates the class parameters. Markov random field model deals with the spatial relations between the labels obtained in an iterative segmentation process. The process assigning pixel labels iteratively can be achieved by maximizing either a posteriori estimation or posterior marginal estimation.

The algorithms based on Markov random field (MRF)/Gibbs random field (GRF) [101] was adopted to segment US images. The Metropolis sampler approach was used, and a new local energy was defined.

The new local neighboring energy is

$$U_{N-4}(x_{i,j}) = U_{local}(x_{i,j}|y_{N_{i,j}}) + U_{local}(x_{i-1,j}|y_{N_{i-1,j}}) + U_{local}(x_{i+1,j}|y_{N_{i+1,j}}) + U_{local}(x_{i,j+1}|y_{N_{i,j+1}}) + U_{local}(x_{i,j-1}|y_{N_{i,j-1}})$$

where $U_{local}(x_{i,j}|y_{N_{i,j}})$, $U_{local}(x_{i-1,j}|y_{N_{i-1,j}})$, $U_{local}(x_{i+1,j}|y_{N_{i+1,j}})$, $U_{local}(x_{i,j+1}|y_{N_{i,j+1}})$ and $U_{local}(x_{i,j-1}|y_{N_{i,j-1}})$ are the local energies of site $x_{i,j}$, and its four first order neighbors, $x_{i-1,j}$, $x_{i+1,j}$, $x_{i,j+1}$, and $x_{i,j-1}$.

Then, ΔU can be represented as

$$\Delta U = U_{N-4}(x'_{i,j}) - U_{N-4}(x_{i,j})$$

where $x'_{i,j}$ is accepted as a new label of site $x_{i,j}$, $U(x)$ is the global energy of each configuration.

The newly defined local energy can fit into the Metropolis sampler algorithm ΔU. The Expectation-Maximization (EM) method is used for parameters estimation of each class.

Ref. [103] used a combination of the maximum a posteriori and Markov random field to estimate the US image distortion field following a multiplicative model while labeling image regions based on the corrected intensity statistics. The MAP was used to estimate the intensity model parameters while the MRF provided a way of incorporating the distributions of tissue classes as a spatial smoothness constraint.

A segmentation algorithm for breast lesion was based on multi-resolution texture adaptive clustering [104], which improved the algorithm in [108] by using a new energy function to measure textural properties of various tissues. The segmentation problem was formulated as a maximum a posteriori estimation problem. The MAP estimation utilized Besag's iterative conditional mode algorithm for minimizing an energy function constraining the region to be close to the data, imposing spatial continuity and considering the texture of various regions. However, the input images for this algorithm were only ROIs.

Ref. [105] used a Markov random field to model the region process and to focus on the adaptive characteristics of the algorithm. It introduced a function to control the adaptive properties of the segmentation process, and took into account both local and global statistics during the segmentation process. A new formulation of the segmentation problem was utilized to control the effective contribution of each statistical component.

Ref. [106] combined EM (expectation maximization) for hyperparameter estimation and MPM (maximization of posterior marginals), and extended the EM/MPM framework to 3D by including pixels from neighboring frames in the Markov random field clique. However, there were many noisy spots in the segmentation results, and the algorithm was quite time-consuming.

The merit of MRF modeling is that it provides a strong exploitation of the pixel correlations. The segmentation results can be further enhanced via the application of maximum a posteriori segmentation estimation scheme based on the Bayesian learning paradigm [102]. However, its iteration process is complex and time-consuming.

3.4. Neural network

Neural network (NN) based methods [11,109,110] are popular in image segmentation, which transform the segmentation problem into classification decision based on a set of input features.

In [109], a NN approach was combined with wavelet analysis for US image segmentation. A multi-layered perceptron (MLP) neural network having one hidden layer was designed with variance contrast and auto-correlation contrast as input features, and trained by error back propagation.

A study [110] integrated neural network classification and morphological watershed segmentation to extract the contours of breast tumors. Textural analysis was employed to find the inputs for the NN. Watershed transformation automatically determined the contour of the tumor. However, how to select the training set was problematic, and training a NN was time-consuming.

A Bayesian neural network (BNN) with five hidden units and an output node were employed for segmentation and detection [11] where input features were the depth-to-width ratio, the radial gradient index (RGI) value, texture, and posterior acoustic behavior of the suspected lesion. At first, a radial gradient index filtering technique was used to locate the ROIs and their centers were documented as the points of interest, and a region growing algorithm was used to determine candidate lesion margins. The lesion candidates were segmented and detected by the BNN. However, the algorithm would fail if the lesion was not compact and round-like. In addition, the appropriate number of hidden units for the neural network was determined empirically.

In order to compare different segmentation methods clearly, descriptions, advantages and disadvantages of different methods are discussed briefly in Table 3.

4. Feature extraction and selection

Feature extraction and selection are important steps in breast cancer detection and classification. An optimum feature set should have effective and discriminating features, while mostly reduce the redundancy of feature space to avoid "curse of dimensionality" problem. The "curse of dimensionality" suggests that the sampling density of the training data is too low to promise a meaningful estimation of a high dimensional classification function with the available finite number of training data [111]. For some advanced classification methods, such as artificial neural network and support vector machine, the dimension of feature vectors not only highly affects the performance of the classification, but also determines the training time of the algorithm. Thus, how to extract useful features and make a good selection of the features is a crucial task for CAD systems.

The features of breast US images can be divided into four categories: texture, morphologic, model-based and descriptor features. We summarize and list the typical and effectiveness-proved features in Table 4. Certainly, one cannot use all of them at the same time. Extraction and selection of effective features is a necessary step. The general guidelines for selecting significant features mainly include four considerations: discrimination, reliability, independence and optimality [112]. However, simply combining the best performed features will not definitely make the systems work well and

Table 3
Summary of segmentation approaches.

Methods	Descriptions	Advantages	Disadvantages
Histogram thresholding method [89–92]	Threshold value is selected to segment the image	Simple and fast	No good results for images with non-bimodel histograms
Active Contour model [93,96–100]	Snake-deformation mode is utilized	It could extract lesion with different shape and keep the boundary correctly	Slow in the iteration speed
MRF [101–106]	It estimates US image distortion field following a multiplicative model while labeling image regions based on the corrected intensity statistics	Precise and accurate	Complex and time-consuming for many iterations
NN [109,110]	Segmentation is regarded as a classification task	It extracts the contours of tumors automatically	How to select the training set is problematic, and training is time-consuming and depending on the image database

effectively. The goal of feature extraction and selection is to maximize the discriminating performance of the feature group.

4.1. Texture features

Most of the texture features are calculated from the entire image or ROIs using the gray level values. FT1 (auto-covariance coefficient) is a basic and traditional texture feature which can reflect the inner-pixel correlation within an image. FT2 (BDIP)–FT3 (BVLC) measure the variation of intensities and texture smoothness, respectively. The higher value of BDIP is, the larger the variance of intensities in a block is, and the larger BVLC value indicates that the ingredients in the block are rough [20]. Both the first and second order of FT2 and FT3 can be used as the features too. FT4 is defined as the ratio of the variance, auto-correlation coefficients or intensity average inside the lesion to that outside the lesion. The larger the ratio is, the lower the possibility of the tumor being malignant is. FT5 is defined as the summation of differences among the real distribution of wavelet coefficients in each high-frequency sub-band and distribution of the expected Laplacian distribution. This feature can reflect the margin smoothness. FT6 is an order statistics-based feature vector extracted from wavelet decomposition sub-bands. After 3rd level wavelet decomposition, the length (length = 20) of order statistics filter is chosen based on Monte Carlo simulation and Akaike's final prediction criterion. Twenty mean values and 20 variance values of order statistics parameters for the 12 wavelet coefficient bands were calculated and formed 480-D feature vectors [116]. The dimension of the feature vector was reduced from 480-D to 7-D by using feature analysis. The stepwise feature selection method or PCA could be a better choice for reducing the feature dimensionality. FT7 and FT8 are defined as

$$CON = \sum_{i,j}(i-j)^2 p(i,j)$$

and

$$COR = \frac{\sum ijp(i,j) - m_x m_y}{\sqrt{S_x^2 S_y^2}}$$

where $p(i,j)$ is the probability that two pixels with gray value i and gray value j are in a fixed distance apart, and

$$m_x = \sum_i i \sum_j p(i,j), \quad m_y = \sum_j j \sum_i p(i,j)$$
$$S_x^2 = \sum_i i \sum_j p(i,j) - m_x^2, \quad S_y^2 = \sum_j j \sum_i p(i,j) - m_y^2$$

Another way to define FT7 is $CON = E\{I(i,j) \cdot I(i+\triangle i, j+\triangle j)\}$, where $I(i,j)$ is the gray value at position (i,j) and $(\triangle i, \triangle j)$ is the distance between two pixels. By the same notation, FT9 is defined as $Diss = E\{I(i,j) - I(i+\triangle i, j+\triangle j)\}$. FT10 is defined as $N_{\text{non-zero}}/N_{\text{all}}$, where $N_{\text{non-zero}}$ is the number of pixels having non-zero average gradients and N_{all} is the total number of pixels in the region. FT7, FT8 and FT10 were features labeled with strong distinguishing ability in [139]. FT11 is calculated from the minimal rectangular ROI containing the lesion:

$$COR = \sum_{n=0}^{N_R-1} \frac{\overline{C_y}(n)}{\overline{C_y}(0)}$$

where

$$\overline{C_y}(n) = \sum_{m=0}^{M_R-1} c_y(m,n), \quad C_y(m,n) = \sum_{p=0}^{N_R-1-n} I^2(m,n+p) I^2(m,p)$$

M_R is the number of pixels in the lateral direction of the ROI and N_R is the number of pixels in the depth direction of the ROI, and I is the gray level value matrix of the ROI. Because COR is a sum, it includes not only the texture information but also the size information.

Based on understanding of the posterior acoustic behavior or posterior shadow, different numeric expressions are proposed to calculate FT12. In [90], three ROIs were defined whose width and depth were the same as the ROI contains the lesion itself. As Fig. 3 shows, the post ROI represents the posterior region of the lesion and the right ROI and left ROI are adjacent tissues at the same depth of the post ROI. The narrow blank boundaries are used to avoid the edge shadows. Finally, the minimum side difference (MSD) is defined as: $\text{MSD} = \min(A_{\text{post}} - A_{\text{left}}, A_{\text{post}} - A_{\text{right}})$, where A_{post}, A_{right} and A_{left} are the average gray-level values of the corresponding ROIs. In [131], another method to calculate posterior shadow was proposed. First a skewness image is built by

$$Skew(x,y) = \frac{1}{N} \sum_{(x',y') \in A} \frac{(I(x',y') - \overline{I(x',y')})^3}{\sigma_A^3}$$

where A is a specified region centered at point (x, y), $I(x',y')$ is the gray value in the original image, N is the total number of data points in region A and σ_A is the standard deviation of the gray values in area A. The skewness image is filtered with a threshold to get the detection points, i.e., the shadow. In [25], the posterior shadow was defined as the difference between the gray scale histograms of the regions inside the lesion and posterior to the lesion. For the same characteristic of breast lesions, we can use different ways to define the numeric expressions. To find more accurate and efficient expressions should be one of the future works.

FT13 is the Boltzmann/Gibbs entropy over the gray scale histogram relative to the maximum entropy. The higher the entropy is, the more homogeneous the lesion is. FT15–FT16 are well-known texture features which have already been well defined. However, they are not frequently used in recent US image characterization. This may be due to their high computation cost. The definition of the fractal dimension (FT17) is similar to the Hausdorff dimension [137]. Informally, the dimension d can be calculated by $N = s^d$, where N is

Table 4
Features.

Feature category	Feature description
Texture features	FT1: Auto-covariance coefficients [3,13,16,20,22,125–127,132-134] FT2: Block difference of inverse probabilities (BDIP) [20] FT3: Block variation of local correlation coefficients (BVLC) [20] FT4: Variance, auto-correlation, or average contrast [18,109,120] FT5: Distribution distortion of wavelet coefficients [109] FT6: Mean and variance of the order statistics after wavelet decomposition [116] FT7: Contrast of grey level values [22,132,134,139] FT8: Correlation of the co-occurrence matrix [139] FT9: Dissimilarity [22,132,139] FT10: Relative frequency of the edge elements [139] FT11: Auto-correlation in depth of R (COR) [11,90,113,129,130,140,141] FT12: Posterior acoustic behavior, minimum side difference (MSD) or posterior acoustic shadow [9,11,25,90,113,117,129,130,140,141] FT13: Homogeneity of the lesion [25] FT14: Standard deviation of gray value and its gradient of the lesion [129] FT15: SGLD matrix based features: correlation, energy, entropy, sum entropy, difference entropy, inertia and local homogeneity [16,21,24] FT16: GLD matrix based features: contrast, mean, entropy, inverse difference moment and angular second moment [16] FT17: Fractal dimension and related features [21,150]
Morphologic features	FM1: Spiculation [18,119,141] FM2: Depth to width ratio (or width to depth ratio) [9,10,11,18,90,113,114,119,121,136,140-143] FM3: Branch pattern [18,119] FM4: Number of lobulations [18,25,119,141,152] FM5: Margin sharpness [17,123,129,130] FM6: Margin echogenicity [17,123] FM7: Angular variance in margin [17,123] FM8: Number of substantial protuberances and depressions (NSPD) [10] FM9: Lobulation index (LI) [10] FM10: Elliptic-normalized circumference (ENC) [10] FM11: Elliptic-normalized skeleton (ENS) [10] FM12: Long axis to short axis ratio (L:S) [10] FM13: Area of lesion [10,25,143] FM14: Normalized radial gradient (NRG) along the margin [11,90,113,114,129,130,140,141] FM15: Margin circularity [25] FM16: Degree of abrupt interface across lesion boundary [152] FM17: Angular characteristic [152]
Model-based features	FB1: β_c, β_s, and β_{env} of PLSN model [122,144]2 FB2: m and Ω of Nakagami model based features [120,122]2 FB3: Single and combined parameters of GS model [122] FB4: $b_{inverse}$ and M of K distribution model based features [120]2 FB5: Normalized skewness K [124] FB6: Signal to noise ratio of the envelope η [124] FB7: Normalized spectral power α [124] FB8: Margin strength [124] FB9: Quality of margin β [120,124] FB10: Speckle factor [84,118]
Descriptor features	FD1: Non-circumscribed or spiculated margins [9,14,115,119,121,135,136,138,142,143] FD2: Shape (round, oval or irregular) [9,14,115,119,121,135,136,138,142] FD3: Presence of calcifications [9,119,115,135,136] FD4: Posterior shadow or posterior echo [119,121,135,136,142] FD5: Decreased sound transmission or acoustic transmission [14] FD6: Echogenicity [14,121,135,136,142] FD7: Heterogeneous echo texture [115,121,135,136,143] FD8: Duct extension [119] FD9: Thickened cooper ligaments [143] FD10: Antiparallel orientation [14,135,138] FD11: Distortion, echogenic halo or rim of surrounding tissue [14,136] FD12: Bilateral refraction sign [121] FD13: Microlobution [119,136]

the number of similar pieces, s is the magnification factor, and d is the "dimension" of the scaling law, known as the Hausdorff dimension, and the fractal dimension-based features are verified to be the valuable features [21].

4.2. Morphologic features

Unlike texture features extracted from the rough ROIs, the morphologic features focus on some local characteristics of the lesion, such as the shape and margin.

In the polar coordinates (r,θ), each boundary pixel is represented as $r(\theta)$ and FM1 (spiculation) is the ratio of low-frequency component (area under the graph $|R(w)|$ from 0 to $\pi/4$) to high-frequency component (area under the graph $|R(w)|$ from $\pi/4$ to π), where $|R(w)|$ is the Fourier transform of $r(\theta)$ and the cutoff frequency $\pi/4$ was experimentally chosen [18]. The larger the value is, the lower possibility of the tumor being malignant is. FM2 is one of the most effective distinguishing features mentioned in many papers. Malignant lesions tend to have the ratio bigger than 1 while benign lesions usually have the ratio smaller than 1. FM3–FM4 are the

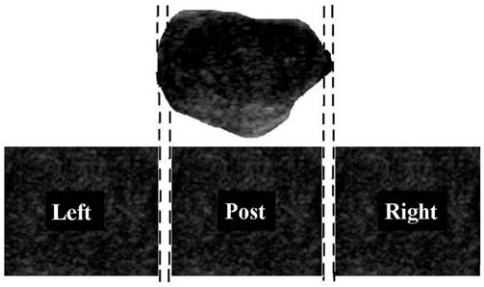

Fig. 3. ROIs used to define the posterior acoustic behavior [90].

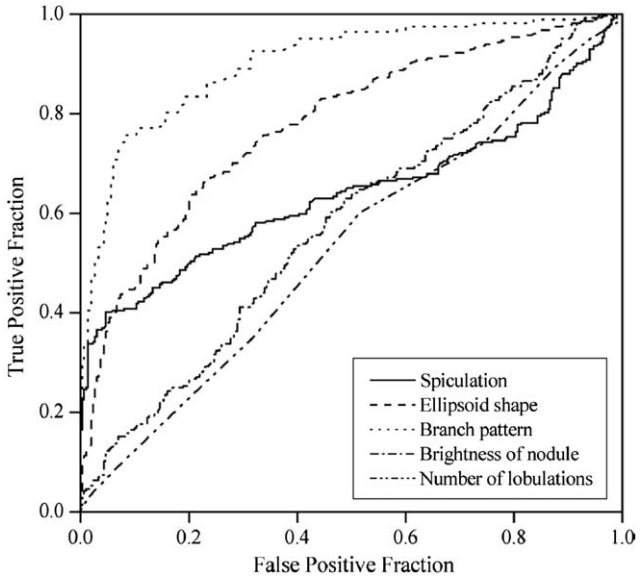

Fig. 4. ROC curves of classification using five features [18].

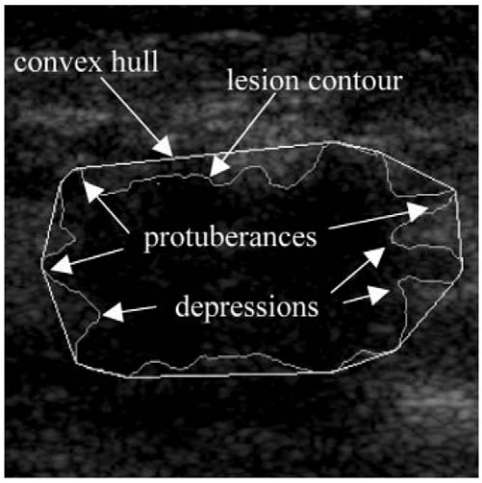

Fig. 5. Convex hull of a lesion (FM8) [10].

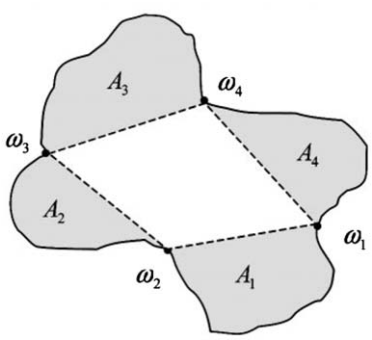

Fig. 6. Concave hull of a lesion (FM8 and FM9) [10].

Table 5
Quantitative margin features comparison [123].

Cases	Margin sharpness	Margin echogenicity	Angular variance
Benign	73.7 ± 10.0	22.9 ± 13.2	0.39 ± 0.13
Malignant	66.5 ± 12.0	14.8 ± 7.7	0.27 ± 0.06
P (2-tailed t-test)	0.027	0.0048	0.000009

Values are mean ± SD.

numbers of local extremes in the low-pass-filtered radial distance graph and the curve-fitted radial distance graph, respectively. Malignant lesions tend to have higher value of FM3 or FM4. The ROC curves of classification using FM1–FM4 are compared in Fig. 4. For FM5–FM7, the lesion is divided into N sectors, and in each sector, the mean gray levels of the pixels in the inner and outer shells are compared. By using a user-defined threshold, some of the sectors are chosen distinctly. The margin sharpness is calculated as (number of distinct sectors)∗100/N. Margin echogenicity is the mean gray level difference of the inside and outside of the sector. Angular variance in margin is the ratio SD/mean for the difference in the mean gray level of the inside and outside of each sector. All of the above three features are proved to be significantly different by Student t-test when they are used to distinguish benign and malignant lesions [123]. Table 5 summarizes the mean ± SD and p values of the three features.

FM8–FM12 are newly proposed morphologic features [10]. As shown in Figs. 5 and 6, a breast lesion is delineated by a convex hull and concave polygon. Given a threshold angle of θ ($\theta \in \{20^\circ, 30^\circ, 40^\circ, 50^\circ, 60^\circ\}$), let $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_p\}$ and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_p\}$ denote the set of representative convex and concave points of a lesion boundary, respectively, where p and d are the numbers of points in each set. Thus, the NSPD is defined as p+d. Ideally, a malignant breast lesion has a larger NSPD. FM9 (LI) is defined as $(A_{\max} - A_{\min}) * N / \sum Ai$ where $A_{\max}$ and $A_{\min}$ are the sizes of maximum and minimum lobes as illustrated in Fig. 6 and N is the total number of the lobes. LI is an effective complement of NSPD, and can correctly characterize the benign lesions with multiple large lobes of similar sizes which are easily to be misclassified by NSPD [10]. FM10 (ENC) is defined as the circumference ratio of the lesion to its equivalent ellipse (Fig. 7), and it represents the anfractuosity of a lesion which is a characteristic of malignant lesions. FM11 (ENS) is defined as the number of skeleton points normalized by the circumference of the equivalent ellipse of the lesion. The calculation cost of this feature is relatively high. Same as FM10, a malignant lesion tends to have a higher value of FM11. These four features capture mainly the contour and shape characteristics of the lesion. FM12 (L:S) is the ratio of the long- to short-axis, where the long and short axes are determined by the major and minor axes of the equivalent ellipse. Therefore, L:S ratio is different from the traditional depth/width ratio (FM2) because it is independent of the scanning angle. For both FM12 and FM13 (lesion area), clinically, the larger the value is, the lesion is more likely malignant. Among the five newly proposed features (FM8–12), NSPD is proved to be the most important feature, and NSPD, LI, ENS and ENC are better than lesion size, L:S ratio and depth/width ratio. FM14 is used to measure the average orientation of the gray level

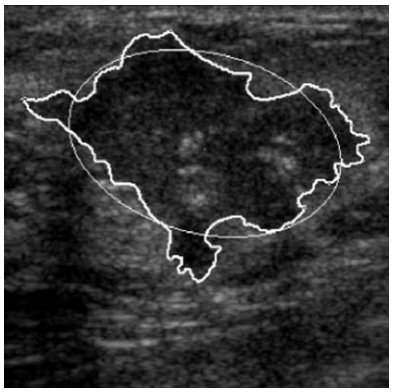

Fig. 7. Equivalent ellipse of a lesion (FM10 and FM11) [10].

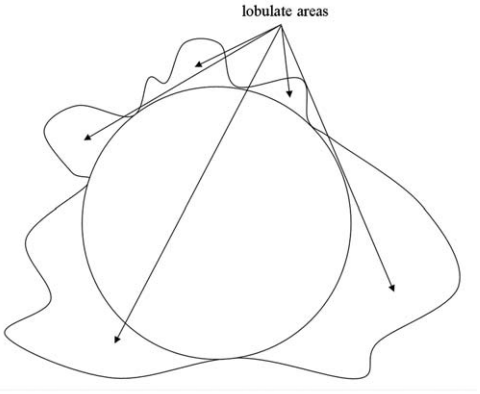

Fig. 8. Maximum inner circle and lobulate areas of a lesion [152].

gradients along the margin. The formula for FM14 is

$$NRG = \frac{\sum_{j=0}^{J-1} \nabla I(x_j, y_j) \cdot \hat{r}(x_j, y_j)}{\sum_{j=0}^{J-1} \|\nabla I(x_j, y_j)\|}$$

where ∇I is the gradient image computed using Sobel filter, J is the number of points on the margin and $\hat{r}(x_j, y_j)$ is the unit vector in the radial direction from the geometric center to the point (x_j, y_j). FM15 is defined as the standard deviation of the normalized distances of every boundary point to the ROI's geometric centers [25]. A high value of FM15 is a sign of malignancy. FM16 and FM17 are proved to be two important features in [152]. To define FM16 and FM17, a distance map should be first calculated. For each pixel in the image, its value in the distance map is defined as the shortest distance from the pixel to the lesion boundary. FM16 is designed to estimate the degree of abrupt interface across the lesion boundary. The formula is $LB_D = avg_{Tissue} - avg_{Mass}$,

$$avg_{Tissue} = \frac{\sum_{dis(n)=1}^{k} I(n)}{N_{Tissue}}, \quad avg_{Mass} = \frac{\sum_{dis(n)=1}^{k} I(n)}{N_{Mass}}$$

where $I(n)$ is the gray level value of pixel n, $dis(n)$ is the value of pixel n in the distance map, N_{Tissue} is the number of pixels in the surrounding tissue and N_{Mass} is the number of pixels in the outer mass. Both the surrounding tissue and outer mass areas are composed of pixels whose distances to the lesion boundary are no more than k in the distance map. Width k was set to 3 in [152]. The likelihood of malignancy was decreased as the increasing of FM16. FM17 (angular characteristics) is defined as the sum of numbers of local maxima in each lobulate area [152]. Fig. 8 displays an example of lobulate areas. By the maximum inside circle, some lobulate areas are partitioned from the mass. Some small lobulate areas with maximum distance < 4 are discarded. For the remained lobulate areas, the local maxima in each lobulate area are grouped as follows: when a new local maximum is discovered, the Euclidean distance from it to the center of the grouped local maxima is calculated. If the distance is larger than a predefined threshold, this local maximum is regarded as a new grouped local maximum. At last, the total number of local maxima in the lobulate areas represents the angular characteristics. In [152], the predefined threshold was set to 10. The larger the value is, the lesion is more likely malignant.

4.3. Model-based features

Model-based feature is a special type of US features which focuses on the backscattered echo from the breast tissue. Different models are developed to simulate the echo of the backscattered envelope. Once a model is chosen and the echo is modeled, the parameters of the model can be used as the features to distinguish malignant and benign lesions. The models have been used for breast cancer diagnosis including power-law shot noise (PLSN) model, Nakagami model, K distribution model and generalized spectrum (GS) model. Comparing with texture features and morphologic features, the advantages of model-based features are that they are not influenced by the experience of the radiologists, and not influenced by the ways in which the images are collected. They are operator- and machine-independent [120]. The disadvantage of model-based features is that the background of the models is quite complex and the estimation of the parameters is very complicated as well.

4.4. Descriptor features

Descriptor features are easier to understand because they are actually the empirical classification criteria of the radiologists. Most of them are descriptive and have no numeric expressions. The reason we list this type of features is that some of the useful features for current CAD systems are transferred from the descriptor features. There are still useful descriptor features that have not been transferred to numeric expressions, and cannot be used in CAD systems. In this subsection, we have chosen the most frequently used descriptor features which are proved to be effective in distinguishing benign and malignant lesions. Except FD2 and FD12, all the listed descriptor features are malignant characteristics.

FD1 and FD2 are the most powerful features to characterize malignant lesions. For FD2, oval or round shape is a sign for benign and "taller than wide" or other irregular shape is a sign for malignancy. FD3 describes calcifications or microcalcifications in the lesion. FD4 is called posterior shadow or posterior echo, and it focuses on the region posterior to the lesion ROI which has darker gray value than that of the surrounding. FD5 is defined as the shadow effect of surrounding tissues. FD6 works well for differentiating large tumors but not small tumors. FD7 is an argumentative feature. This might be caused by the subjective nature of this feature so that an accurate numeric expression of FD7 is needed. FD8 is a projection from the nodule that extends radially within or around a duct and toward the nipple [119]. FD9 represents thickened suspensory ligaments of the breast which tend to stretch over time. FD11 describes the echogenicity of the surrounding tissue of the tumor. FD12 is an acoustic phenomenon that mostly occurs in benign tumors. FD13 is recognized by the presence of many small lobulations on the surface of the solid lesion. Most of these descriptor features are included in the Breast Imaging Reporting and Data System (BI-RADS) [146].

4.5. Other features

Sometimes, other information can be integrated to help the diagnosis. Patient's age is proved to be an effective feature to diagnose malignancy [14,17,122,123]. Also, family disease history is another useful feature for the diagnosis.

4.6. Reduce feature space dimension

With so many features available, the crucial task is to find an optimal set of features with relative low dimension. The feature extraction transforms the coordinate system to improve a determined goal; whereas feature selection only reduces the dimensionality, i.e., does not change the coordinate system of the data [156].

4.6.1. Feature extraction

Feature extraction, linearly or nonlinearly, transforms the coordinate system of the original variables [156]. The most well-known feature extraction technique is principal component analysis (PCA). PCA performs on the symmetric covariance matrix or symmetric correlation matrix, and solves the eigenvalues and eigenvectors of the matrix. PCA is good at reducing the high dimensional correlated features into low dimensional features. The feature vector of the autocovariance coefficients can be optimized by PCA effectively [13,16]. Other feature extraction techniques such as factor analysis (FA) [162], independent component analysis (ICA) [158], and discriminant analysis (DA) [159] can also be used to reduce the feature dimension.

4.6.2. Feature selection

Generally, algorithms for feature selection can be categorized into two classes: wrapper and filter. Filter approach (such as FOCUS and Relief algorithms [157]) selects features using a preprocessing step and does not take into account the bias of induction algorithms. On the contrary, to search for a good subset of the features, wrapper approach uses the induction algorithm as a part of the evaluation function. Ref. [157] provided detailed explanations and summaries of these two classes of feature selection algorithms. As the wrapper approach has obvious advantages over filter approach, especially for complex feature data set, wrapper approach has more applications in breast cancer detection [9,21,113,114]. For example, [114] applied a wrapper approach (linear stepwise feature selection) to a feature set composed of 15 sonographic features of breast cancer and found that the two most significant features were the average orientation of gray level gradients along the margin and depth-to-width ratio.

5. Classifiers

After the features have been extracted and selected, they are input into a classifier to categorize the images into lesion/non-lesion or benign/malignant classes. Majority of the publications focuses on classifying malignant and benign lesions (usually called lesion classification), and some of the articles focus on classifying lesions and non-lesions (usually called lesion detection), and only a few of them focus on both. Lesion detection is necessary before lesion classification. We summarize the different classifiers commonly used in breast cancer detection and classification in Table 6.

5.1. Linear classifiers

Frequently used linear classifiers for breast cancer detection and classification are linear discriminant analysis [160] and logistic regression (LOGREG) [161]. The main idea of LDA is to find the linear combination of the features which best separate two or more classes of the data. If there are n classes, and LDA classifies the data by the following linear functions:

$$f_i = \mu_i C^{-1} x_k^T - \frac{1}{2}\mu_i C^{-1}\mu_i^T + \ln(P_i), \quad 1 \le i \le n$$

where

$$C = \frac{1}{N}\sum_{i=1}^{n} n_i C_i, \quad P_i = \frac{n_i}{N}$$

n_i is the number of samples in the ith class, N is the number of total samples, μ_i is the mean of class i, and C_i is the covariance matrix of class i.

The above parameters can be obtained from the training data. When a new data x_k is in, it is assigned to class i with the highest f_i.

Logistic regression is a model for predicting the probability of an event occurring as a function of other factors. The probability of $X = x_1, x_2, \ldots, x_n$ is formulated as

$$\text{logit}(P) = \log\frac{P}{1-P} = b_0 + \sum_{i=1}^{n} b_i x_i$$

where $b_0, \ldots, bn$ are model parameters which could be estimated from the training data. When LOGREG is used to classify two-class problem, for each feature vector x_i, threshold $= 0.5$ is used to decide which class X belongs to.

In [90], LDA was applied to the data set of 400 cases with four automatically extracted features. The average A_z under ROC curve was 0.87 over eleven independent trials. In [123], LOGREG was used to determine the probability of malignancy in a database of 58 cases. Three margin-based features were evaluated and the area under the ROC curve with the best feature combination of age, margin echogenicity and angular variation was 0.87 ± 0.05. Here, we can see that the performances of LDA and LOGREG are not high because the classifiers are linear, and for nonlinear separable data, the methods have intrinsic limits.

5.2. Artificial neural networks

Artificial neural networks are the collection of mathematical models that imitate the properties of biological nervous system and the functions of adaptive biological learning [2]. It is a self-learning system that changes its parameters based on external or internal information that flows through the network during the learning phase. ANN is composed of an input layer, an output layer and one or more hidden layers. Layer is composed of neurons. In the field of breast cancer detection and classification, three types of artificial neural networks are frequently used: Back-propagation neural network, self-organizing map (SOM) and hierarchical ANN [10,17,18,109,125–127].

5.2.1. Back-propagation neural network

Back-propagation (BP) neural network is a feed-forward ANN with supervised learning process. Frequently used back-propagation neural networks have one or two hidden layers and 2–10 neurons in each layer. There is no universal rule to decide the number of layers or number of neurons in each layer. In [10], a BP neural network was used in the proposed CAD and the result was compared with that in [125]. The CAD in [10] achieved $A_z = 0.959 \pm 0.005$ with the selected morphologic features and outperformed the one in [125]. Ref. [150] combined K-means classification with BP neural network. K-means classification was used to select training samples for BP neural network and only those samples within a distance to the cluster center would be used for training. The performance of back-propagation neural network is better than that of linear classifiers. However, the training process is stochastic and unrepeatable even with the same data and same initial conditions.

Table 6
Classifiers.

Classifier	Features used	Advantage	Disadvantage
Linear classifiers: Construct decision boundaries by optimizing certain criteria: LDA and LOGREG [9,90,116,120,121,122,123,124,139,143,152]	Text features (FT6–FT8, FT10, FT12), morphologic features (FM2, FM5–FM7, FM14), descriptor features (FD1–FD4, FD6–FD7, FD9, FD12)	Simple and effective for linearly separable data	Poor performance for nonlinearly separable data. Poor adaptability for complex problem
ANNs: Construct nonlinear mapping functions: Back-propagation, SOM and hierarchical ANN [10,16,17,18,109,125,126,127,150]	Texture features (FT1, FT4, FT5), morphologic features (FM1–FM4, FM8–FM13)	Robustness, no rule or explicit expression is needed, and widely applicable	Long training time, initial value dependent, unrepeatable, over-parameterization and over-training
BNN: A probabilistic approach to estimate the class conditional probability density functions [11,129,130,141]	Texture features (FT11, FT12, FT14), morphologic features (FM2, FM5, M14)	Priori information can incorporate in models, useful when there is finite training data	Need to construct model and estimate the associated parameters
Decision tree: A tree structure with classification rules on each node [22,132]	Texture features (FT1, FT7, FT9)	Low complexity	Accuracy depends fully on the design of the tree and the features
Support vector machines: Map the input data into a higher dimension space and seek an optimal hyperplane to separate samples [3,20,21,25,133]	Text features (FT1–FT3, FT12, FT13), morphologic features (FM4, FM13, FM15)	Training process is faster than NN's, Repeatable training process, good performance	Supervised learning (training data should be labeled), parameter-dependent
Template matching: Uses retrieval technique to find the most alike image in the database and assign the query image to the class of the most alike image [13,16,134]	Texture features (FT1, FT7, FT9, FT15, FT16)	No training process needed, new data can be directly added to the system	Requiring large size database, images should come from the same platform to archive better performance
Human classifiers: Physicians/radiologists, use empirical criteria to classify US images [14,115,135,136,138,142]	Descriptor features (FD1–FD14), morphologic features (FM2)	Incorporate human knowledge and use the features that cannot be used by computers	Interobserver variability, unstable and inaccurate, human error, and subjectiveness

5.2.2. SOM

SOM can automatically classify input data into different classes (number of classes could be more than 2) and it is a totally unsupervised method. The disadvantages are that the number of parameters grows exponentially with the dimension of input space, and user cannot decide the number of classes. In [126], SOM was employed as the classification method with 24-D auto-correlation texture features. By 10-folder cross-validation, A_z was 0.9357 ± 0.0152 on a data set of 243 lesions images.

5.2.3. Hierarchical ANNs

In a hierarchical ANN structure, individual ANNs are combined into a tree structure, and each node is associated with an ANN. In [127], a 2-layer hierarchical ANN was developed with 4 ANNs in the first layer and one ANN in the second layer. The performance of the proposed hierarchical ANN system was $A_z = 0.9840 \pm 0.0072$ on a data set of 1020 images. Although the performance is high, the extraction of the 4 ROIs is troublesome and the whole training process is time-consuming.

5.3. Bayesian neural networks

Bayesian neural network is a kind of neural network using Bayesian method to regularize the training process [145]. The idea behind BNN is to cast the task of training a network as a problem of inference, which is solved using Bayes' theorem [128]. Bayesian neural network is more optimal and robust than conventional neural networks, especially when the training data set is small.

A BNN with one hidden layer and five neurons in the hidden layer was chosen to detect lesions [11]. This work focused on distinguishing true lesions from non-lesions. The performance was $A_z = 0.84$ on the database of 757 images. Using the same database, in [129], two BNNs were trained and tested separately with different tasks. One was used to classify true lesions from non-lesions, and the other was used to classify malignant lesions from other detections. The performance of these two BNNs were $A_z = 0.91$ and 0.81, respectively. In [130], a 3-way BNN was used to classify the data into three classes (malignant, benign and non-lesion). To evaluate the performance, the output can be projected to 2-way classifiers. In this way, on database of 858 cases (1832 images), the performance of classifying lesions from non-lesions was $A_z = 0.92$, and the performance of classifying malignant from other detections was $A_z = 0.83$. The BNN model is easy to incorporate priori information, but to estimate those statistical parameters requires a relatively huge database.

5.4. Decision tree

A decision tree is a simple tree structure where non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes. Each non-terminal node has a threshold associating with one or more features to divide the data into its descendents, and the process stops when each terminal node only contains one class. Thus decision tree can be used as a classification tool after the thresholds are set in the training process. Comparing with neural networks, the decision tree approach is much simpler and faster [2]. However, it highly relies on the design of classification rules on each non-terminal node and the set of threshold values.

A well known algorithm for constructing decision trees is C4.5 [131]. This algorithm has been incorporated into the free classifier package WEKA (it is called J48 in WEKA) and widely used in artificial intelligence. An updated version C5.0 provides a number of improvements on C4.5. In [132], algorithm C5.0 was used to build the decision tree with 153 training samples and 90 testing samples. Covariance coefficients of the ROIs were features inputting to the decision tree, and the performance on the testing data set was accuracy = 96% (86/90), sensitivity = 93.33% (28/30) and specificity = 96.67% (58/60), respectively. The performance was compared with that of an experienced physician on the same testing data set and experiment result showed that the proposed CAD did a better job. Ref. [22] used bootstrap technique to train the decision tree with small size training sets which were parts of the database in [132]. Bootstrap technique was proved to be effective and useful, especially, for the case that a huge database was not available.

Table 7
Different classification targets: lesion/non-lesion and malignant/benign.

Target	Features used	Classification method	References
Lesion detection: Distinguish lesions from non-lesions	Texture features (FT6, FT11–FT13), morphologic features (FM5, FM14), model-based feature (FB1)	Linear classifier, BNN	[11,116,129,130,144]
Lesion classification: Distinguish malignant lesions from benign lesions	Texture features (FT1–FT5, FT7–FT16), morphologic features (FM1–FM13, FM15), model-based feature (FB1–FB10), descriptor features (FD1–FD2, FD4, FD6, FD7, FD9, FD12), patient's age	LDA, LOGREG, ANN, BNN, decision tree, SVM, template matching	[3,6,9,10,11,13,16–18,20,24–26,84,90, 109,113,120,121,122,124–127,129,130, 132–134,139,143,144,150]

5.5. Support vector machine

Support vector machine, is a supervised learning technique that seeks an optimal hyperplane to separate two classes of samples. Kernel functions are used to map the input data into a higher dimension space where the data are supposed to have a better distribution, and then an optimal separating hyperplane in the high-dimensional feature space is chosen.

In [3,20,21,133], SVM was applied to classify malignant and benign lesions. Both the performance and time cost of SVM were compared with ANN on the same data set [133]. Experiment result showed that SVM ($A_z = 0.970$) not only outperformed the ANN ($A_z = 0.956$), but also was almost 700 times faster than ANN. Ref. [21] proposed a fuzzy support vector machine (FSVM) based on a regression model. The performance of the FSVM outperformed both the ANN and SVM with classification accuracy = 94.25%. The drawback of SVM is that the parameters such as the cost parameter C which controls the trade-off between allowing training errors and forcing rigid margins, and the kernel type needs to be tuned. Also, the mapping to higher dimension is complex and training time increases exponentially with the input data dimension.

5.6. Template matching

Image retrieval technique can be used to classify malignant and benign lesions. The methods use feature vectors to represent the query image and the images in the database. Based on the similarity formula, the distance between the query image and each image in the database was calculated. The final class of the query image was decided by combining the first K retrieved images with the K highest similarity scores.

In [134], texture features were used directly as the feature vector to calculate the similarity score and the disadvantage of the method is that it requires the images in the database come from the same platform. In [13], the principle component analysis (PCA) was applied to the entire database to form a basic set of the images and each image was represented by a weighted linear combination of the images in the basic set. The weight vector was the new feature vector used to calculate the similarity score. This method was robust with the images from different sources.

The advantage of using image retrieval technique to classify breast lesions is that no training is needed and new images can be incorporated into the system easily. The disadvantages are that for some systems the images in the database have to come from the same platform to guarantee that the similarity measure is fair, and the running time of the algorithm increases if the size of database increases. However, to get a better performance, the method requires that the database is big enough to include various lesion types. There is a trade-off between database's size and algorithm's efficiency.

5.7. Human classifier

Human classifiers imply the radiologists who classify the lesions using empirical criteria. They are not a component of CAD systems.

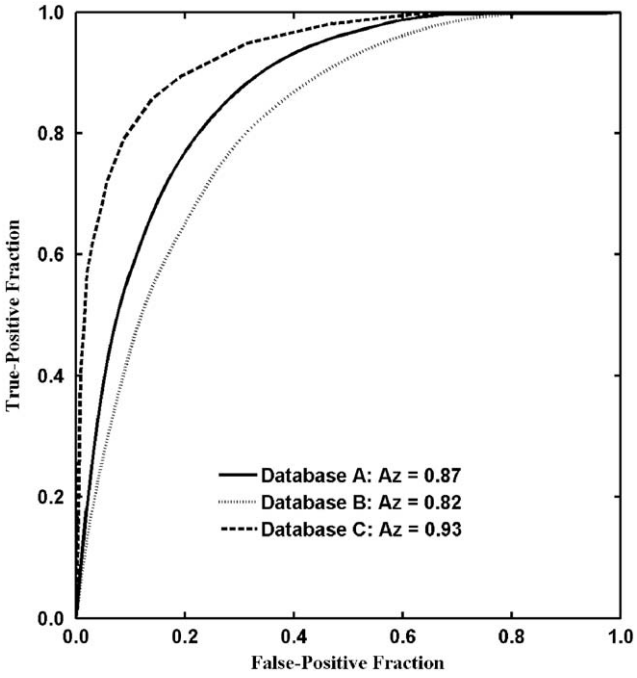

Fig. 9. ROC analysis of three databases [90].

CAD has the advantage over human classifiers since CAD is fast, stable, accurate, objective and consistent. In [9], experiment result showed that with CAD, the average diagnosis accuracy of five radiologists was improved to $A_z = 0.90$ from 0.83.

We summarize the classification targets and the results in Table 7.

6. Evaluations

As we all know, to evaluate different CAD systems fairly and correctly, the same database should be used. However, there is no a benchmark of US image database available to the public yet, and most of the works in this field are done by using their own databases. Not only the sizes of these databases are different, but also the portions of each class are different, and the sources of images are different as well. Different image sources imply that the US images are acquired by different equipments or techniques. For example, the images obtained with or without spatial compounding technique perform differently in the same CAD system [113], and they should be treated separately. Without a public database accessible by the researchers, even though the same evaluation criteria are used, it is still hard to make the evaluation fair and justified.

Next we study several frequently used evaluation criteria. A receiver operating characteristic (ROC) curve is most frequently used because of its comprehensive and fair evaluation ability. A ROC curve is a plotting of true positive fraction (TPF) as a function of false

Table 8
Databases used by CAD systems.

Database	Description	Source	Performance
DB1 [129]	The database consists of two parts: 1. 757 images, including 229 complex cysts, 334 benign solid lesions and 194 malignant lesions 2. 1740 images, including 258 complex cysts, 520 simple cysts, 210 benign solid lesions, 87 other benign breast disease, 87 malignant lesions and 578 normal images	The first subset of DB1 is obtained from Northwestern University, Chicago, IL. The second subset of DB1 is obtained from University of Chicago, Chicago, IL	$A_z = 0.94$ and 0.91 on training and testing data sets, respectively
DB2 [116]	The database has 204 US images. 204 ROIs are labeled with lesions and 816 ROIs are labeled without lesions	Images are acquired at Thomas Jefferson University Hospital, Philadelphia	$A_z = 0.91$
DB3 [133]	The database consists of two subsets obtained from different periods: 1. 88 benign lesions and 52 malignant lesions 2. 215 benign lesions and 35 malignant lesions	Taiwan, China	$A_z = 0.97$
DB4 [10]	This database consists of two subsets: 1. 91 benign lesions and 69 malignant lesions 2. 40 benign lesions and 71 malignant lesions	Images are collected from a database of a medical center in Taiwan, China	$A_z = 0.95$
DB5 [18]	The database consists of two subsets: 1. 300 benign lesions and 284 malignant lesions 2. 167 benign lesions and 99 malignant lesions	Images are provided by the Seoul National University Hospital, Seoul, Korea	$A_z = 0.95$
DB6 [21]	51 Benign lesions and 36 malignant lesions	Images are acquired from 2nd Affiliated Hospital of Harbin Medical University, Harbin, China	Accuracy = 94.25%, sensitivity = 91.67%, specificity = 96.08%, PPV = 94.29%, NPV = 94.23%

As mentioned before, there is no a benchmark accessible to the public yet, therefore, the results listed just have some reference value.

positive fraction (FPF). The area (A_z) under the ROC curve can be used as a criterion. Fig. 9 shows an example of ROC curve evaluation of the performance of CAD systems using three different data sets. Other frequently used criteria are [10,14,17,18,21,115]:

$$Overall\ accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Positive\ predictive\ value\ (PPV) = \frac{TP}{TP + FP}$$

$$Negative\ predictive\ value\ (NPV) = \frac{TN}{TN + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

The last formula is Matthew's correlation coefficient (MCC), which has seldom been used for breast cancer CAD performance evaluation. However, MCC is a powerful accuracy evaluation criterion of machine learning methods. Especially, when the number of negative samples and positive samples are obviously unbalanced, MCC gives a better evaluation than overall accuracy. As more and more breast cancer CAD systems employed machine learning methods, such as SVM, ANN and BNN, MCC should be used as an additional evaluation criterion.

The performance of some CAD systems and the databases used are listed in Table 8.

7. Future directions

Masses and microcalcifications are both important signs of breast cancer [1,2]. Currently, in the field of breast cancer CAD systems using US images, most of the works focus on mass detection and classification since the ordinary US images can hardly show microcalcifications. One of the future directions is high-resolution US imaging devices which can support microcalcification detection [27,28]. Some successful experiments proved that high-resolution US can show microcalcifications within breast cancer with a sensitivity of 95%. With the development of US equipments and refinement of image techniques, the rate of microcalcification detection and characterization will be higher. Besides, such advancement will also improve the detection of blood flow, an indicator of malignancy [147].

Three-dimensional ultrasound imaging is another future direction which has been paid more and more attention. Three-dimensional ultrasound imaging can provide more comprehensive information of the breast lesion than 2D imaging and incorporate all 2D characteristics. The advantages of 3D US are especially obvious in a CAD system because CAD system is good for processing a large amount of data in a short time, which can greatly reduced the variability of the observations and the work load of radiologists. Most of the 2D techniques can be directly applied to 3D US images with some preprocessing or post-processing methods [9,16,23–25,136]. Some newly developed methods, especially for 3D US images, can be found in [26,82,83,93,96].

In order to handle the fuzzy and uncertainty nature of the US images, some new techniques and approaches based on fuzzy logic, rough set and neutrosophic logic have been developed. Neutrosophic logic, a new powerful theory which handles indeterminate and uncertain characteristics in different sets, could be applied to medical image processing [148]. Some research works using fuzzy logic and fuzzy entropy have obtained good results [19,21]. Quantitative ultrasound (QUS) technique is recently used for breast cancer detection and diagnosis. A research group implemented a multi-parameter approach using QUS technique and the experimental result showed

that this approach improved the diagnostic potential of ultrasound for breast cancer detection [149]. Besides, whole breast US images which provide more breast information are available to detect breast cancer with the support of advanced scanner [151]. Also developing more accurate numeric expressions for image features is another future work to improve the performance of CADs.

Comparative studies of different preprocessing methods, segmentation methods, features and classifiers should be carried out fairly, deeply and accurately. As we stressed above, the most urgent task is to build a benchmark database of US images accessible to the public to support the comparison and evaluation of different algorithms and CAD systems. Besides, more clinical trials of the currently developed CAD systems for breast cancer should be conducted, which can provide important second opinion to physicians and can provide the feedback from the physicians to the CAD system designers.

8. Conclusions

In this paper, we reviewed CAD systems for breast cancer detection and classification using ultrasound images in the literature. The techniques developed in the four stages (preprocessing, segmentation, feature extraction and selection, classification) are summarized, and their advantages and disadvantages are discussed. Different performance evaluation metrics are studied, and the future developments and trends are also investigated. The paper will be useful for the researches in BUS imaging, computer vision, image processing and radiology.

References

[1] H. Cheng, X. Cai, X. Chen, L. Hu, X. Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey, Pattern Recognition 36 (2003) 2967–2991.

[2] H. Cheng, X. Shi, R. Min, L. Hu, X. Cai, H. Du, Approaches for automated detection and classification of masses in mammograms, Pattern Recognition 39 (4) (2006) 646–668.

[3] R.F. Chang, W.J. Wu, W.K. Moon, D.R. Chen, Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis, Ultrasound in Medicine and Biology 29 (5) (2003) 679–686.

[4] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, M.J. Thun, Cancer Statistics 2008, CA: A Cancer Journal for Clinicians 58 (2) (2008) 71–96.

[5] J. Jesneck, J. Lo, J. Baker, Breast mass lesions: computer-aided diagnosis models with mammographic and sonographic descriptors, Radiology 244 (2) (2007) 390–398.

[6] P. Shankar, C. Piccoli, J. Reid, J. Forsberg, B. Goldberg, Application of the compound probability density function for characterization of breast masses in ultrasound B scans, Physics in Medicine and Biology 50 (10) (2005) 2241–2248.

[7] K. Taylor, C. Merritt, C. Piccoli, R. Schmidt, G. Rouse, B. Fornage, E. Rubin, D. Georgian-Smith, F. Winsberg, B. Goldberg, E. Mendelson, Ultrasound as a complement to mammography and breast examination to characterize breast masses, Ultrasound in Medicine and Biology 28 (1) (2002) 19–26.

[8] H. Zhi, B. Ou, B. Luo, X. Feng, Y. Wen, H. Yang, Comparison of ultrasound elastography, mammography, and sonography in the diagnosis of solid breast lesions, Journal of Ultrasound in Medicine 26 (6) (2007) 807–815.

[9] B. Sahiner, Malignant and benign breast masses on 3D US volumetric images: effect of computer-aided diagnosis on radiologist accuracy, Radiology 242 (3) (2007) 716–724.

[10] C.M. Chen, Y.H. Chou, K.C. Han, G.S. Hung, C.M. Tiu, H.J. Chiou, S.Y. Chiou, Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks, Radiology 226 (2003) 504–514.

[11] K. Drukker, M.L. Giger, K. Horsch, M.A. Kupinski, C.J. Vyborny, E.B. Mendelson, Computerized lesion detection on breast ultrasound, Medical Physics 29 (7) (2002) 1438–1446.

[12] M.P. Andr, M. Galperin, L.K. Olson, K. Richman, S. Payrovi, P. Phan, Improving the accuracy of diagnostic breast ultrasound, Acoustical Imaging 26 (2002) 453–460.

[13] Y.L. Huang, D.R. Chen, Y.K. Liu, Breast cancer diagnosis using image retrieval for different ultrasonic systems, in: International Conference on Image Processing, vol. 5, 2004, pp. 2598–2960.

[14] M. Costantini, P. Belli, R. Lombardi, G. Franceschini, A. Mule, L. Bonomo, Characterization of solid breast masses use of the sonographic breast imaging reporting and data system lexicon, Journal of Ultrasound in Medicine 25 (5) (2006) 649–659.

[15] K.H. Hwang, J.G. Lee, J.H. Kim, H.J. Lee, K.S. Om, M. Yoon, W. Choe, Computer aided diagnosis (CAD) of breast mass on ultrasonography and scintimammography, in: Proceedings of Seventh International Workshop on Enterprise networking and computing in Healthcare Industry, HEALTHCOM 2005, 2005, pp. 187–189.

[16] Y.-L. Huang, Computer-aided diagnosis applied to 3-D US of solid breast nodules by using principal component analysis and image retrieval, in: Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005, pp. 1802–1805.

[17] J.H. Song, S.S. Venkatesh, E.F.C. Md, T.W. Cary, P.H. Md, Artificial neural network to aid differentiation of malignant and benign breast masses by ultrasound imaging, in: Proceedings of SPIE, vol. 5750, 2005, pp. 148–152.

[18] J. Segyeong, S.Y. Yoon, K.M. Woo, C.K. Hee, Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features, IEEE Transactions on Medical Imaging 23 (10) (2004) 1292–1300.

[19] Y.H. Guo, H.D. Cheng, J.H. Huang, J.W. Tian, W. Zhao, L.T. Sun, Y.X. Su, Breast ultrasound image enhancement using fuzzy logic, Ultrasound in Medicine and Biology 32 (2) (2006) 237–247.

[20] Y. Huang, K. Wang, D. Chen, Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines, Neural Computing & Applications 15 (2) (2006) 164–169.

[21] H.D.C. Xiangjun Shi, Liming Hu, Mass detection and classification in breast ultrasound images using fuzzy SVM, in: JCIS-2006 Proceedings, 2006.

[22] D.R. Chen, W.J. Kuo, R.F. Chang, W.K. Moon, C.C. Lee, Use of the bootstrap technique with small training sets for computer-aided diagnosis in breast ultrasound, Ultrasound in Medicine and Biology 28 (7) (2002) 897–902.

[23] R.F. Chang, K.C. Chang-Chien, E. Takada, J.S. Suri, W.K. Moon, J.H.K. Wu, N. Cho, Y.F. Wang, D.R. Chen, Breast density analysis in 3-D whole breast ultrasound images, in: Proceedings of the 28th IEEE EMBS Annual International Conference, 2006, pp. 2795–2798.

[24] B. Sahiner, H. Chan, G. LeCarpentier, N. Petrick, M. Roubidoux, P. Carson, Computerized characterization of solid breast masses using three-dimensional ultrasound images, in: Proceedings of SPIE, Medical Imaging 1998: Image Processing, vol. 3338, 1998, pp. 301–312.

[25] P.S. Rodrigues, A new methodology based on q-entropy for breast lesion classification in 3-D ultrasound images, in: Proceedings of the 28th IEEE EMBS Annual International Conference, 2006, pp. 1048–1051.

[26] D.R. Chen, R.F. Chang, W.M. Chen, W.K. Moon, Computer-aided diagnosis for 3-dimensional breast ultrasonography, Archives of Surgery 138 (2003) 296–302.

[27] T. Nagashima, H. Hashimoto, Ultrasound demonstration of mammographically detected microcalcifications in patients with ductal carcinoma in situ of the breast, Breast Cancer 12 (3) (2005) 216–220.

[28] W.T. Yang, In vivo demonstration of microcalcification in breast cancer using high resolution ultrasound, British Journal of Radiology 70 (835) (1997) 685–690.

[29] L. Weng, J.M. Reid, P.M. Shankar, K. Soetanto, Ultrasound speckle analysis based on the k distribution, The Journal of the Acoustical Society of America 89 (6) (1991) 2992–2995.

[30] E. Jakeman, R.J.A. Tough, Generalized k distribution: a statistical model for weak scattering, Journal of the Optical Society of America A 4 (9) (1987) 1764–1772.

[31] F. Ossant, F. Patat, M. Lebertre, M.L. Teriierooiterai, L. Pourselot, Effective density estimators based on the k distribution: interest of low and fractional order moments, Ultrasonic Imaging 20 (1998) 243–259.

[32] V. Dutt, J.F. Greenleaf, Ultrasound echo envelope analysis using a homodyned k distribution signal model, Ultrasonic Imaging 16 (1994) 265–287.

[33] K.Z. Abd-Elmoniem, A.B.M. Youssef, Y.M. Kadah, Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion, IEEE Transactions on Biomedical Engineering 49 (2002) 997–1014.

[34] C.P. Loizou, C.S. Pattichis, C.I. Christodoulou, R.S.H. Istepanian, N. Pantziaris, A. Nicolaides, "Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery, IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control 52 (2005) 1653–1669.

[35] D.T. Kuan, A.A. Sawchuk, T.C. Strand, P. Chavel, Adaptive noise smoothing filter for images with signal-dependent noise, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-7 (1985) 165–177.

[36] J.S. Lee, Digital image enhancement and noise filtering by use of local statistics, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2 (1980) 165–168.

[37] V.S. Frost, J.A. Stiles, K.S. Shanmugan, J.C. Holtzman, A model for radar images and its application to adaptive digital filtering of multiplicative noise, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-4 (1982) 157–166.

[38] A. Lopes, R. Touzi, E. Nezry, Adaptive speckle filters and scene heterogeneity, IEEE Transaction on Geoscience and Remote Sensing 28 (1990) 992–1000.

[39] V. Dutt, J.F. Greenleaf, Adaptive speckle reduction filter for log-compressed B-scan images, IEEE Transactions on Medical Imaging 15 (1996) 802–813.

[40] Y. Dong, A.K. Milne, B.C. Forster, Toward edge sharpening: a SAR speckle filtering algorithm, IEEE Transactions on Geoscience and Remote Sensing 39 (2001) 851–863.

[41] P.B. Caliope, F.N.S. Medeiros, R.C.P. Marques, R.C.S. Costa, A comparison of filters for ultrasound images, Telecommunications and Networking 3124 (2004) 1035–1040.

[42] R.C. Gonzalez, R.E. Woods, Digital Image Processing, second ed, Prentice-Hall, Englewood Cliffs, NJ, 2002.
[43] R.W. Prager, A.H. Gee, G.M. Treece, L.H. Berman, Analysis of speckle in ultrasound images using fractional order statistics and the homodyned *k*-distribution, Ultrasonics 40 (2002) 133–137.
[44] Y. Yu, J.A. Molloy, S.T. Acton, Three-dimensional speckle reducing anisotropic diffusion, in: Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers, vol. 2, 2003, pp. 1987–1991.
[45] Y.H. Guo, H.D. Cheng, J.W. Tian, Y.T. Zhang, A novel approach to speckle reduction in ultrasound imaging, Ultrasound in Medicine and Biology 35 (4) (2009) 628–640.
[46] A.N. Evans, M.S. Nixon, Mode filtering to reduce ultrasound speckle for feature extraction, IEEE Proceedings on Vision Image and Signal Processing 142 (1995) 87–94.
[47] T. Loupas, W.N. McDicken, P.L. Allan, An adaptive weighted median filter for speckle suppression in medical ultrasonic images, IEEE Transactions on Circuits and Systems 36 (1989) 129–135.
[48] R.N. Czerwinski, D.L. Jones, W.D. O'Brien Jr., Ultrasound speckle reduction by directional median filtering, in: International Conference on Image Processing, vol. 1, 1995, pp. 358–361.
[49] D. Kuan, A. Sawchuk, T. Strand, P. Chavel, Adaptive restoration of images with speckle, IEEE Transactions on Acoustics, Speech, and Signal Processing 35 (1987) 373–383.
[50] F.N.S. Medeiros, N.D.A. Mascarenhas, L.F. Costa, Evaluation of speckle noise MAP filtering algorithms applied to SAR images, International Journal of Remote Sensing 24 (2003) 5197–5218.
[51] A. Lopes, E. Nezry, R. Touzi, H. Laur, Maximum a posteriori speckle filtering and first order texture models in SAR images, in: Geoscience and Remote Sensing Symposium, 1990, pp. 2409–2412.
[52] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 629–639.
[53] Y.J. Yu, S.T. Acton, Speckle reducing anisotropic diffusion, IEEE Transactions on Image Processing 11 (2002) 1260–1270.
[54] C.Y. Xiao, Z. Su, Y.Z. Chen, A diffusion stick method for speckle suppression in ultrasonic images, Pattern Recognition Letters 25 (2004) 1867–1877.
[55] Z. Yang, M.D. Fox, Multiresolution nonhomogeneous anisotropic diffusion approach to enhance ultrasound breast tumor image legibility, in: Proceedings of SPIE, Medical Imaging 2004: Ultrasonic Imaging and Signal Processing, vol. 5373, 2004, pp. 98–107.
[56] R. Wang, J.L. Lin, D.Y. Li, T.F. Wang, Edge enhancement and filtering of medical ultrasonic images using a hybrid method, in: The First International Conference on Bioinformatics and Biomedical Engineering, 2007, pp. 876–879.
[57] T.R. Crimmins, Geometric filter for speckle reduction, Optical Engineering 25 (5) (1986) 651–654.
[58] Y. Chen, R.M. Yin, R. Flynn, S. Broschat, Aggressive region growing for speckle reduction in ultrasound images, Pattern Recognition Letters 24 (2003) 677–691.
[59] R.N. Czerwinski, D.L. Jones, W.D. O'Brien, Detection of lines and boundaries in speckle images—application to medical ultrasound, IEEE Transactions on Medical Imaging 18 (1999) 126–136.
[60] F. Abramovich, T. Sapatinas, B.W. Silverman, Wavelet thresholding via a Bayesian approach, Journal of the Royal Statistics Society 60 (1998) 725–749.
[61] A. Khare, U.S. Tiwary, Soft-thresholding for denoising of medical images—a multiresolution approach, International Journal of Wavelets Multiresolution and Information Processing 3 (2005) 477–496.
[62] D.L. Donoho, De-noising by soft-thresholding, IEEE Transactions on Information Theory 41 (1995) 613–627.
[63] D.F. Zha, T.S. Qiu, A new algorithm for shot noise removal in medical ultrasound images based on alpha-stable model, International Journal of Adaptive Control and Signal Processing 20 (2006) 251–263.
[64] H.A.M. Mohamad Forouzanfar, M. Dehghani, Speckle reduction in medical ultrasound images using a new multiscale bivariate Bayesian MMSE-based method, in: IEEE 15th SIU on Signal Processing and Communications Applications, 2007, pp. 1–4.
[65] S. Gupta, R.C. Chauhan, S.C. Saxena, Locally adaptive wavelet domain Bayesian processor for denoising medical ultrasound images using speckle modeling based on Rayleigh distribution, IEEE Proceedings on Vision Image and Signal Processing 152 (2005) 129–135.
[66] S. Gupta, R.C. Chauhan, S.C. Sexana, Wavelet-based statistical approach for speckle reduction in medical ultrasound images, Medical & Biological Engineering & Computing 42 (2004) 189–192.
[67] S. Gupta, L. Kaur, R.C. Chauhan, S.C. Saxena, A versatile technique for visual enhancement of medical ultrasound images, Digital Signal Processing 17 (2007) 542–560.
[68] S. Gupta, R.C. Chauhan, S.C. Saxena, Robust non-homomorphic approach for speckle reduction in medical ultrasound images, Medical & Biological Engineering & Computing 43 (2005) 189–195.
[69] A. Achim, A. Bezerianos, P. Tsakalides, Novel Bayesian multiscale method for speckle removal in medical ultrasound images, IEEE Transactions on Medical Imaging 20 (2001) 772–783.
[70] A. Pizurica, W. Philips, I. Lemahieu, M. Acheroy, A versatile wavelet domain noise filtration technique for medical imaging, IEEE Transactions on Medical Imaging 22 (2003) 323–331.
[71] H. Xie, L.E. Pierce, F.T. Ulaby, SAR speckle reduction using wavelet denoising and Markov random field modeling, IEEE Transactions on Geoscience and Remote Sensing 40 (2002) 2196–2212.
[72] A. Pizurica, A.M. Wink, E. Vansteenkiste, W. Philips, J. Roerdink, A review of wavelet denoising in MRI and ultrasound brain imaging, Current Medical Imaging Reviews 2 (2006) 247–260.
[73] B. Li, T.G. Zhuang, A speckle suppression method based on nonlinear threshold wavelet packet in ultrasound images, Journal of Infrared and Millimeter Waves 20 (2001) 307–310.
[74] X.H. Hao, S.K. Gao, X.R. Gao, A novel multiscale nonlinear thresholding method for ultrasonic speckle suppressing, IEEE Transactions on Medical Imaging 18 (1999) 787–794.
[75] W. Fourati, F. Kammoun, M.S. Bouhlel, Medical image denoising using wavelet thresholding, Journal of Testing and Evaluation 33 (2005) 364–369.
[76] S.G. Chang, B. Yu, M. Vetterli, Spatially adaptive wavelet thresholding with context modelling for image denoising, IEEE Transactions on Image Processing 9 (2000) 1522–1531.
[77] J.R. Sveinsson, J.A. Benediktsson, Speckle reduction and enhancement of SAR images in the wavelet domain, in: International Geoscience and Remote Sensing Symposium, IGARSS '96, 'Remote Sensing for a Sustainable Future', vol. 13, 1996, pp.725–735.
[78] Y. Rangsanseri, W. Prasongsook, Speckle reduction using Wiener filtering in wavelet domain, in: Proceedings of the Ninth International Conference on Neural Information Processing, ICONIP '02, vol. 2, 2002, pp. 792–795.
[79] Y. Yong, M.M. Croitoru, A. Bidani, J.B. Zwischenberger, J.W. Clark Jr., Nonlinear multiscale wavelet diffusion for speckle suppression and edge enhancement in ultrasound images, IEEE Transactions on Medical Imaging 25 (2006) 297–311.
[80] V. Behar, D. Adam, Z. Friedman, A new method of spatial compounding imaging, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 41 (2003) 377–384.
[81] D. Adam, S. Beilin-Nissan, Z. Friedman, V. Behar, The combined effect of spatial compounding and nonlinear filtering on the speckle reduction in ultrasound images, Ultrasonics 44 (2006) 166–181.
[82] P. Stetson, F. Graham, A. Macovski, Lesion contrast enhancement in medical ultrasound imaging, IEEE Transactions on Medical Imaging 16 (1997) 416–425.
[83] A.G.a.L.B. Robert Rohlinga, Three-dimensional spatial compounding of ultrasound images, Medical Image Analysis 1 (1997) 177–193.
[84] P. Shankar, C. Piccoli, J. Reid, J. Forsberg, B. Goldberg, Application of the compound probability density function for characterization of breast masses in ultrasound B scans, Physics in Medicine and Biology 50 (10) (2005) 2241–2248.
[85] X.J. Shi, H.D. Cheng, A simple and effective histogram equalization approach to image enhancement, Digital Signal Processing 14 (2004) 158–170.
[86] J. Awad, T.K. Abdel-Galil, M.M.A. Salama, H. Tizhoosh, A. Fenster, K. Rizkalla, D.B. Downey, Prostate's boundary detection in transrectal ultrasound images using scanning technique, IEEE Conference on Electrical and Computer Engineering 2 (2003) 1199–1202.
[87] H.D. Cheng, X.H. Jiang, Y. Sun, J.L. Wang, Color image segmentation: advances and prospects, Pattern Recognition 34 (12) (2001) 2259–2281.
[88] E. Littmann, H. Ritter, Adaptive color segmentation—a comparison of neural and statistical methods, IEEE Transactions on Neural Networks 8 (1) (1997) 175–185.
[89] J. Segyeong, K.M. Woo, C.K. Hee, Computer-aided diagnosis of solid breast nodules on ultrasound with digital image processing and artificial neural network, in: 26th Annual International Conference of the Engineering in Medicine and Biology Society, vol. 1, 2004, pp. 1397–1400.
[90] K. Horsch, M.L. Giger, L.A. Venta, C.J. Vyborny, Computerized diagnosis of breast lesions on ultrasound, Medical Physics 29 (2) (2002) 157–164.
[91] K. Horsch, M.L. Giger, L.A. Venta, C.J. Vyborny, Automatic segmentation of breast lesions on ultrasound, Medical Physics 28 (8) (2001) 1652–1659.
[92] D.R. Chen, R.F. Chang, Y.L. Huang, Computer-aided diagnosis applied to us of solid breast nodules by using neural networks, Radiology 213 (1999) 407–412.
[93] R.F. Chang, W.J. Wu, W.K. Moon, W.M. Chen, W. Lee, D.R. Chen, Segmentation of breast tumor in three-dimensional ultrasound images using three-dimensional discrete active contour model, Ultrasound in Medicine and Biology 29 (2003) 1571–1581.
[94] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics 9 (1) (1979) 62–66.
[95] M. Kass, A. Witkin, D. Terzopoulos, Snakes—active contour models, International Journal of Computer Vision 1 (4) (1987) 321–331.
[96] D.R. Chen, R.F. Chang, W.J. Wu, W.K. Moon, W.L. Wu, 3-D breast ultrasound segmentation using active contour model, Ultrasound in Medicine and Biology 29 (7) (2003) 1017–1026.
[97] A. Madabhushi, D.N. Metaxas, Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions, IEEE Transactions on Medical Imaging 22 (2) (2003) 55–169.
[98] A. Sarti, C. Corsi, E. Mazzini, C. Lamberti, Maximum likelihood segmentation of ultrasound images with Rayleigh distribution, IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control 52 (6) (2005) 947–960.
[99] C.M. Chen, H.H. Lu, Y.C. Lin, An early vision-based snake model for ultrasound image segmentation, Ultrasound in Medicine and Biology 26 (2) (2000) 273–285.
[100] R.F. Chang, W.-J. Wu, C.-C. Tseng, D.-R. Chen, W. Moon, 3-D snake for US in margin evaluation for malignant breast tumor excision using mammotome, IEEE Transactions on Information Technology in Biomedicine 7 (3) (2003) 197–201.
[101] H.D. Cheng, L. Hu, J. Tian, L. Sun, A novel Markov random field segmentation algorithm and its application to breast ultrasound image analysis, in: The

Sixth International Conference on Computer Vision, Pattern Recognition and Image Processing, Salt Lake City, USA, 2005.
[102] J.A. Noble, D. Boukerroui, Ultrasound image segmentation: a survey, IEEE Transactions on Medical Imaging 25 (8) (2006) 987–1010.
[103] G.F. Xiao, M. Brady, J.A. Noble, Y.Y. Zhang, Segmentation of ultrasound B-mode images with intensity inhomogeneity correction, IEEE Transactions on Medical Imaging 21 (1) (2002) 48–57.
[104] D. Boukerroui, O. Basset, N. Gu, A. Baskurt, Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation, European Journal of Ultrasound 8 (1998) 135–144.
[105] D. Boukerroui, A. Baskurt, J.A. Noble, O. Basset, Segmentation of ultrasound images-multiresolution 2D and 3D algorithm based on global and local statistics, Pattern Recognition Letters 24 (4–5) (2003) 779–790.
[106] L.A. Christopher, E.J. Delp, C.R. Meyer, P.L. Carson, 3-D Bayesian ultrasound breast image segmentation using the EM/MPM algorithm, in: Proceedings of the 2002 IEEE International Symposium on Biomedical Imaging, 2002, pp. 86–89.
[107] D.R. Chen, R.F. Chang, W.J. Wu, W.K. Moon, W.L. Wu, 3-D breast ultrasound segmentation using active contour model, Ultrasound in Medicine and Biology 29 (2003) 1017–1026.
[108] E.A. Ashton, K.J. Parker, Multiple resolution Bayesian segmentation of ultrasound images, Ultrasonic Imaging 17 (4) (1995) 291–304.
[109] D.R. Chen, R.F. Chang, W.J. Kuo, M.C. Chen, Y.L. Huang, Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks, Ultrasound in Medicine and Biology 28 (10) (2002) 1301–1310.
[110] Y.L. Huang, D.R. Chen, Watershed segmentation for breast tumor in 2-D sonography, Ultrasound in Medicine and Biology 30 (2004) 625–632.
[111] V.S. Cherkassky, F. Mulier, Learning from Data: Concepts, Theory, and Methods, Wiley, New York, NY, USA, 1998.
[112] H. Li, Y. Wang, K.J.R. Liu, S.B. Lo, M.T. Freedman, Computerized radiographic mass detection—part II: lesion site selection by morphological enhancement and contextual segmentation, IEEE Transactions on Medical Imaging 20 (4) (2001) 302–313.
[113] K. Drukker, C.A. Sennett, M.L. Giger, The effect of image quality on the appearance of lesions on breast ultrasound: implications for CADx, in: Proceedings of SPIE, Medical Imaging 2007: Computer-Aided Diagnosis, vol. 6514, 2007, p. 65141E.
[114] K. Horsch, A.F. Ceballos, M.L. Giger, I.R. Bonta, Z. Huo, C.J. Vyborny, E.R. Hendrick, L. Lan, Optimizing feature selection across a multimodality database in computerized classification of breast lesions, Progress in biomedical optics and imaging 3 (22) (2002) 986–992.
[115] J.W. Tian, L.T. Sun, Y.H. Guo, H.D. Cheng, Y.T. Zhang, Computerized-aid diagnosis of breast mass using ultrasound image, Medical Physics 34 (2007) 3158–3164.
[116] K. Mogatadakala, K. Donohue, C. Piccoli, F. Forsberg, Detection of breast lesion regions in ultrasound images using wavelets and order statistics, Medical Physics 33 (4) (2006) 840–849.
[117] K. Drukker, M. Giger, E. Mendelson, Computerized analysis of shadowing on breast ultrasound for improved lesion detection, Medical Physics 30 (7) (2003) 1833–1842.
[118] P. Shankar, The use of the compound probability density function in ultrasonic tissue characterization, Physics in Medicine and Biology 49 (6) (2004) 1007–1015.
[119] A. Stavros, D. Thickman, C. Rapp, M. Dennis, S. Parker, G. Sisney, Solid breast nodules—use of sonography to distinguish benign and malignant lesions, Radiology 196 (1) (1995) 123–134.
[120] P. Shankar, V. Dumane, T. George, C. Piccoli, J. Reid, F. Forsberg, B. Goldberg, Classification of breast masses in ultrasonic B scans using Nakagami and *K* distribution, Physics in Medicine and Biology 48 (14) (2003) 2229–2240.
[121] S. Chen, Y. Cheung, C. Su, M. Chen, T. Hwang, S. Hsueh, Analysis of sonographic features for the differentiation of benign and malignant breast tumors of different sizes, Ultrasound in Medicine and Biology 23 (2) (2004) 188–193.
[122] S. Gefen, O. Tretiak, C. Piccoli, K. Donohue, A. Petropulu, P. Shankar, V. Dumane, L. Huang, M. Kutay, V. Genis, F. Forsberg, J. Reid, B. Goldberg, ROC analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis, IEEE Tractions on Medical Imaging 22 (2) (2003) 170–177.
[123] C. Sehgal, T. Cary, S. Kangas, S. Weinstein, S. Schultz, P. Arger, E. Conant, Computer-based margin analysis of breast sonography for differentiating malignant and benign masses, Journal of Ultrasound in Medicine 23 (9) (2004) 1201–1209.
[124] P. Shankar, V. Dumane, C. Piccoli, J. Reid, F. Forsberg, B. Goldberg, Computer-aided classification of breast masses in ultrasonic B-scans using a multiparameter approach, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 50 (8) (2003) 1002–1009.
[125] D. Chen, R. Chang, Y. Huang, Computer-aided diagnosis applied to US of solid breast nodules by using neural networks, Radiology 213 (2) (1999) 407–412.
[126] D.R. Chen, R.F. Chang, Y.L. Huang, Breast cancer diagnosis using self-organizing map for sonography, Ultrasound in Medicine and Biology 26 (3) (2000) 405–411.
[127] D. Chen, R. Chang, Y. Huang, Y. Chou, C. Tiu, P. Tsai, Texture analysis of breast tumors on sonograms, Seminars in Ultrasound CT and MRI 21 (4) (2000) 308–316.
[128] P.C. Bhat, H.B. Prosper, Bayesian Neural Networks, in: Statistical Problems in Particle Physics, Astrophysics and Cosmology: Proceedings of PHYSTAT05, 2006, pp. 151–155.
[129] K. Drukker, M.L. Giger, C.J. Vyborny, E.B. Mendelson, Computerized detection and classification of cancer on breast ultrasound 1, Academic Radiology 11 (5) (2004) 526–535.
[130] K. Drukker, D.C. Edwards, M.L. Giger, R.M. Nishikawa, C.E. Metz, Computerized detection and 3-way classification of breast lesions on ultrasound images, in: Proceedings of SPIE, Medical Imaging 2004: Image Processing, vol. 5370, 2004, pp. 1034–1041.
[131] J.R. Quinlan, C4. 5: Programs for Machine Learning, vol. 95, Morgan Kaufmann, Los Altos, CA, 1993.
[132] W. Kuo, R. Chang, D. Chen, C. Lee, Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images, Breast Cancer Research and Treatment 66 (1) (2001) 51–57.
[133] Y.L. Huang, D.R. Chen, Support vector machines in sonography application to decision making in the diagnosis of breast cancer, Clinical Imaging 29 (3) (2005) 179–184.
[134] W. Kuo, R. Chang, C. Lee, W. Moon, D. Chen, Retrieval technique for the diagnosis of solid breast tumors on sonogram, Ultrasound in Medicine and Biology 28 (7) (2002) 903–909.
[135] W. Berg, J. Blume, J. Cormack, E. Mendelson, Operator dependence of physician-performed whole-breast US: lesion detection and characterization, Radiology 241 (2) (2006) 355–365.
[136] N. Cho, W. Moon, J. Cha, S. Kim, B. Han, E. Kim, M. Kim, S. Chung, H. Choi, J. Im, Differentiating benign from malignant solid breast masses: comparison of two-dimensional and three-dimensional US, Radiology 240 (1) (2006) 26–32.
[137] C.C. Chen, J.S. Daponte, M.D. Fox, Fractal feature analysis and classification in medical imaging, IEEE Transactions on Medical Imaging 8 (2) (1989) 133–142.
[138] A. Hong, E. Rosen, M. Soo, J. Baker, BI-RADS for sonography: positive and negative predictive values of sonographic features, American Journal of Roentgenology 184 (4) (2005) 1260–1265.
[139] B. Garra, B. Krasner, S. Horii, S. Ascher, S. Mun, R. Zeman, Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis, Ultrasonic Imaging 15 (4) (1993) 267–285.
[140] M. Giger, Y. Yuan, H. Li, K. Drukker, W. Chen, L. Lan, K. Ho, Progress in breast CADx, in: Biomedical Imaging: Fourth IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007, pp. 508–511.
[141] K. Drukker, M. Giger, C. Metz, Robustness of computerized lesion detection and classification scheme across different breast US platforms, Radiology 238 (1) (2006) 834–840.
[142] M. Mainiero, A. Goldkamp, E. Lazarus, L. Livingston, S. Koelliker, B. Schepps, W. Mayo-Smith, Characterization of breast masses with sonography—can biopsy of some solid masses be deferred?, Journal of Ultrasound in Medicine 24 (2) (2005) 161–167.
[143] R. Paulinelli, R. Freitas Jr., M. Moreira, V. de Moraes, J. Bernardes Jr., C. Vidal, A. Ruiz, M. Lucato, Risk of malignancy in solid breast nodules according to their sonographic features, Journal of Ultrasound in Medicine 24 (5) (2005) 635–641.
[144] M. Kutay, A. Petropulu, C. Piccoli, Breast tissue characterization based on modeling of ultrasonic echoes using the power-law shot noise model, Pattern Recognition Letters 24 (4–5) (2003) 741–756.
[145] M.A. Kupinski, D.C. Edwards, M.L. Giger, C.E. Metz, Ideal observer approximation using Bayesian classification neural networks, IEEE Transactions on Medical Imaging 20 (9) (2001) 886–899.
[146] American College of Radiology. ACR Standards 2000–2001, American College of Radiology, Reston, VA, 2000.
[147] W. Yang, P. Dempsey, Diagnostic breast ultrasound: current status and future directions, Radiologic Clinics of North America 45 (2007) 845–861.
[148] H.D. Cheng, Y. Guo, A new neutrosophic approach to image thresholding, New Mathematics and Natural Computation 4 (2008) 291–308.
[149] M. Oelze, W. O'Brien, J. Zachary, 11B-4 quantitative ultrasound assessment of breast cancer using a multiparameter approach, in: Ultrasonics Symposium, 2007, pp. 981–984.
[150] K. Zheng, T.F. Wang, J.L. Lin, D.Y. Li, Recognition of breast ultrasound images using a hybrid method, in: IEEE/ICME International Conference on Complex Medical Engineering, 2007, pp. 640–643.
[151] Y. Ikedo, D. Fukuoka, T. Hara, Development of a fully automatic scheme for detection of masses in whole breast ultrasound images, Medical Physics 34 (2007) 4378–4388.
[152] W. Shen, R. Chang, W. Moon, Y. Chou, C. Huang, Breast ultrasound computer-aided diagnosis using BI-RADS features, Academic Radiology 14 (2007) 928–939.
[153] B. Anderson, R. Shyyan, A. Eniu, R. Smith, C. Yip, Breast cancer in limited-resource countries: an overview of the breast health global initiative 2005 guidelines, The Breast Journal 12 (2006) S3–15.
[154] J.M. Park, W.J. Song, W.A. Pearlman, Speckle filtering of SAR images based on adaptive windowing, IEEE Proceedings on Visual Image Processing 146 (1999) 191–197.
[155] L. Gagnon, A. Jouan, Speckle filtering of SAR images—a comparative study between complex-wavelet-based and standard filters, in: Proceedings of SPIE: Conference on Wavelet Applications in Signal and Image Processing, vol. 3169, 1997.
[156] O. Uncu, I.B. Turksen, A novel feature selection approach: combining feature wrappers and filters, Information Sciences 177 (2007) 449–466.
[157] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artificial Intelligence 97 (1/2) (1997) 273–324.
[158] U. Fayyad, K. Irani, Multi-interval discretization of continuous-valued attributes for classification learning, in: Proceedings 13th International Joint Conference on Artificial Intelligence, 1993, pp. 1022–1027.

[159] R.C. Holte, Very simple classification rules perform well on most commonly used datasets, Machine Learning 11 (1993) 63–90.
[160] P.A. Lachenbruch, Discriminant Analysis, Hafner, New York, 1975.
[161] J.C. Rice, Logistic regression: an introduction, in: B. Thompson (Ed.), Advances in Social Science Methodology, vol. 3, JAI Press, Greenwich, CT, 1994, pp. 191–245.
[162] R.L. Gorsuch, Factor Analysis, Erlbaum, Hillsdale, NJ, 1983.

About the Author—HENG-DA CHENG received Ph.D. degree in Electrical Engineering from Purdue University (Supervisor: Prof. K.S. Fu), West Lafayette, IN, 1985. Now, he is a Full Professor, Department of Computer Science, and an Adjunct Full Professor, Department of Electrical Engineering, Utah State University, Logan, Utah. Dr. Cheng is an Adjunct Professor and Doctorial Supervisor of Harbin Institute of Technology. He is also a Guest Professor of the Institute of Remote Sensing Application, Chinese Academy of Sciences, a Guest Professor of Wuhan University, a Guest Professor of Shantou University, and a Visiting Professor of Northern Jiaotong University.
Dr. Cheng has published more than 250 technical papers and is the Co-editor of the book, Pattern Recognition: Algorithms, Architectures and Applications (World Scientific Publishing Co., 1991). His research interests include image processing, pattern recognition, computer vision, artificial intelligence, medical information processing, fuzzy logic, genetic algorithms, neural networks, parallel processing, parallel algorithms, and VLSI architectures.
Dr. Cheng is the General Chair of the 11th Joint Conference on Information Sciences (JCIS 2008), was the General Chair of the 10th Joint Conference on Information Sciences (JCIS 2007),was the General Chair of the Ninth Joint Conference on Information Sciences (JCIS 2006), was the General Chair of the Eighth Joint Conference on Information Sciences (JCIS 2005), General Chair and Program Chair of the Sixth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2007), General Chair and Program Chair of the Sixth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2006), General Chair and Program Chair of the Sixth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2005), and was the General Chair and Program Chair of the Fifth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2003), and was the General Chair and Program Chair of the Fourth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2002), the General Chair and Program Chair of the Third International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2000), the General Chair and Program Chair of the First International Workshop on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 1998), and the Program Co-Chairman of Vision Interface 1990. He served as program committee member and Session Chair for many conferences, and as reviewer for many scientific journals and conferences.
Dr. Cheng has been listed in Who's Who in the World, Who's Who in America, Who's Who in Communications and Media, Who's Who in Finance and Industry, Who's who in Science and Engineering, Men of Achievement, 2000 Notable American Men, International Leaders in Achievement, 500 Leaders of Influence, International Dictionary of Distinguished Leadership, etc.
Dr. Cheng is also an Associate Editor of Pattern Recognition, an Associate Editor of Information Sciences and Associate Editor of New Mathematics and Natural Computation.

About the Author—JUAN SHAN received her B.S. degree from Computer Science School of Harbin Institute of Technology (HIT), Harbin, China, in 2004. She is a Ph.D. student in the Department of Computer Science, Utah State University from 2004 to now. Her research interests include pattern recognition, computer vision and medical image processing.

About the Author—WEN JU received her B.S. degree in Computer Science from University of Science and Technology of China in 2002. Currently, she is a Ph.D. student in the Department of Computer Science, Utah State University. Her research interests include: pattern recognition, computer vision and image processing.

About the Author—YANHUI GUO was born in 1979, PR China. He received B.S. degree in Automatic Control from Zhengzhou University, PR China, in 1999, and the M.S. degree in Pattern Recognition and Intelligence System from Harbin Institute of Technology, Harbin, Heilongjiang Province, PR China, in 2002. He is currently working toward the Ph.D. degree in the Department of Computer Science, Utah State University. His research interests include image processing, pattern recognition, medical imaging processing, fuzzy logic, and neural networks.

About the Author—LING ZHANG received the Bachelor of Engineering degree from the Department of Computer Science, Shandong University of Technology, Jinan, China, in 1993 and the Master of Science degree from the Department of Computer Science, Shandong University of Technology, Jinan, China, in 1997. From 1993 to now, she worked as a teacher in Shandong University. Now she is a Ph.D. candidate in School of Computer Science and Technology in Shandong University. Her research interests include artificial intelligence, nature language processing, and database.