Computer-Aided Breast Cancer Detection Using Mammograms: A Review

Karthikeyan Ganesan, U. Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, K. Thomas Abraham, and Kwan-Hoong Ng

Methodological Review

***Abstract*—The American Cancer Society (ACS) recommends women aged 40 and above to have a mammogram every year and calls it a gold standard for breast cancer detection. Early detection of breast cancer can improve survival rates to a great extent. Inter-observer and intra-observer errors occur frequently in analysis of medical images, given the high variability between interpretations of different radiologists. Also, the sensitivity of mammographic screening varies with image quality and expertise of the radiologist. So, there is no golden standard for the screening process. To offset this variability and to standardize the diagnostic procedures, efforts are being made to develop automated techniques for diagnosis and grading of breast cancer images. A few papers have documented the general trend of computer-aided diagnosis of breast cancer, making a broad study of the several techniques involved. But, there is no definitive documentation focusing on the mathematical techniques used in breast cancer detection. This review aims at providing an overview about recent advances and developments in the field of Computer-Aided Diagnosis (CAD) of breast cancer using mammograms, specifically focusing on the mathematical aspects of the same, aiming to act as a mathematical primer for intermediates and experts in the field.**

***Index Terms*—Breast cancer, classifiers, computer-aided diagnosis (CAD), digital mammography, feature extraction techniques.**

I. Introduction

CANCER refers to the uncontrolled multiplication of a group of cells in a particular location of the body [1]. A group of rapidly dividing cells may form a lump, microcalcifications or architectural distortions which are usually referred to as tumors [2]. Breast cancer is any form of malignant tumor which develops from breast cells [3]. Breast cancers are traditionally known to be one of the major causes of death among women [4]. Mortality rates due to breast cancer have been reducing due to better diagnostic facilities and effective treatments [5]. One of the leading methods for diagnosing breast cancer is screening mammography. This method involves X-ray imaging of the breast. Screening mammography examinations are performed on asymptomatic women to detect early, clinically unsuspected breast cancer [2]. The need for early detection of breast cancer is highlighted by the fact that incidence rates for breast cancer is one of the highest among all cancers according to the American Cancer Society which quotes a morbidity of 230 000 and a mortality of 40 000 according to the latest figures gathered for the American population [6].

Important signs to look for in the case of breast cancer are clusters of microcalcifications, masses and architectural distortions [2]. Following the results of screening mammography, a follow up study is made for patients according to the level of suspicion of the abnormality. This stage is referred to as diagnostic mammography. Both screening mammography and diagnostic mammography are performed by radiologists who visually inspect the mammograms.

Early detection of breast cancer through screening and diagnostic mammography increases breast cancer treatment options and survival rates. Unfortunately, due to the human factor involved in the screening process, detection of suspicious abnormalities is prone to a high degree of error. Studies have shown that radiologists have an error rate between 10%–30% for detection of cancer in screening studies [7], [8]. Misinterpretation of breast cancer signs result in 52% of the errors and 43% of the errors are caused due to overlooking signs in abnormal scans [8]. As a result of this error rate, biopsies are frequently performed on benign lesions, resulting in unwarranted expenditure and anxiety for the patient involved. The cost associated with errors due to misclassification of mammograms is considerable. This is because of the fact that false negatives are a huge problem in screening mammography as early detection can reduce treatment cost, time and effectiveness to a great extent. False negatives affect all three parameters as early detection is not an option with an incorrect diagnosis. A study by [9] found that double reading of screening mammograms provided greater sensitivity than single reading without increasing recall rates. But, manpower is a major drawback with this approach. The number of radiologists required for double reading of mammograms will be huge. As a result, many nations might not be able to meet the manpower requirements for such an approach.

A major reason for these errors is due to the fact that radiologists depend on visual inspection. During manual screening of a large number of mammograms, radiologists may get easily worn out, missing out vital clues while studying the scans. To

Manuscript received June 13, 2012; revised October 22, 2012; accepted November 12, 2012. Date of publication December 11, 2012; date of current version March 29, 2013.

K. Ganesan, C. K. Chua, and L. C. Min are with the Department of Electrical and Computer Engineering, Ngee Ann Polytechnic, 599489 Singapore.

U. R. Acharya is with the Department of Electrical and Computer Engineering, Ngee Ann Polytechnic, 599489 Singapore, and also with the Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia.

K. T. Abraham is with SATACommHealth, Singapore.

K.-H. Ng is with the Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia.

Digital Object Identifier 10.1109/RBME.2012.2232289

offset these effects, tremendous effort is being made to automate the process of mammographic screening. Automated screening of mammograms or computer-aided diagnosis (CAD) of breast cancer is a vast field of research. Sampat *et al.*, and Rangayyan *et al.*, provide an extensive review on different stages of a CAD methodology for breast cancer.

Classifier systems have been widely used in medical diagnosis [3]. Though the most important factor in diagnosis is evaluation of data taken from patients by human experts, expert systems and various artificial intelligence techniques for classification aid radiologists to a great extent [3]. As yet, there is no definitive literature which focuses on an elaborate discussion on the feature extraction, feature selection and classification methodologies used in breast cancer detection. The current study aims at filling this gap by documenting developments in that aspect.

II. Computer-Aided Diagnosis System Pipeline

Any computer-aided diagnosis system is based on artificial intelligence (AI) techniques. The pipeline used in a CAD system for breast cancer detection is similar to any other AI-based system and consists of preprocessing, breast region segmentation, feature extraction and classification. A major difference between computer-aided detection of breast cancer and other AI-based technologies is that breast cancer detection using CAD systems requires human intervention for interpreting the final results [10], [13], [16].

Preprocessing of mammograms is done to improve the contrast of mammograms which will be helpful in further stages of the detection pipeline. This step also includes denoising of the images. Segmenting the breast region from pectoral muscle and surrounding regions is carried out in order to make it easier to extract the suspicious tissues from breast segments. Feature extraction and classification steps are similar to other AI and pattern recognition systems with not much of a difference between commonly used methods [17], [20], [24]. Fig 2 shows a standard CAD system pipeline.

III. Preprocessing of Mammograms

Denoising and enhancement of mammograms are very important for both the manual inspection stage and for the computer-aided second reading stage. Mammograms do not provide a very good contrast between normal glandular and malignant tissues. This is because X-ray attenuation between these two tissues does not vary much, especially in younger women with denser breast tissues [11]. This fact is seen quite evidently in the case of smaller malignancies where it becomes more difficult for the radiologist to manually delineate between normal and cancerous tissues. This issue can be understood better by knowing the linear absorption coefficients of various tissues which define the image contrast. Image contrast obtained through the linear absorption coefficients is determined by the Beer-Lambert law $(I = I_0 e^{-\alpha x})$ which relates the intensity of the incident electromagnetic (EM) wave (I_0) with that of the transmitted EM wave (I), the attenuation coefficient of the material (α) and the length of the material (x) through which it is transmitted.

Contrast enhancement procedures are well known for enhancement of mammograms. Radiological images contain random fluctuations due to the statistics of X-ray quantum absorption. This noise makes the detection of small and subtle structures more difficult [11]. It has been observed that noise tends to increase with pixel intensity in images where local contrast and image intensity are interdependent [12]. This is indeed the case in mammograms. A solution for this problem is proposed by [12] in the form of a noise equalization procedure for obtaining images where local contrast is approximately equal at all image intensities. Reference [14] improved this technique to obtain better noise estimates which is further enhanced by [11]. In this method, a neighborhood $\wp$ of an image location (x, y) is considered. The local contrast of this neighborhood is estimated as

$$c(x, y) = f(x, y) - \text{median}_{\wp}(x, y) \tag{1}$$

where $c(x, y)$ is the estimated local contrast, $f(x, y)$ is the image gray level at (x, y) and $median_{\wp}(x, y)$ is the median gray level within the neighborhood $\wp$ of (x, y). Equation (1) can be equated to a high-pass spatial filter. The local contrast provides a measure of the high-frequency image noise. The noise associated with each image gray level I can be measured by the local contrast standard deviation $\sigma_c(I) \equiv \sigma\{c(I)\}$. The contrast enhancement function is then defined as

$$f_{ceq}(I_i) = \begin{cases} \frac{\sigma(I_i)}{\sigma_c(I_i)}, & \text{if } \sigma_c(I_i) > 0 \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

While executing the contrast enhancement function, the grayscale is divided into several overlapping bins $i = 1, \ldots, N$ where N is the number of bins. Interpolation of the estimated $f_{ceq}(I_i)$ values provides an estimate of the function $f_{ceq}(I)$ for all image intensities I. This method of contrast enhancement is a nonlinear gray level rescaling technique. This transfer function $f_{ceq}(I_i)$ is then normalized such that the total value adds up to 1, the same way it is done in the case of a probability density function.

A Bayesian estimator-based discriminator for image enhancement by separating image and noise by assuming the noise as *a priori* Gaussian additive noise was proposed by [15]. This is a semi-blind noise removal algorithm which is based on a steerable wavelet pyramid. A spatially adaptive statistical model for image denoising was proposed by [18]. In this method, wavelet coefficients are modelled as Gaussian random variables with high local correlation. Posteriori probability rules are applied to estimate the original coefficients from noisy observations. An experiment using two image measurements, namely the local regularity and geometric constraints of the image, was developed by [19] to filter out noise. The two measurements were then combined using a Bayesian probabilistic framework and executed using a Markov random field model.

Apart from intensity and probabilistic-based methods, wavelet-based techniques have also been used for image enhancement in mammograms [21]–[23]. The idea of processing images at several scales arises from the concept of short-time Fourier Transform and has been adapted well in wavelet techniques for both denoising and feature extraction from

mammograms. Microcalcification enhancement methods based both on discrete wavelet transforms (DWT) [21] and continuous wavelet transforms (CWT) [23] are widely used with considerable rates of success. Conventional contrast enhancement techniques can be found in [25]. Since wavelet-based denoising and contrast enhancement methods are not discussed in a concise manner in literature, we shall perform a short review of the basics.

Image denoising based on wavelet transforms need wavelets and smoothing functions. A good primer to this can be found in [11]. In this technique, a smoothing function $\phi(x,y)$ and two wavelets $\psi^i(x,y)$ are considered

$$\phi_s(x,y) = \frac{1}{s^2}\phi\left(\frac{x}{s} \quad \frac{y}{s}\right) \tag{3}$$

$$\psi^i_s(x,y) = \frac{1}{s^2}\psi^i\left(\frac{x}{s} \quad \frac{y}{s}\right), \quad i = 1,2 \tag{4}$$

and the dyadic wavelet transform $f(x,y)$, at a scale $s = 2^j$, has two detail components given by

$$W^i_{2^j} f(x,y) = \left(f * \psi^i_{2^j}\right)(x,y), \quad i = 1,2 \tag{5}$$

and one low-pass component, given by

$$S_{2^j} f(x,y) = (f * \phi_{2^j})(x,y). \tag{6}$$

The coefficients $W^1_{2^j} f(x,y)$ and $W^2_{2^j} f(x,y)$ represent the details in the x and y directions, respectively. The representation of an image according to a family of wavelet functions is discussed in [27]. According to this discussion, a function $\psi \in L^2(\Re^2)$, which is the mother wavelet, is chosen. This mother wavelet must satisfy the admissibility condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\left|\hat{\psi}(\omega_x,\omega_y)\right|^2}{\omega_x\omega_y} d\omega_x d\omega_y < \infty \tag{7}$$

where $\hat{\psi}(\omega_x,\omega_y)$ is the Fourier transform of ψ and ω_x, and ω_y are the spatial frequency components along the x,y axis. This condition is a regularity constraint so that ψ has a zero mean value and local oscillation quickly decaying to zero. From the above admissibility condition, the discrete wavelet transform can be defined as a decomposition on the following family of functions:

$$\psi_{n,m,l} = \left[\frac{1}{s^l_0}\psi\left(\frac{x - nu_0 s^l_0}{s^l_0}, \frac{y - mu_0 s^l_0}{s^l_0}\right)\right]_{n.m.l} \tag{8}$$

where $(n,m,l) \in Z^3$, s^l_0 is the discrete scaling based on a dilation step $s_0 > 1$ and u_0 is a translational step. With these basic definitions forming a wavelet decomposition system for denoising, from [11], we can model a function $p(x|\text{noise})$ that models the coefficient distribution for a Gaussian noise with a standard deviation σ_{noise}

$$p(x|\text{noise}) = \frac{1}{\sigma_{\text{noise}}\sqrt{2\pi}} e^{-x^2/2\sigma^2_{\text{noise}}}. \tag{9}$$

Also, the distribution of noise-free wavelet coefficients $W_f[n,m]$ is approximated by a generalized Laplacian probability density function, given by

$$p(x|\text{edge}) = \frac{e^{-|\frac{x}{\alpha}|^\beta}}{C(\alpha,\beta)}, \quad \text{for } -\infty < x < \infty,\ \beta > 0,\ \alpha > 0 \tag{10}$$

where

$$C(\alpha,\beta) = 2\frac{\alpha}{\beta}\Gamma\left(\frac{1}{\beta}\right) \tag{11}$$

where Γ is the gamma function. With the values of $p(x|\text{noise})$ and $p(x|\text{edge})$, a function $g(x)$ called "shrinkage" is calculated

$$g(x) = \frac{(1-\omega)p(x|\text{edge})}{(1-\omega)p(x|\text{edge}) + \omega p(x|\text{noise})}. \tag{12}$$

This shrinkage function is decomposed to several levels, from which denoising and enhancement of the images is done. The advantage of this method is that low contrast features will be more enhanced than high contrast image features and artifacts will not be introduced. From [28], we can see that contrast enhancement could be achieved by changing the gradient magnitude using the wavelet coefficient in each subimage. This arises from the reasoning that the relative change of image intensity at each point is expressed by the image gradient magnitude which is highly correlated with the contrast [30]. The magnitude and phase of the wavelets of the form seen in (5) can be represented as $M_{2^j}(n,m)$ and $A_{x^j}(n,m)$, respectively. These coefficients can be transformed into polar coordinates as follows:

$$M_{2^j}(n,m) = \sqrt{W_{2^j}(n,m) + W_{2^j}(n,m)} \tag{13}$$

$$A_{2^j}(n,m) = \arctan\left(\frac{W^2_{2^j}(n,m)}{W^1_{2^j}(n,m)}\right). \tag{14}$$

Linear stretching as a technique for contrast enhancement with linear or nonlinear mapping of wavelet coefficients is discussed in [31]. Reference [28] discusses a way to enhance coarse information consisting of sizeable structures in the mammograms. This enhancement is obtained from the second and third level decomposed wavelet images along with a background level approximation which is extracted from the fourth decomposition level.

Local range modification is a technique developed by [32] based on linear stretching. This technique makes use of the straight line function $y = ax + b$ for image enhancement. In this equation, y is the enhanced image, x is the original grayscale image and a, b are parameters depending on the local contrast,

computed by an interpolation procedure using overlapping image blocks. The technique processes the whole image twice, with the first pass calculating the local parameters and the second performing the contrast enhancement. Interpolation of neighboring grid points for a four point system to estimate the local maximum and minimum pixel values is obtained by

$$\max = \left[\frac{s_y}{s}M_4 + \left(\frac{s-s_y}{s}\right)M_1\right]\left(\frac{s-s_x}{s}\right) + \left[\frac{s_y}{s}M_5 + \left(\frac{s-s_y}{s}\right)M_2\right]\left(\frac{s_x}{s}\right) \quad (15)$$

where s is the size of the block, s_x and s_y are the horizontal and vertical distances of the examined point, respectively, from the grid points, and M_i are the intensity values of the surrounding grid points. With this, the output value of each pixel with coordinates $[m, n]$ is calculated by linear stretching

$$y[m,n] = \frac{L-1}{(\max - \min)}(x[m,n] = \min) \quad (16)$$

where L is the number of grayscales and $\max$ and $\min$ are the margins of the local input grayscale range, respectively. It is to be noted that two major methods of enhancements are popularly used: statistical methods based on intensity difference between the microcalcifications and its surroundings and wavelet-based decomposition methods. While statistical methods tend to make use of the fact that microcalcifications tend to be brighter than their surroundings, wavelet-based decomposition methods are based on the difference in frequency content of the bright microcalcification spots from their surrounding background [33].

Reference [34] discusses the application of Iris filter to mammograms. Iris filters are adaptive filters which are effective in enhancing approximately rounded opacities, no matter what their contrasts might be. In this technique, the filter uses the orientation map of gradient vectors instead of applying it directly to the image. Gradients are generated in two orthogonal directions. The convergence index of the gradient vector g at an arbitrary pixel Q_i, towards the pixel of interest P, is given by

$$f(Q_i) = \begin{cases} \cos\theta, & |g| \neq 0 \\ 0 & |g| \neq 0 \end{cases} \quad (17)$$

where θ is the orientation of the gradient vector g at Q_i with respect to the ith half line. The convergence degree of gradient vectors on the line PQ_i, C_i can be defined as the average of convergence indexes over the length PQ_i as

$$C_i = \frac{\int_P^{Q_i} f(Q)dQ}{\overline{PQ_i}}. \quad (18)$$

With the maximum convergence degree being given by C_{im}, the output of the iris filter at the pixel of interest (x, y) can be defined as

$$C(x,y) = \frac{1}{N}\sum_{i=0}^{N-1} C_{im}. \quad (19)$$

Apart from these methods, [35] lists and reviews other algorithms which are widely used for image enhancement. Some of these algorithms are: manual intensity windowing (MIW), histogram-based intensity windowing (HIW), mixture-model intensity windowing (MMIW), contrast-limited adaptive histogram equalization (CLAHE), unsharp masking, peripheral equalization and Trex processing. Since the preprocessing technique used depends on the type of automated diagnosis method to be adapted further down the pipeline, there is no definitive standard or effectiveness measure of how different preprocessing techniques score against each other. It is subjective and varies according to the image acquisition technique and instrument used [36], [37].

IV. Feature Generation and Extraction

The goal of a statistical pattern recognition technique is to choose those features that allow pattern vectors belonging to different categories to occupy compact and disjoint regions in a d-dimensional feature space [38]. The effectiveness of the feature set is determined by how well patterns from different classes can be separated. The decision boundaries for separating patterns or features belonging to different classes are determined by the probability distributions of these patterns. As a result, it is essential to formulate the patterns in such a way that they can be classified in the most accurate and computationally efficient manner [26], [29].

A main step in any kind of pattern recognition problem is the representation of data in a reduced number of dimensions. This is carried out for a number of reasons such as improved classification, stable representation or ease of computation. Geometrical methods such as principal component analysis and multidimensional scaling are used for representing data in a reduced dimension. These methods, apart from several others such as linear discriminant analysis and Karhunen-Loeve expansion, are collectively referred to as feature selection and feature extraction methods [39]. In any classification task, it is imperative to identify those variables that do not contribute to the classification. It is essential to get rid of these variables which do not contribute to the effectiveness of the classification. In other words, it is essential to reduce a d-dimensional feature vector into an m-dimensional vector $(m \leq d)$ such that m represents the most effective set of feature measurements for the given classification problem. This can be done in two ways. One method is to seek d features out of the available p measurements. This process is called feature selection in the measurement space or simply feature selection [39]. The second technique is to find a transformation from the p measurements to a lower dimensional feature space. This is called feature selection in the transformed space or feature extraction. The transformation is either a linear or nonlinear combination of the original variables and may be supervised or unsupervised [39].

To put it in perspective, for feature selection, the optimization is over the set of all possible subsets of size d, χ_d, of the p possible measurements, $x_1, \ldots, x_p$. Thus, we seek the subset $\bar{X}_d$ for which

$$J(\bar{X_d}) = \max_{X \in \chi_d} J(X). \quad (20)$$

In feature extraction, the class of transformation is usually specified and we seek the transformation $\bar{A}$, for which

$$J(\bar{A}) = \max_{A \in \Lambda} J\left(A(x)\right) \tag{21}$$

where Λ is the set of allowable transformations, letting the feature vector be $y = \bar{A}(x)$. Reference [42] defines the purpose of feature extraction and selection as keeping the number of features as small as possible to design classifiers with good generalization capabilities. Reference [38] argues, since feature selection is done in an offline manner, optimality of the feature subset is more critical than the execution time of a particular algorithm. As a result, many feature extractions and selections have a preprocessing step with three major processes: outlier removal, data normalization and handling missing data [42].

The concept of feature generation is akin to transforming a given set of measurements to a new set of features. Suitably chosen transforms can provide information packing to a great extent, thereby reducing information redundancies present in the original data.

For an M-class problem with feature vectors distributed according to $p(x|\omega_i)$, $i = 1, 2, \ldots, M$, the likelihood functions are given in a parametric form and the corresponding parameters form the vectors θ_i which are unknown. The dependence on θ_i is written as $p(x_i|\theta_i)$. To estimate the unknown parameters using a set of known feature vectors in each class, several techniques such as maximum likelihood estimators, maximum *a posteriori* probability estimators, Bayesian interference, maximum entropy estimation, mixture model expectation maximization (EM) algorithms are used [42]. For handling missing data, probability density functions (pdf) can be efficiently calculated using the expectation maximization (EM) algorithm. In this case, pdf of the incomplete data is given by

$$p_x(\mathbf{x}; \boldsymbol{\theta}) = \int_{Y(x)} p_y(y; \theta) dy. \tag{22}$$

The maximum likelihood estimate of θ is given by

$$\hat{\theta}_{ML} : \sum_k \frac{\partial ln\left(p_y(y_k; \theta)\right)}{\partial \theta} = 0. \tag{23}$$

Here, the complete data samples are represented by $y \in Y \subseteq \Re^m$, and the corresponding pdf by $p_y(y; \theta)$, where θ is an unknown parameter vector. However, the samples y cannot be directly observed and are hence referred to as incomplete data. Instead, we are able to observe samples $x = g(y) \in X_{ob} \subseteq \Re^l$, $l < m$. The corresponding pdf is represented by $p_x(x; \theta)$. Also, $Y(x) \subseteq Y$ is assumed to be the subset of all the y's corresponding to a specific x.

However, in (20), y's are not available. So, the EM algorithm maximizes the expectation of the log-likelihood function, conditioned on the observed samples and the current iteration estimate of θ [42]. The two steps in the algorithm are the expectation step $(E - \text{step})$

$$Q\left(\theta; \theta(t)\right) = E\left[\sum_k ln\left(p_y(y_k; \theta | X; \theta(t)\right)\right] \tag{24}$$

and the maximization step $(M - \text{step})$ which involves maximizing $Q(\theta; \theta(t))$ obtained from the E-step, where differentiability is assumed

$$\theta(t+1) : \frac{\partial Q\left(\theta; \theta(t)\right)}{\partial \theta} = 0. \tag{25}$$

For the EM algorithm, an initial estimate $\theta(0)$ is assumed and iterations are terminated if $\|\theta(t+1) - \theta(t)\| \leq \epsilon$.

With the pdfs of data available, it is important to extract meaningful features from them in order to proceed with further processing. Most of the features used in describing pictorial information are those which are similar to the ones human beings use in interpreting images [44]. Major features used in describing mammographic images are spectral, textural and contextual. The definition of these features vary greatly, but a widely accepted definition is provided by [44]. Here, spectral features are described as the average tonal variations in various bands of the visible and invisible bands of the electromagnetic spectrum. Similarly, textural features are said to contain information about the spatial distribution of tonal variations within a band and contextual features are said to contain information derived from blocks of pictorial data surrounding the area being analyzed.

While spectral and contextual features are useful in other image processing and analysis techniques, textural features are the most sought after in mammographic image analysis. This is due to the fact that mammograms are images which are obtained using a single medium of acquisition and spatial distribution of features in these images can be found within a single band, making it imperative for the use of textural features. In textural feature analysis, texture can be evaluated as being fine, coarse or smooth, rippled, molled, irregular or lineated [44]. All these types of textures play an important role in extracting meaningful information from the mammograms. Feature extraction and projection methods can be broadly classified into techniques based on linear mapping and nonlinear mapping. The techniques can either be iterative or noniterative [38]. Based on this, several textural features are defined for use in mammogram analysis [45], [46], [49], [50], [51]. In all of the following definitions, $P(i, j)$ is the image in consideration and N_g is the set of quantized gray values.

A simple energy-based feature which describes the energy contained in a subregion of the image can be used to find the intensity level variations in the image

$$\text{Energy} = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij}^2. \tag{26}$$

Angular second-moment feature (ASM) is a measure of homogeneity of the image. In a homogeneous image, there are very few dominant gray-tone transitions

$$\text{ASM} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(\frac{P(i,j)}{R}\right)^2. \tag{27}$$

Contrast features are difference moments and are a measure of the contrast or the amount of local variations present in an image

$$\text{CF} = \sum_{n=0}^{N_g-1} n^2 \left\{\sum_{|i-j|=n}\left(\frac{P(i,j)}{R}\right)\right\} \tag{28}$$

where R is a normalizing constant.

Correlation features are a measure of gray-tone linear dependencies in the image

$$\text{CorrF} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[\frac{ijP(i,j)}{R}\right] - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{29}$$

where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of the marginal distributions associated with $P(i,j)/R$.

Many of the texture features described here are functions of distance and angle. The angular dependencies present specific problems. For instance, consider an image A with features a, b, c, d for angles 0°, 45°, 90°, 135° respectively, and image B which is identical to A except that B is rotated 90° with respect to A. Hence, B will produce features c, d, a, b. Since both A and B are identical with similar texture features, any decision rules must produce the same result in a classifier. But, in order to guarantee this, it has been recommended that the angularly dependent features obtained from both the images, instead of being used directly, have to be averaged for each function and then used as input to a classifier [44].

Apart from the aforementioned textural features, several other features can be used in mammographic gray scale images. A measure to study the homogeneity of a subregion in an image can be defined as

$$\text{Homogeneity} = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{P_{ij}}{1+|i-j|}. \tag{30}$$

Just as in the case of mechanical objects, inertia of an image can be calculated with the help of image intensities

$$\text{Inertia} = \sum_{i=1}^{n}\sum_{j=1}^{m}(i-j)^2 P_{ij}. \tag{31}$$

Sum of squares (SoS) can be used to find out the variance between the pixels of interest and their mean to extract relevant features

$$\text{SoS} = \sum_i\sum_j (i-\mu)^2 p(i,j) \tag{32}$$

where μ is the mean of a subset of pixels considered.

Geometric moments are widely used to extract useful information from images for pattern classification. The popularity of this geometric moments lies in the fact that moments provide an equivalent representation of an image, such that a whole image can be reconstructed from its moments [52]. As a result, each moment conveys certain useful information about the image. A general representation of the geometric moment (GM) of order $k+q$ of an image $P(i,j)$ can be represented as

$$\text{GM} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^q P(i,j)dxdy. \tag{33}$$

Though geometric moments can be used to represent image features in an efficient manner, it is also to be noted that they are not free from the invariance property which is important in geometric transformations for pattern generation. Hence, it is necessary to normalize the moments such that they are invariant to translation, scaling and rotations.

The moments found in (33) can also be viewed as projections of the image $P(i,j)$ on the basis function formed by the monomials $x^k y^q$ [52]. Since these monomials need not necessarily be orthogonal, the resulting geometric moment features might have information redundancy. To overcome this problem, a set of polynomial functions called the Zernike polynomials formed over a unit circle are generated [54]. The Zernike moments for an image are defined as

$$A_{mn} = \frac{p+1}{\pi}\sum_k P(i_k, j_k)V*(\rho_k,\theta_k), \quad i_k^2 + j_k^2 \le 1 \tag{34}$$

where i runs over all the image pixels. Similar to Zernike moments, another set of moments referred to as the Hu moments [56] are defined. A set of seven moments, the features obtained using Hu's technique are invariant under the actions of translation, scaling and rotation.

A measure of local homogeneity referred to as difference moments for usage in mammograms is discussed in [57]

$$DM = \sum_{n=0}^{N-1} n^2 \sum_{i-j=n}\left(\frac{P(i,j)}{R}\right). \tag{35}$$

Inverse difference moment (IDM) is almost similar to SoS and is a measure of the local homogeneity

$$\text{IDM} = \sum_i\sum_j \frac{1}{1+(i-j)^2}p(i,j). \tag{36}$$

Sum average (SA) is found from the pixel in consideration and the size of the gray scale

$$\text{SA} = \sum_{i=2}^{N_g} i p_{x+y}(i). \tag{37}$$

Entropy is computed from the second order histogram and provides a measure of nonuniformity

$$\text{Entropy} = -\sum_i \sum_j p(i,j) log\left(p(i,j)\right). \tag{38}$$

Sum entropy (SE) is calculated as a logarithmic function of the image in consideration. In the case of Sum entropy, some of the probabilities might be zero and $log(0)$ is not defined. Hence, the term $log(p+\epsilon)$, where ϵ is an arbitrarily small positive constant, is used in place of $log(p)$. The same logic holds true for all types of entropy calculations

$$\text{SE} = -\sum_{i=2}^{2N_g} p_{x+y}(i) log\left\{p_{x+y}(i)\right\}. \tag{39}$$

Sum variance (SV) is calculated from the original image and the sum entropy SE calculated previously

$$\text{SV} = \sum_{i=2}^{2N_g} (i - SE)^2 p_{x+y}(i). \tag{40}$$

A Haralick feature [44] referred to as texture probability (TP), of run length of 2, is defined as

$$\text{TP} = \sum_{i=1}^{n} \frac{(P_i - P_{ii})^2 P_{ii}}{P_i^2}. \tag{41}$$

A gray level run is a set of consecutive pixels having the same gray value. The length of the run is the number of pixels in the run. Depending on the total number of runs, several types of run emphases such as short run, long run, gray level nonuniformity, run length nonuniformity and run percentages can be obtained.

Cluster tendency (CT) is yet another Haralick feature which can be tuned with user defined parameter. In this case, the parameter to be adjusted is k

$$\text{CT} = \sum_{i=1}^{n} \sum_{j=1}^{m} (i + j - 2\mu)^k C_{ij}, \quad k = 1, 2, \ldots (i,j). \tag{42}$$

Difference variance (DV) is a variance measure between the image intensities calculated as a function of the SE calculated previously

$$\text{DV} = \sum_{i=2}^{2N_g} (i - SE)^2 p_{x-y}(i). \tag{43}$$

Difference entropy (DE) is an entropy measure which provides a measure of nonuniformity while taking into consideration a difference measure obtained from the original image

$$\text{DE} = -\sum_{i=0}^{N_g - 1} p_{x-y}(i) log\left\{p_{x-y}(i)\right\}. \tag{44}$$

A set of measures known as the information measures of correlation [IMCorr_1, IMCorr_2 and maximum correlation coefficient (MCC)] obtained from the entropy of P_x and P_y is calculated to bring out certain properties not easily identified using the rectangular correlation measure CorrF described previously.

$$\text{IMCorr}_1 = \frac{\text{Entropy}_{x,y} + \sum_i \sum_j p(i,j) log\{p_x(i) p_y(j)\}}{max\{\text{Entropy}_x, \text{Entropy}_y\}} \tag{45}$$

$$\text{IMCorr}_2 = \left(1 - exp\left[-2\left(-\sum_i \sum_j p_x(i) p_y(j) \log\{p_x(i) p_y(j) - \text{Entropy}_{x,y}\right)\right]\right)^{\frac{1}{2}} \tag{46}$$

$$\text{MCC} = (\text{Second largest eigenvalue of } Q)^{\frac{1}{2}} \tag{47}$$

and

$$Q(i,j) = \sum_k \frac{p(i,k) p(j,k)}{p_x(i) p_y(k)}. \tag{48}$$

Apart from these information measures of correlation, another measure of correlation for use in mammograms discussed in [57] provides a measure of linear dependency of brightness

$$\text{Corr} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \left[\frac{ijP(i,j)}{R}\right] - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{49}$$

where μ and σ are the mean and standard deviation of the group of pixels in consideration.

Deviation measures from the second order histogram, $H_m(y_q, d)$, can be defined as

$$\text{Deviation} = \sqrt{\sum_{y_q=y_1}^{y_t} \left[y_q - \sum_{y_p=y_1}^{y_t} y_p H_m(y_p, d)\right] H_m(y_q, d)}. \tag{50}$$

This feature is a measure of the density of distribution of $H_m(y_q, d)$ about the mean. Its value is small if the histogram is concentrated around the mean.

With textural features playing a major role in many mammographic analysis techniques, gradient-based measures have also been found useful in breast cancer detection [45], [57]. Most discussions in literature with respect to gradient-based techniques are based on the motivation that malignant breast lesions often permeate larger areas than apparent on mammograms. So, traditional gradient-based techniques might not be able to reveal clear-cut transitions or gradient information. To counter this and

to provide global gradient information about the lesion masses, directional gradient features that tolerate uncertainties in the location of mass boundaries have been developed [57]. These directional derivatives are used to find a measure of variations in intensities referred to as acutance [45]. Acutance is measured as

$$A_{dg} = \frac{1}{(f_{\max} - f_{\min})} \frac{\sum_{i=1}^{N} d_i}{N} \tag{51}$$

where $f_{\max}$ and $f_{\min}$ are the local maximum and minimum pixel values in the stretch of pixels in the region of interest (ROI), N is the number of pixels along the boundary of the ROI, d_i is the rms gradient at the ith boundary point and is defined as

$$d_i = \sqrt{\frac{\sum_{j=0}^{(n_i-1)} [f_i(j) - f_i(j+1)]^2}{n_i}} \tag{52}$$

where $f_i(j)$, $j = 0, 1, 2 \ldots, n_i$ are $(n_i + 1)$ number of pixels available along the perpendicular at the ith boundary point including the boundary point.

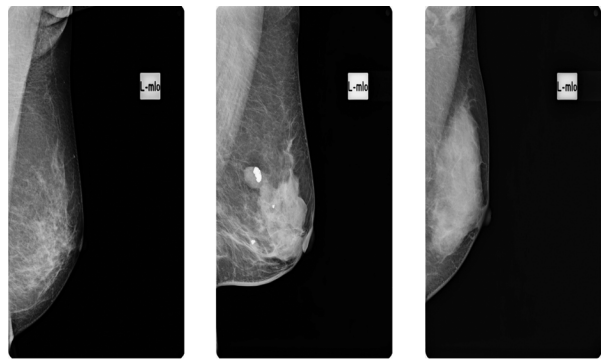
Spiculation features used to identify malignant tissues have been discussed in [46]. Spicules are a radiating pattern of linear spikes surrounding the irregularly shaped malignant densities. Spicular features are identified by a multiscale approach where, at any given scale, accurate line-based orientation estimates are obtained from the output of 3-D second-order, Gaussian derivative operators. The orientation at the scale at which these operators have maximum response is selected. This method provides an estimate of the orientation of the structure. The orientation map is used to obtain radial patterns of straight lines. With this information, stellar patterns or presence of malignant tissue is identified using classifiers.

Morphological features based on the physical characteristics of lesions are discussed in [49]. According to this study, the circularity of suspicious masses can be studied to find information about the malignancy of lesions. The idea that benign lesions are more or less round or oval, while malignant lesions are irregularly shaped, proves to be a valid ground for using circularity as a morphological feature. In this study, the authors define circularity as

$$c = \frac{p^2}{A} \tag{53}$$

where p is the perimeter and A the size of the region.

Reference [49] also discusses difference and similarity features based on prior views of mammograms. In this technique, previously obtained mammographic images are analyzed in conjunction with the most recent mammographic image. Difference and similarities between the two images are studied to find if the suspicious masses have grown in size or have remained unchanged. Registration is a common method used in this technique, with the prior mammogram registered to the current mammographic image. This registration can then be used to obtain registration scores, with a very high score



(a) (b) (c)

Fig. 1. Typical mammogram images (a) normal, (b) benign, and (c) malignant.

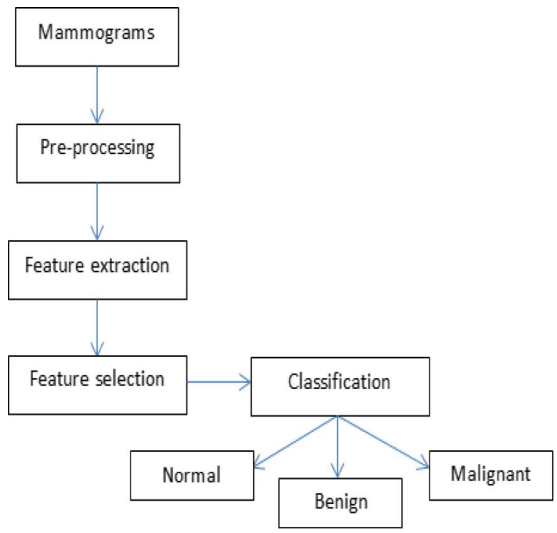


Fig. 2. Computer-aided diagnosis system pipeline.

indicating an unchanged mass or a benign tissue, while a low score indicates a malignant tissue.

Another feature that can be obtained using prior mammograms is the relative grey level change (RGLC). In this technique, the cumulative histograms of prior and current images are calculated. Following this, relative grey level changes are studied between a similar region on the prior and the current view. The relative grey level change between the prior and current views is defined by

$$\text{RGLC} = \frac{1}{N} \sum_{(m,n) \in C} (\bar{y}_c(m,n) - y_p(m',n')) \tag{54}$$

where N denotes the number of pixels inside C, $\bar{y}_c(m,n)$ denotes the transformed gray level at location (m,n) in C, $y_p(m',n')$ the gray level at the same relative location in the prior region with center (m_s, n_s), while$\bar{y}$, which is the histogram matched gray level, is defined as

$$\bar{y} = f_P^{-1}(f_C(y)) \tag{55}$$

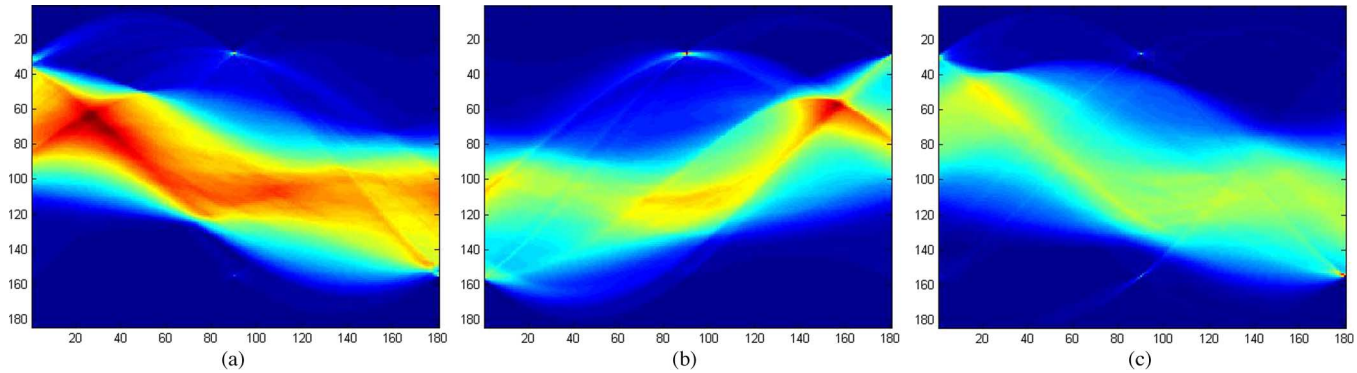


Fig. 3. Features extracted for mammogram images (a) normal, (b) benign, and (c) malignant.

where f_C and f_P are the cumulative histograms of the current and prior images respectively, which are defined as

$$f_C(y) = \sum_{i=0}^{y} H_C(i) \tag{56}$$

$$f_P(y) = \sum_{i=0}^{y} H_P(i) \tag{57}$$

where H_C and H_P are the gray value histograms calculated inside the breast area on the current and the prior image, respectively.

Usage of fractal theory for extracting features in computer-aided diagnosis of breast cancer has been discussed in [50]. In this work, fractal dimensions of mammograms are obtained by using the differential box counting method. The fractal dimension D of a set A is calculated as

$$D = \frac{log(N_r)}{log\left(\frac{1}{r}\right)}, \quad (1 = N_r.r^D) \tag{58}$$

where A is a bounded set in a n-dimensional space, where A can be represented as its own nonoverlapping unity N_r, with each copy being similar to A, scaled down by a factor of r. Reference [50] used multilevel fractal dimensions by analyzing the images at different scales and obtaining a D for images at each scale and using every D as a feature to form a combined feature vector. This feature vector can then be used for further classification.

Apart from the feature extraction methods, wavelet features have been known to be used in extracting useful information from mammographic images [58], [68]. Wavelets can be defined in the way described by (7) and (8). The energy features for these decomposed wavelet packets for Level 0 decomposition and the wavelet packets of Level 1 decomposition are defined as [68]

$$\text{Energy} = \frac{\sum_i \sum_j x_{ij}^2}{\text{length} * \text{breadth}} \tag{59}$$

where x_{ij} is the computed wavelet packet value at the ith row and jth column of the wavelet packet and length and breadth are the dimensions of the wavelet packet. Similarly, entropy features are defined as

$$\text{Entropy} = -\sum_i \sum_j \left[\frac{x_{ij}^2}{\text{norm}}\right] \log_{10} \left[\frac{x_{ij}^2}{\text{norm}}\right] \tag{60}$$

where norm $= \sum_i \sum_j x_{ij}^2$.

The above-mentioned feature generators are quite exhaustive in terms of the techniques used in general computer-aided detection of breast cancer. Several other feature generators, though available, have yet to be tested on mammographic images. The reason for a very selective usage of techniques in the case of mammograms is that the image properties are such that not all feature generators can provide optimal information to be used by classifiers specific to computer-aided diagnosis.

Fig. 3 shows an example of a feature extraction algorithm (trace transform) presenting good differentiation between the normal, benign and malignant mammograms seen in Fig. 1.

V. Curse of Dimensionality

It is an established fact that performance of a classifier depends on the features selected for classification. It is also known that size of the feature vector greatly affects classification rates. It is not always necessary for a large feature vector to translate into better accuracy rates. It has been observed that with added features, the classifier performance might actually degrade if the number of training samples that is used to design the classifier is small in relation to the number of features used [38], [59]. This phenomenon is referred to as the curse of dimensionality or the peaking phenomenon. Reference [38] provides an explanation of this behavior by stating that the most commonly used parametric classifiers estimate the unknown parameters and plug them in for the true parameters in the class-conditional densities. For a fixed sample size, as the number of features is increased along with corresponding increase in the number of unknown parameters, the reliability of the parameter estimates decreases. Consequently, the performance of the resulting classifiers for a fixed sample size might degrade with an increase in the number of features.

From this discussion, it can be seen that careful selection of feature vectors is necessary for a proper classification result. One way to view dimensionality reduction is to define it as a

process that involves identification of variables that do not contribute to the classification task. Every dimensionality reduction problem looks at solving the question of finding the best subset of size d given a set of p measurements. Reducing the number of variables also helps in reducing redundancy, eliminating unnecessary variables. From available literature, it is seen that feature selection is not widely used in mammographic image analysis. But it is imperative to use feature selection techniques to identify the best features from a given set of measurements. Several concrete discussions about methods for feature selection have been made in [38] and [39].

In order to choose a good feature set $\Re$, the ability of $\Re$ to discriminate accurately between two or more classes is to be studied. To study the accuracy rates of different feature subsets $\mathbb{Z} : \Re$, $\Re < \infty$, a class separability measure optimized for $\mathbb{Z}$ needs to be found. This can be done by selecting a subset of $\mathbb{Z}$ for which a particular classifier performs well. For a different classifier, another feature set can be chosen. But, this cannot be an optimal solution for an unsupervised or a real-time system. So, a better way to find an optimal feature set is to choose the best $\mathbb{Z}$ for $\mathbb{Z}_x < \mathbb{Z}_y$; $\mathbb{Z}_x$, $\mathbb{Z}_y : \mathbb{Z}$.

Class Separability Measures: If individual features are tested for their ability to discriminate between classes, correlation between different vectors will not be taken into account. So, the ability of combined feature vectors to discriminate between classes is to be considered more than the ability of individual features. Class separability measures help in determining the separability criterion of various feature classes. These measures are based on the familiar Bayes rule. Given two classes ω_1 and ω_2, with a feature vector x, class ω_1 is chosen if

$$P(\omega_1|x) > P(\omega_2|x). \tag{61}$$

The ratio between $P(\omega_1|x)$ and $P(\omega_2|x)$ can hence provide crucial information about the discriminatory properties of the classifier and in turn the feature vectors. Let the distance between the two classes be D_{12}. This distance can be obtained from (61), as follows. For completely overlapping classes, $D_{12} = 0$. This value for D_{12} can be found for varying values of x and the mean value over class ω_1 can be considered as

$$D_{12} = \int_{-\infty}^{\infty} p(x|\omega_1) ln \frac{p(x|\omega_1)}{p(x|\omega_2)} dx. \tag{62}$$

Similarly, for class ω_2 the distance D_{21} can be calculated in the same way as above, and total sum d_{12} can be expressed as a sum of D_{12} and D_{21}. The sum d_{12} is referred to as the divergence and is a separability measure for multiclass problems. In multiclass problems, the divergence is computed by [42]

$$d_{ij} = D_{ij} + D_{ji} \tag{63}$$

$$= \int_{-\infty}^{\infty} \left(p(x|\omega_i) - p(x|\omega_j)\right) ln \frac{p(x|\omega_i)}{p(x|\omega_j)} dx. \tag{64}$$

The above relation which is usually used to calculate the divergence, is also referred to as the Kullback-Leibler distance measure between density functions [61]. Yet another distance measure which is used in a variety of applications is the Mahalanobis distance. Considering the case of two Gaussian distributions with equal covariance matrices $\Sigma_i = \Sigma_j = \Sigma$ and means μ_i and μ_j, the divergence, according to Mahalanobis distance can be defined as

$$d_{ij} = (\mu_i - \mu_j)^T \Sigma^{-1} (\mu_i - \mu_j). \tag{65}$$

With the Bayes classifier and divergence, the minimum classification error for a two class problem is defined by

$$P_e = \int_{-\infty}^{\infty} \min \left[P(\omega_i)p(x|\omega_i), P(\omega_j)p(x|\omega_j)\right] dx. \tag{66}$$

Extending this with the Chernoff bound and Battacharya distance, the Chernoff bound for classifier errors is given in [42]

$$\epsilon_{CB} = \sqrt{P(\omega_i)P(\omega_j)} \exp(-B) \tag{67}$$

where B is the *Battacharya distance* and is defined as

$$B = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} ln \frac{\left|\frac{\Sigma_i + \Sigma_j}{2}\right|}{\sqrt{|\Sigma_i||\Sigma_j|}}. \tag{68}$$

Feature Selection: Class separability measures are used to estimate the best feature sets for a classification problem. It is necessary to carefully choose features since the choice of features is critical in the final accuracy. Watanabe's Ugly Duckling theorem [62] proves a good point in favor of a good choice of feature sets by proving that two totally different datasets can be made to look similar by choice of the wrong features. This also holds true for a choice of wrong subset of features for classification. So, feature selection is an important step in any classification problem.

The most basic approach to a feature selection problem is analyzing all the subsets of a feature vector $\Re$ and choosing the subsets with the lowest error rate ϵ_{CB} as calculated from (67). This can be a computationally intensive task for huge datasets. But, feature selection is done in the development stage of an algorithm, which means that the step is done offline in most cases. As a result, computational efficiency is not as important as an optimal solution to the problem. Even then, exhaustive feature subset evaluations can become prohibitively expensive for large datasets.

For a given set of features with dimensionality d and desired number of features m, every possible subset of the feature vectors can be studied individually in the *best individual features* method of feature selection. This is an exhaustive search algorithm and can produce an optimal subset of features. But, this would take a tremendous amount of computational complexity to complete it. As a result, suboptimal selection tech-

niques which have a balanced tradeoff between optimality and computational efficiency have been proposed and developed. This can be done both in a *bottom-up* or a *top-down* approach.

When single features are selected, the correlation between the features is studied and features with heavy overlap are ranked lower in the feature vector. If x_{nk} is the kth feature of the nth pattern, with $n = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, m$, the cross correlation between any two of them is given by

$$\rho_{ij} = \frac{\Sigma_{n=1}^{N} x_{ni} x_{nj}}{\sqrt{\Sigma_{n=1}^{N} x_{ni}^2 \Sigma_{n=1}^{N} x_{nj}^2}}. \tag{69}$$

The feature selection procedure is carried out by selecting a class separability criterion C and computing its value for all available features x_k, $k = 1, 2, \ldots, m$. The C value is then ranked in descending order. The one with the best C value is named x_{i1}. The cross-correlation coefficient between x_{i1} and the remaining features as in (69) is calculated to select the second best feature. The feature x_{i2} for which $i_2 = \arg\max_j \{\alpha_1 C(j) - \alpha_2 |\rho_{i_1 j}|\}$, for all $j \neq i_1$, is calculated to find the next best feature and the procedure is carried on further for all values of k to rank the best features. From this ranked list of features, the required number of features is then chosen.

The only optimal search procedure which is not exhaustive is the *branch and bound* procedure which is a top-down approach. The method involves starting a search procedure beginning with the set of p variables and constructing a tree deleting variables successively. The feature selection criterion used in this procedure is

$$X \subset Y \Rightarrow J(X) < J(Y). \tag{70}$$

In other words, the performance of a feature subset should optimally improve whenever a new feature is added to it. This is the *monotonicity property*. All methods other than the branch and bound are referred to as suboptimal as the best pair of features found in other techniques need not necessarily contain the best single feature according to [64]. There are several types of suboptimal feature selection techniques such as best individual N, sequential forward selection, generalized sequential forward selection, sequential backward selection, generalized sequential backward selection, plus l-take away r selection, generalized plus l-take away r selection, sequential forward floating selection (SFFS) and sequential backward floating selection (SBFS). The last two methods are generally referred to as floating search methods. Detailed discussions about each of these techniques can be found in [39].

Another simple yet effective method for selection of features is through dimensionality reduction using principal component analysis (PCA). References [66] and [67] discuss the use of PCA using eigenvalue decomposition on covariance matrices of the observed regions and finds their principal components. Once the principal components are found, the first n significant components can be extracted to reduce the dimensionality of the feature vector. Though PCA is a widely used method for dimensionality reduction, it is not a popular method for analysis of mammograms.

VI. Classification

After the patterns in a data have been extracted to form feature representations, classifiers can be developed using several approaches depending on the features available and the pattern classification problem in hand. The choice of a classifier is usually not a simple task. In general, several classifiers are evaluated and tested before arriving at a final choice. Given a classification task, conditional probabilities referred to as *a posteriori probabilities* are calculated. In other words, the probability that the unknown pattern belongs to a particular class is calculated using these conditional probabilities. Classifying an object x is assigning it to a class ω_i with the highest class posterior probability $p(\omega_i|x)$. This class posterior probability can also be written in terms of the class conditional probability as $p(x|\omega)$.

Equations (22)–(25) describe the estimation of maximum likelihood parameters for the probability densities. These estimations can be classified into parametric and nonparametric estimations. Mixture modelling, Bayesian interference, and maximum likelihood estimation are examples of parametric modelling of the probability distribution functions (pdf). Nonparametric techniques are variations of histogram approximation of an unknown pdf. The basic idea of a nonparametric estimation technique is to divide the complete histogram into several bins and calculate the probability of a random sample belonging to one particular bin in the histogram. Parzen windows are very good examples of nonparametric estimation models in a multidimensional sense. In this case, instead of dividing the histogram into a number of bins, the m-dimensional space is divided into hypercubes with length of side h and volume h^l. In the case of a Parzen window, the probability $\hat{p}(x)$ of a variable belonging to a particular hypercube can be given by

$$\hat{p}(x) = \frac{1}{h^l}\left(\frac{1}{N}\sum_{i=1}^{N}\phi\left(\frac{x_i - x}{h}\right)\right). \tag{71}$$

It can be seen that this is for a discontinuous case. But, we need to look at the relation from the point of view of a smooth function, where $\phi(x) \geq 0$ and $\int_x \phi(x)dx = 1$. By considering this case, a smooth function known as the kernel or potential function or a Parzen window for estimation of class densities is obtained. When the fit for such a density model is evaluated, an error representation similar to the log-likelihood $LL(\mathbf{X})$ is to be estimated

$$LL(\mathbf{X}) = \log\left\{\prod_i \hat{p}(x_i)\right\} = \sum_i \log\left(\hat{p}(x_i)\right). \tag{72}$$

The better the data x fits with the probability density model $\hat{p}$, the higher the values of $\hat{p}(x)$ will be. This will result in a high value of $\sum_i log(\hat{p}(x_i))$. The probability density estimate $\hat{p}$ with the highest value of LL is to be used in calculations. From (71), it can be seen that with varying values of the width h, we might get varying values of $\hat{p}$.

In Parzen estimation of pdf, the volume around the points x is considered fixed with a representation of h^l and the number of points k_N falling inside the volume is left to vary randomly from point to point. By reversing this functionality by having the number of points k_N fixed, with the size of volume around x varying each time to include k points, we obtain the k nearest neighbor (k-NN) density estimation. The property of a k-NN density estimation is that, in low-density areas, the volume will be large and in high-density areas, it will be smaller. With this property, the estimator can be defined as

$$\hat{p}(x) = \frac{k}{NV(x)}. \tag{73}$$

In this relation, the dependence of the volume $V(x)$ is clearly seen. Or in other words, the probability $\hat{p}$ is inversely proportional to the volume V. In the case of mammographic image analysis, k-NN density estimates have been widely used. Reference [68] provides a discussion on the use of k-NN density estimators for microcalcification estimates in mammograms. When the value of $k = 1$, the k-NN estimator is evaluated as the *nearest neighbor* estimator. For this estimation, a feature vector x is assigned to the class of its nearest neighbor. Distance measures such as Euclidean and Mahalanobis distance can be used to evaluate the class of each feature vector. Though the nearest neighbor estimator is seemingly simple, it is quite effective when the training sample is large enough [71].

References [72], [75], and [76] discuss parametric multivariate methods for breast cancer detection. This theory is based on linear classifier theory. The general idea of these classifiers is that, considering a training pattern $x_1, x_2, \ldots\ldots, x_n$, with each of these being assigned to one of the two classes ω_1 or ω_2, a weight vector $\mathbf{w}$ and a threshold ω_0 is determined such that

$$\mathbf{w^T}x + \omega_0 \begin{cases} > 0 \\ < 0 \end{cases} \Rightarrow x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}. \tag{74}$$

In this relation, $\mathbf{w^T}x + \omega_0$ can be rewritten as $\nu^T z$ where z is the augmented pattern vector and ν is a multidimensional vector with elements $(\omega_0, \omega_1, \ldots\ldots, \omega_n)^T$. If $\nu^T z > 0$ for all samples in a single class, the data is said to be linearly separable. This is the basis of a linear discriminator design. The error patterns, the absolute correction and fractional correction rules assigned to linear discriminators, determine the error percentages of these classifiers apart from the variable margins which classify the samples to their respective classes.

A hybrid classifier involving LDA along with adaptive resonance theory (ART) was proposed by [78] for mammographic breast cancer detection. This classifier utilized a neural network scheme along with LDA for classification. The reason the authors have proposed using a neural network-based classifier in combination with LDA is that supervised classifiers such as LDA tend to overlook previously learned expert knowledge during classification. If an ART scheme is used in combination with LDA, the performance can be improved according to the arguments provided by the authors. This ARTLDA hybrid classifier is defined as

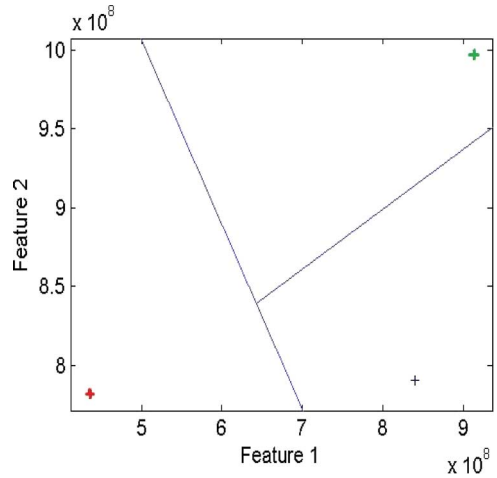


Fig. 4. Scattered plot of the three classes of mammograms along with their classification hyperplanes.

$$y_{AL} = g\left(f_2(x)\right) f_1(x) + 1 - g\left(f_2(x)\right) \tag{75}$$

where x is the input vector, $f_1(x)$ is the LDA classifier, $f_2(x)$ is the ART classifier and $g(x)$ is a binary membership function which labels the classes identified by ART to be either malignant or mixed. A class is said to be mixed if samples in it have both malignant and benign features. The membership function in this classifier can be defined as

$$g(c) = \begin{cases} 0, & \text{if } c \text{ is a malignant class} \\ 1, & \text{if } c \text{ is a mixed class} \end{cases}. \tag{76}$$

This hybrid classifier is unique in the way that the classification of the first level of classifier, namely the ART classifier, is checked for the class. If the class is malignant according to the classification, no further processing is done. But, if the class is labelled as mixed, the data is then processed through a LDA classifier to further classify the data as malignant or benign. A combination of support vector machines (SVM) along with LDA was utilized by [82] to obtain better classification in computer-aided mammographic analysis.

Fig. 4 shows a scattered plot of the three classes of mammograms seen in Fig. 1. The blue lines indicate the classification hyperplanes separating the classes.

There are numerous criteria which are adapted to maximize the separation between classes in a linear classification scheme. For instance, Fisher criterion provides a ratio of between-class to within-class variances represented as

$$J_F = \frac{\left|\omega^T(m_1 - m_2)\right|^2}{\omega^T S_w \omega} \tag{77}$$

where the main criterion to increase discrimination between classes is to find a value for ω such that the ratio is a maximum.

In the previous equation, m_1 and m_2 are the group means and S_w is the pooled within-class covariance matrix, in its bias-corrected form given by

$$\frac{1}{n-2}(n_1\Sigma_1 + n_2\Sigma_2) \tag{78}$$

where Σ_1 and Σ_2 are the maximum likelihood estimates of the covariance matrices of classes ω_1 and ω_2 and n is the number of samples in class ω_i. The term linear discriminant analysis refers to the type of Bayesian classifier used for classification. If the covariance matrices are equal, the classifier is referred to as linear, while different covariance matrices give rise to quadratic discriminant analysis. The advantage with LDA is that no assumption is made with respect to the normality of the data. An extension of linear discrimination is logistic discrimination where the difference between the logarithms of the class-conditional density functions are studied [83]

$$\log\left(\frac{p(x|\omega_1)}{p(x|\omega_2)}\right) = \omega_0 + \omega^T x. \tag{79}$$

While the previous representation of a logistic classifier is for a two-class problem, it can easily be extended for a multiclass problem as in the case of breast cancer analysis, by adapting the form

$$\log\left(\frac{p(x|\omega_s)}{p(x|\omega_C)}\right) = \omega_{s0} + \omega_s^T x, \quad s = 1, \ldots\ldots C-1. \tag{80}$$

Yet another popularly used linear classifier is the support vector machine (SVM) which has been widely used in mammographic analysis [82], [92]. SVMs were first proposed by [100] and have since been used in a variety of applications including medical image analysis. The rationale of this method is to design a hyperplane for separating linearly separable data which can be of two or more classes [101]. This hyperplane can be defined by a linear discriminant function as

$$g(x) = \omega^T x + \omega_0 = 0. \tag{81}$$

This hyperplane is to be optimized such that the classes are separated with a maximal distance margin between the sample vectors and the *canonical* hyperplane

$$\begin{aligned} \omega^T x_i + \omega_0 &\geq +1, \quad \forall\, x \epsilon \omega_1 \\ \omega^T x_i + \omega_0 &\leq -1, \quad \forall\, x \epsilon \omega_2. \end{aligned} \tag{82}$$

These are the two hyperplanes with the two classes and $g(x) = 0$ being the separating hyperplane. The value of $1/\omega$ which gives the distance between the two hyperplanes and the separating hyperplane is referred to as the *margin*. Support vectors are the sample feature points that lie on the canonical hyperplane and classifiers which make use of these hyperplanes for classification are referred to as the support vector machines. The support vectors classify new patterns according to the sign of $\omega^T x + \omega_0$ from (82). For a good classification using support vectors, logic demands that the margin has to be maximized, which from the above discussion would mean that the value of ω has to be minimized. Using the Karush-Kuhn-Tucker conditions and Lagrange multipliers, we can arrive at a linear discriminant to be given by [39]

$$\sum_{i \in S\nu} \alpha_i y_i x_i^T x - \frac{1}{n_{s\nu}} \sum_{i \in S\nu} \sum_{j \in S\nu} \alpha_i y_i x_i^T x_j + \frac{1}{n_{s\nu}} \sum_{i \in S\nu} y_i > 0. \tag{83}$$

Here, $S\nu$ is the set of support vectors with associated values of α_i. In the case of multiclass problems, (81) can be generalized to the form [100]

$$g_k(x) = \left(\omega^k\right)^T x + \omega_0^k \; k = 1, \ldots\ldots, C. \tag{84}$$

In this generalized form, the decision rule becomes

$$g_i(x) = \max_j g_j(x). \tag{85}$$

This decision rule is used to separate the training data into classes, with solutions for ω_k, ω_0^k, for $k = 1, \ldots\ldots, C$ being given by

$$\left(\omega^k\right)^T x + \omega_0^k - \left(\left(\omega^j\right)^T x + \omega_0^j\right) \geq 1, \; \forall\, x \in \omega_k, \; j \neq k. \tag{86}$$

While the above classifier is for a linear case, in case of a mapping of the form $x \in \Re^l \rightarrow y \in \Re^k$, where a feature sample is mapped from an l-dimensional space to a k-dimensional space, a slightly different SVM methodology is to be adapted for implicit mappings in a multidimensional space. Functions known as *kernels* and *Hilbert spaces* or *reproducing kernel Hilbert spaces* [102] are used for mappings of SVM in a nonlinear space. The most commonly used kernels can be found in [103] as follows:

Polynomials

$$K(x,z) = \left(x^T z + 1\right)^q, \quad q > 0 \tag{87}$$

Radial Basis Functions

$$K(x,z) = exp\left(-\frac{\|x-z\|^2}{\sigma^2}\right) \tag{88}$$

Hyperbolic Tangent

$$K(x,z) = tanh(\beta x^T z + \gamma). \tag{89}$$

TABLE I
LISTING OF POPULAR FEATURE EXTRACTION AND CLASSIFICATION METHODS

Author	Method used	Accuracy (%)
Kimme at al., [165]	Normalized statistics and texture features	74
Petrosian et al., [166]	Spatial Gray Level Dependence and textural features with a decision tree classifier	76-89
Kinoshita et al., [167]	Shape and texture features with a three layer feed-forward neural network	81
Rangayyan et al., [45]	Region based edge-profile acutance measure	92
Polakowski et al., [168]	Model based vision algorithm. Difference of Gaussians and texture features	92
Priebe et al., [169]	Fractal texture measures	88
Sameti et al., [170]	Optical density, photometric and textural features	72
Chitre et al., [171]	Texture measures with artificial neural network	87
Mudigonda et al., [57]	Gray level co-occurence matrices, polygonal modeling with jack-knife classification	83
Brijesh et al., [172]	Statistical features with fuzzy neural network	83
Yoshida et al., [173]	Wavelet features in combination with a difference image technique	90
Liyang Wei et al., [92]	Statistical features in a multiple view mammogram with SVM and KFD	85
Oliver A et al., [174]	Eigen faces approach	82-90
Szekeley et al., [175]	Texture features and a combining classifier of decision trees and multiresolution markov random models	88-94
Alolfe et al., [82]	Forward stepwise linear regression method with a combined classifier of SVM and LDA	82.5-90

Appropriate values for β and γ are to be chosen for Mercer's conditions to be satisfied [102]. With a suitable kernel function chosen, the classifier can be rewritten as

$$g(x) = \sum_{i=1}^{N_s} \lambda_i y_i K(x_i, x) + \omega_0 \tag{90}$$

where x is assigned to class ω_1 or ω_2 according to the value of $g(x)$ being greater than or less than zero, respectively. In the case of mammographic image analysis, it can be assumed safely that the classes are more often than not going to be nonlinear. This is also seen from experimental results [92]. There have been new approaches to the use of SVM as seen in [82] where SVM classifiers have been used in combination with LDA to produce classifier results which are comparable or even better than the individual classification capabilities of the conventional single classifier. Reference [85] uses a soft clustering technique based on SVMs to classify abnormalities in mammograms.

Since an SVM works by transforming the nonlinear feature space into a linearly separable feature space with the help of kernels, it is practically useful in classifying mammograms due to the fact that mammograms are more often than not highly overlapping and nonlinear in their feature space. Due to the inherent capacity of SVM kernels to map this highly overlapping, nonlinear feature space into a more manageable feature space, it is highly successful in classification problems as seen by the references and discussions provided. It is also because of this inherent property of mammograms that LDA and other linear classifiers do not perform very well as seen from Table I.

I) Neural Network: This method of parallel distributed information processing is another widely used method in breast cancer detection [110], [111], [125]. The structure of a neural network is built in such a way that each element of the network possesses a local memory that carries out localized information processing operations. Multilayer perceptrons (MLP), as they are known, can provide a solution to common problems found in perceptrons which are nothing but linear hyperplanes [138]. The basic block of MLP creates a transformation of a pattern $x \in \Re^p$ to an n'-dimensional space according to the relation given by [138]

$$g_j(x) = \sum_{i=1}^{m} \omega_{ji} \phi_i \left(\alpha_i^T x + \alpha_{i0}\right) + \omega_{j0},\ j = 1, \ldots, n' \tag{91}$$

where ϕ_i is a fixed nonlinearity, usually to be taken to be of the logistic form representing the firing rate of a neuron as a function of the input

$$\phi_i(z) = \phi(z) = \frac{1}{1 + \exp(-z)}. \tag{92}$$

From (91) and (92), we can see that the basic function of a neural network consists of projecting the data onto each of the directions described by the vectors α_i, then transforming the projected data by the nonlinear functions $\phi_i(z)$ and finally forming a linear combination using the weights ω_{ji} [39]. In the structure of a neural network, there are input nodes and output nodes, with weights associated between the input nodes and the hidden nodes that accept the weighted combinations which perform the nonlinear transformation. The output nodes take a linear combination of the outputs of the hidden nodes and deliver them as outputs. Owing to the multilayered structure of an MLP, neural networks are a nonlinear model where the output is a nonlinear function of its parameters and the inputs. As a result, a nonlinear optimization scheme must be employed to minimize the optimization criterion.

The most commonly used technique for classification using neural networks is the back-propagation network which has been used in breast cancer detection [68]. Back propagation is the calculation of the derivative of an error function in multilayer networks. In neural networks, the synaptic weights of nodes are computed in such a way that the cost function is minimized. Reference [140] discusses the design of a three-layered

feedforward neural network for automated classification of microcalcifications for breast cancer detection. The authors train the network with a backpropagation algorithm with eight input neurons for eight corresponding features. Neural networks have been used to a considerable extent in mammographic image analysis due to the adaptable nature of neural networks which can be tuned to classify data in any kind of a feature space including the highly nonlinear feature space of mammograms.

II) Fuzzy Logic: Fuzzy logic is based on probabilistic logic, but differing from traditional logic in a way that they can have truth values which are not strictly binary in nature. This works well especially in cases where the truth value does not have a definite description. Fuzzy logic is unique due to the fact that it is internally robust and does not require precise, noise-free inputs [141]. This can be particularly useful in mammographic analysis where noise-free inputs are extremely rare and in many cases, noise in the input is a common feature for mammograms due to the acquisition process and also due to the preprocessing techniques adapted. The rule-based operations in fuzzy logic can be used for quite a number of inputs and more importantly, several outputs or classes, which in some cases can be more than four. In the case of mammograms, this can be useful since five-class problems are quite common in computer-aided diagnosis of mammograms. A comprehensive discussion and application of fuzzy logic to mammographic image analysis can be found in [146].

III) Bayesian Networks: Based on the theory of acyclic graphs, Bayesian networks are probabilistic classifiers which provide an optimal solution to classification problems. Bayesian networks have been used in CAD of breast cancer with good results [147]. Broadly speaking, a Bayesian network is a directed acyclic graph where the nodes correspond to random variables. Each node is associated with a set of conditional probabilities, $P(x_i|A_i)$, where x_i is the variable associated with the specific node and A_i is the set of its parents in the graph [42]. For a Bayesian network to be complete, the marginal probabilities of root nodes and conditional probabilities of nonroot nodes, given their parents for all possible combination of their values, are to be known. Bayesian networks allow efficient calculation of the conditional probability of any node in the graph, given that the values of some other nodes have been observed. This has an efficient implication in the case of CAD in breast cancer detection. For instance, if there are training values with certain nodes indicating the probability of a certain tissue being cancerous or noncancerous, other unknown node values can be found by *probability inference*. The values of the known variables are referred to as the *evidence*, while the conditional probabilities for other variables in the graph can be computed from the evidence. Mathematically, this can be represented as

$$p(x_1, \ldots, x_p) = \prod_{i=1}^{p} p(x_i|\pi_i) \tag{93}$$

where π_i is the set of parents of x_i. For a six-variable problem, the multivariate density can be represented as

$$p(x_1, \ldots\ldots, x_6) = p(x_6|x_4, x_5)p(x_5|x_3)p(x_4|x_1, x_3)p(x_3) \times p(x_2|x_1)p(x_1). \tag{94}$$

Training of a Bayesian network is a two-step process. The first step is to learn the network topology, that can either be fixed by the user who would incorporate details about dependencies or by using optimization techniques based on the training set. Once the topology is designed, conditional probabilities and marginal probabilities are estimated from the available training data points. A practical example of a two level Bayesian classifier scheme has been discussed in [148] for the case of tumor class classification.

IV) Decision Trees: Decision trees are multistage decision making processes where instead of using a complete set of features jointly to make a decision, different subsets of features are used at different levels of the tree. In this technique, decisions on choosing the classes are done in a sequential manner. Decision trees are often referred to as nonlinear classifiers. Decision trees greatly resemble Bayesian networks in their structure. Decision trees start from one parent node and continue splitting until a final result is obtained. Splitting criteria are set for each of the nodes and adherence to the splitting criterion is necessary for node splits. With each split of the node, the node is declared as a leaf and a particular class label is to be given to this node. To label this node, rules such as the majority rule can be used

$$j = \arg \max_i P(\omega_i|t). \tag{95}$$

The relation states that a leaf t is assigned to that class to which a majority of vectors in the node tree belongs. Quite a few authors have used decision trees for breast cancer analysis [149]. Combinations of fuzzy sets have been shown to improve the accuracy of decision trees as seen from the discussions in [150]. Due to the fact that mammograms are highly overlapping in a nonlinear feature space, decision trees perform considerably well as seen from Table I. This is due to the inherent properties of decision trees where they tend to act more robust in conditions where the feature vectors are highly overlapping [149].

V) k-Means Clustering: k-means clustering aims at segmenting the available data into k clusters so that the within-group sum of squares is minimized. The functioning of k-means is pretty simple which is the assignment of objects to the group whose mean is closest to its value. The closeness is usually calculated as an Euclidean distance. Allocation of samples take place individually on a sample-by-sample basis instead of taking place after sorting through the whole data set. Mathematically, a sample x_i is assigned to a group β if

$$\frac{n_\alpha}{n_\alpha - 1} d_{i\alpha}^2 > \frac{n_\beta}{n_\beta + 1} d_{i\beta}^2 \tag{96}$$

where $d_{i\alpha}$ is the distance to the αth centroid and n_α is the group number. The greatest decrease in the sum-squared error is achieved by choosing the group for which $n_\beta d_{i\beta}^2/(n_\beta + 1)$ is a minimum. Usage of k-means clustering is extremely useful in mammographic analysis since the number of classes in mammograms are usually limited to less than four with three classes being the norm in most of the cases. In such instances, with clearly distinguishable intensity levels, it is quite intuitive that

k-means clustering would provide a considerably good classification accuracy level as seen from [151]. An improvement on normal k-means technique is shown in [152] by usage of clustering techniques with adaptive region growing.

VII. Real-Time CAD of Breast Cancer

With the previous mathematical techniques and pipeline, we can discuss the general flow of real-time breast cancer detection used by several authors. Double reading of mammograms can prove to be almost impossible given the limited amount of human resources. As a result, CAD is often proposed as a second reader in most occasions. CAD systems usually point out to suspicious areas which can then be read more carefully by the radiologist aiding him in the diagnostic procedure. A typical CAD session consists of the following steps [153].

1) First reading of the mammograms is carried out by the radiologist to find out suspicious clues in the scan.
2) The CAD system then scans the mammograms to detect suspicious features. This procedure acts as a second reading by pointing out anomalous features in the mammogram.
3) The radiologist then analyzes prompts given by the mammogram about suspicious regions.

CAD systems also provide a numerical estimate of the likelihood of a lesion being benign or malignant, which can later be corroborated by an expert radiologist. Commercial CAD systems have started entering the market with an aim of tapping the potential of using these systems for screening and diagnosis [154], [155]. Though the potential and benefits of these systems are still in discussion, it is nevertheless an important step ahead in the usage of computers for breast cancer detection.

The advantages of using a CAD system have been studied by [156]. In this study, 12 860 screening mammograms were analyzed with the help of a CAD system and it was found that, compared to visual inspection, early-stage malignancy detection increased from 73% to 78% and the number of cancers detected increased by 19.5%. The results of other studies which have been conducted to verify the sensitivity and accuracy of computer-aided detection of breast cancer can be found in [157]–[160]. These systems, in general, try to detect masses, calcifications and architectural distortions in screening mammography. All the aforementioned studies acknowledge the fact that CAD systems are able to detect and mark potentially malignant cells that are overlooked by radiologists.

A contradictory view point is provided by [161] where the authors argue that a group of 24 radiologists were not able to find any considerable increase in mammography recall and cancer detection rates after introduction of a CAD system into a clinical radiology practice in an academic setting. Reference [162] concluded their study about the effectiveness of a CAD system with a similar finding. Reference [163] conducted a study with senior and junior radiologists to find the difference a CAD system made in their ability to detect cancer. It was found that with the help of a CAD system while the sensitivity increased by 8% for senior radiologists, it increased by a numerically significant 22% in the case of junior radiologists. Reference [164] provided a new point of view by considering sensitivity of reading craniocaudal (CC) views and mediolateral oblique (MLO) views as separate parameters in evaluation of the system. According to this study, it has been documented that the craniocaudal view is more sensitive than the MLO view. Though there are opposing views about the usage of computer-aided systems, it is clear that positive reviews clearly outweigh negative reviews. But it is imperative to note that studies indicating that CAD systems do not improve accuracy rates do mention the fact that it helps increase true positive percentages.

VIII. Combining Classifiers

It has been observed that every classifier has its own advantages in classifying specific data [92]. This is due to the fact that every classifier has its feature space where it performs the best. Hence, it is a common practice to combine classifier outputs from different classifiers to improve final classification accuracy. Yet another reason to combine classifiers is the fact that designing a classifier specific to the data set might not result in a good generalization of the classifier. The goal of any pattern recognition system must be to classify unknown data sets in the best possible manner. Hence, optimizing a classifier might not provide the most optimal results in an unknown test set. Hence, classifiers can indeed be combined to provide a better generalization capability for the recognition system. Several classifier combination schemes including parallel, serial combination or cascading, hierarchal or trees, have been proposed [176]. Most combination schemes in literature are of the parallel architecture type which involves invoking independent classifiers individually and then combining their results by using a combination scheme [38]. There are gated variants of the parallel combination scheme where the outputs of individual classifiers are weighted by a gating mechanism before they are combined.

In parallel combination methods, we have several simple methods such as the geometric averaging rule, arithmetic averaging rule and the majority voting rule. In the geometric averaging rule, $P(\omega_i|x)$ is chosen in such a way that the average Kullback-Leibler distance between probabilities is minimized

$$D_{\text{average}} = \frac{1}{L}\sum_{j=1}^{L} D_j \tag{97}$$

where

$$D_j = \sum_{i=1}^{M} P(\omega_i|x) ln \frac{P(\omega_i|x)}{P_j(\omega_i|x)} \tag{98}$$

with $\sum_{i=1}^{M} P_j(\omega_i|x) = 1$. Similarly, arithmetic averaging works in the same way as (98), with the difference that the ratio inside the summation works out to be $ln(P_j(\omega_i|x)/P(\omega_i|x))$ instead of the form seen in geometric averaging.

While the averaging rules are brute force methods, simpler and robust techniques such as the majority voting rule are used too. The majority voting rule is simple, where the result depends on the class for which there is a consensus when at

least l_c of the classifiers agree on the class label of the unknown pattern. Though majority voting has been widely used in combining classifiers, it has been shown that in the case of combining dependent classifiers, there is no definite proof that combination through majority voting can improve final classification performance [177]. Classifier combinations can be highly useful if the individual classifiers are independent [38]. Instead of using different training sets for individual classifiers, various resampling techniques like bootstrapping can be used to virtually induce that kind of a scenario in the classification system. Boosting and bagging are methods which can be used to achieve this goal [178], [179]. Bagging is the technique by which different datasets are created by bootstrapping of the original dataset, which are then combined by a combination rule. Reference [38] provides an exhaustive study on the theoretical analysis of combination schemes in a practical sense. In the case of mammographic image analysis, not many combination schemes are seen. The reason is not discussed in literature. A few works which have discussed classifier combination can be found in [180]–[182]. Though, it seems plausible that classifier schemes can work for mammograms, it can be deduced that since mammographic image analysis consists of at most three to five classes, classifier combination might actually be a burden on the computational scheme given the meager percentage of accuracy increase it might produce.

IX. Conclusion

From the various results and discussions seen in this paper, we find that the results of CAD, though encouraging, are not yet conclusive enough to warrant a credible clinical usage. Literature has been provided to show that the accuracy of cancer detection has indeed improved with introduction of CAD-based diagnostic procedures. But, there is still a long way to go for implementation of the same in a clinical setting. This can be seen from Table I with results from various feature extraction and classification techniques. Here, the best results obtained are around 90%, which is not sufficient enough for implementation in clinical trials. Though better results have been reported in literature, it is noted that results better than these are obtained on specific datasets which cannot be generalized to a wide array of data which can be seen in actual practice. It is also seen that the performance of conventional methods such as decision tree with texture features do not provide results as good as nonconventional techniques, such as Eigen faces approach or a model-based vision algorithm and region based edge-profile acutance measure. This can be due to the fact that conventional techniques are tuned to act on specific datasets while nonconventional techniques developed with the nature of mammographic datasets in mind can be well adapted to a wide array of mammographic data [104]–[109].

Though the idea of using computer-aided detection is gaining popularity, it should not be missed that CAD techniques can serve only as a double-reading aid and cannot replace human readers. This assumes great significance in places where expert radiologists cannot be present. As seen from literature, junior radiologists are prone to making more errors than senior radiologists. In this case, CAD-based readings can provide an improved diagnostic accuracy for radiologists. The main goal of CAD must be to increase diagnostic accuracy with advanced mathematical and computational techniques. Though several advances in this respect have been made in the past 30 years, with mathematical advances and recent improvements in computing techniques and improved speed of computation, we can be sure that techniques which were previously hard to implement in a real-time setting can now be implemented easily.

References

[1] W. Schulz, *Molecular Biology of Human Cancers*. New York, USA: Springer, 2007.

[2] M. P. Sampat, M. K. Markey, and A. C. Bovik, *Computer-Aided Detection and Diagnosis in Mammography, Handbook of Image and Video Processing*. London, U.K.: Elsevier, 2003.

[3] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Syst. With Applicat.*, vol. 36, no. 2, pt. 2, pp. 3240–3247, Mar. 2009, 10.1016/j.eswa.2008.01.009, 0957-4174.

[4] J. Eric, L. M. Wun, C. C. Boring, W. Flanders, J. Timmel, and T. Tong, "The lifetime risk of developing breast cancer," *JNCI J. Nat. Cancer Inst.*, vol. 85, no. 11, pp. 892–897, 1993, 10.1093/jnci/85.11.892.

[5] B. E. Sirovich and H. C. Sox, Jr., "Breast cancer screening," *Surgical Clinics N. Amer.*, vol. 79, no. 5, pp. 961–990, Oct. 1, 1999, 10.1016/S0039-6109(05)70056-6, 0039-6109.

[6] Cancer Facts and Figures 2012 American Cancer Society, Inc..

[7] K. Kerlikowske, P. A. Carney, B. Geller, M. T. Mandelson, S. H. Taplin, K. Malvin, V. Ernster, N. Urban, G. Cutter, R. Rosenberg, and R. B. Barbash, "Performance of screening mammography among Women with and without a first-degree relative with breast cancer," *Ann. Intern. Med.*, vol. 133, pp. 855–863, Dec. 5, 2000.

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

[9] R. G. Blanks, M. G. Wallis, and S. M. Moss, "A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: Results from the UK National Health Service breast screening programme," *J. Med. Screening*, vol. 5, no. 4, pp. 195–201, 1998.

[10] C. J. Vyborny, M. L. Giger, and R. M. Nishikawa, "Computer aided detection and diagnosis of breast cancer," *Radiologic Clinics N. Amer.*, vol. 38, no. 4, pp. 725–740, Jul. 1, 2000, 10.1016/S0033-8389(05)70197-4, 0033-8389.

[11] J. Scharcanski and C. R. Jung, "Denoising and enhancing digital mammographic images for visual screening," *Computerized Medical Imag. Graphics*, vol. 30, no. 4, pp. 243–254, Jun. 2006, 10.1016/j.compmedimag.2006.05.002, 0895-6111.

[12] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms," *Int. J. Pattern Recog. Artif. Intell.*, vol. 7, pp. 1357–1376, 1993.

[13] M. L. Giger, "Computer-aided diagnosis in radiology," *Acad. Radiol.*, vol. 9, pp. 1–3, 2002.

[14] K. J. McLoughlin, P. J. Bones, and N. Karssemeijer, "Noise equalization for detection of microcalcification clusters in direct digital mammogram images," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 313–320, Mar. 2004, 10.1109/TMI.2004.824240.

[15] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc, Int. Conf. Image Processing*, Sep. 16–19, 1996, vol. 1, pp. 379–382, 10.1109/ICIP.1996.559512.

[16] C. D. Maggio, "State of the art of current modalities for the diagnosis of breast lesions," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 31, pp. S56–S69, 2004, (Suppl. 1).

[17] M. L. Giger, N. Karssemeijer, and S. G. Armato, "Guest editorial computer-aided diagnosis in medical imaging," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1205–1208, Dec. 2001, 10.1109/TMI.2001.974915.

[18] M. M. Kivanc, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999, 10.1109/97.803428.

[19] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field *a priori* model," *IEEE Trans. Image Process.*, vol. 6, no. 4, pp. 549–565, Apr. 1997, 10.1109/83.563320.

[20] A. K. Hackshaw and E. A. Paul, "Breast self-examination and death from breast cancer: A meta-analysis," *Br. J. Cancer*, vol. 88, pp. 1047–1053, 2003.

[21] H. Yoshida, Z. Wei, W. Cai, K. Doi, R. M. Nishikawa, and M. L. Giger, "Optimizing wavelet transform based on supervised learning for detection of microcalcifications in digital mammograms," in *Proc. Int. Conf. Image Processing* , Oct. 23–26, 1995, vol. 3, pp. 152–155, 10.1109/ICIP.1995.537603.

[22] A. F. Laine, S. Schuler, J. Fan, and W. Huda, "Mammographic feature enhancement by multiscale analysis," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 725–740, Dec. 1994, 10.1109/42.363095.

[23] P. Heinlein, J. Drexl, and W. Schneider, "Integrated wavelets for enhancement of microcalcifications in digital mammography," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 402–413, Mar. 2003, 10.1109/TMI.2003.809632.

[24] B. Cady and M. Chung, "Mammographic screening: No longer controversial," *Amer. J. Clinical Oncol.*, vol. 28, no. 1, pp. 1–4, Feb. 2005.

[25] R. C. Gonzalez and R. E. Woods*, Digital Image Processing*. Reading, MA, USA: Addison-Wesley Longman, 1992, 0201508036.

[26] E. A. Sickles, "Mammographic features of 300 consecutive nonpalpable breast cancers," *Amer. J. Roentgenol.*, vol. 146, no. 4, pp. 661–663, 1986.

[27] G. Boccignone, A. Chianese, and A. Picariello, "Computer aided detection of microcalcifications in digital mammograms," *Computers Biol. Medicine*, vol. 30, no. 5, pp. 267–286, Sep. 1, 2000, 10.1016/S0010-4825(00)00014-7, 0010-4825.

[28] A. Papadopoulos, D. I. Fotiadis, and L. Costaridou, "Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques," *Computers Biol. Medicine*, vol. 38, no. 10, pp. 1045–1055, Oct. 2008, 10.1016/j.compbiomed.2008.07.006, 0010-4825.

[29] C. Burrell, D. M. Sibbering, A. R. M. Wilson, S. E. Pinder, A. J. Evans, L. J. Yeoman, C. W. Elston, I. O. Ellis, R. W. Blamey, and J. F. R. Robertson, "Screening interval breast cancers: Mammographic features and prognostic factors," *Radiology*, vol. 199, no. 4, pp. 811–817, 1996.

[30] A. Laine, J. Fan, and W. Yang, "Wavelets for contrast enhancement of digital mammography," *IEEE Eng. Medicine Biol.*, vol. 14, no. 5, pp. 536–550, Sep./Oct. 1995, 10.1109/51.464770.

[31] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 7, pp. 710–732, Jul. 1992, 10.1109/34.142909.

[32] J. D. Fahnestock and R. A. Schowengerdt, "Spatially variant contrast enhancement using local range modification," *Optical Eng.*, vol. 22, pp. 378–381, May–Jun. 1983.

[33] H. Jing, Y. Yang, and R. M. Nishikawa, "Detection of clustered microcalcifications using spatial point process modeling," *Phys. Med. Biol.*, vol. 56, no. 1, 2011, 10.1088/0031-9155/56/1/001.

[34] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, no. 5, pp. 369–378, May 1999, 10.1109/42.774164.

[35] E. D. Pisano, E. B. Cole, B. M. Hemminger, M. J. Yaffe, S. R. Aylward, A. D. A. Maidment, R. E. Johnston, M. B. Williams, L. T. Niklason, E. F. Conant, L. L. Fajardo, D. B. Kopans, M. E. Brown, and S. M. Pizer, "Image processing algorithms for digital mammography: A pictorial essay," *Radiographics*, vol. 20, pp. 1479–1491, Sep. 2000.

[36] N. Jamal, K. H. Ng, and D. McLean, "A study of mean glandular dose during diagnostic mammography in Malaysia and some of the factors affecting it," *Br. J. Radiol*, vol. 76, pp. 238–245, 2003.

[37] N. Jamal, K.-H. Ng, D. McLean, L.-M. Looi, and F. Moosa, "Mammographic breast glandularity in Malaysian women derived from radiographic data," *Amer. J. Roentgenol.*, vol. 182, pp. 713–717, 2004.

[38] A. K. Jain, R. P. W. Duin, and J. C. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000, 10.1109/34.824819.

[39] A. R. Webb*, Statistical Pattern Recognition*, 2nd ed. New York, NY, USA: Wiley, 2002, 0470845147.

[40] J. Scharcanski and C. R. Jung, "Denoising and enhancing digital mammographic images for visual screening," *Computerized Med. Imag. Graphics*, vol. 30, no. 4, pp. 243–254, Jun. 2006, 10.1016/j.compmedimag.2006.05.002, 0895-6111.

[41] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms," *Int. J. Pattern Recog. Artif. Intell.*, vol. 7, pp. 135713–135776, 1993.

[42] S. Theodoridis and K. Koutroumbas*, Pattern Recognition*, 3rd ed. New Providence, NJ, USA: Elsevier, 2006, 0-12-369531-7.

[43] A. Petrosian, H. P. Chan, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computer-aided diagnosis in mammography: Classification of mass and normal tissue by texture analysis," *Phys. Med. Biol.*, vol. 39, pp. 2273–2288, 1994.

[44] R. M. Haralick, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, pp. 610–621, Dec. 1973.

[45] R. M. Rangayyan, N. M. El-Faramawy, J. E. L. Desautels, and O. A. Alim, "Measures of acutance and shape for classification of breast tumours," *IEEE Trans. Med. Imag.*, vol. 16, no. 12, pp. 799–810, Dec. 1997.

[46] N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms," *IEEE Trans. Med. Imag.*, vol. 15, no. 10, pp. 611–619, Oct. 1996.

[47] Guliato, R. M. Rangayyan, W. A. Carnielli, J. A. Zuffo, and J. E. L. Desautels, "Segmentation of breast tumors in mammograms by fuzzy region growing," in *Proc. 20th Annu. Int. Conf. IEEE Engineering Medicine Biology Soc.*, Hong Kong, Oct. 29–Nov. 1 1998, pp. II:1002–II:1004.

[48] S. L. Kok, J. M. Brady, and L. Tarassenko, "The detection of abnormalities in mammograms," in *Proc. 2nd Int. Workshop Digital Mammography*, York, U.K., Jul. 10–12, 1994, pp. 261–270.

[49] S. Timp, C. Varela, and N. Karssemeijer, "Temporal change analysis for characterization of mass lesions in mammography," *IEEE Trans. Med. Imag.*, vol. 26, no. 7, pp. 945–953, Jul. 2007, 10.1109/TMI.2007.897392.

[50] L. Ke, N. Mu, and Y. Kang, "Mass computer-aided diagnosis method in mammogram based on texture features," in *Proc. 3rd Int. Conf. Biomedical Engineering and Informatics (BMEI)*, Oct. 16–18, 2010, vol. 1, pp. 354–357, 10.1109/BMEI.2010.5639515.

[51] S. H. Amroabadi, M. R. Ahmadzadeh, and A. Hekmatnia, "Mass detection in mammograms using GA based PCA and Haralick features selection," in *Proc. 19th Iranian Conf. Electrical Engineering (ICEE)*, May 17–19, 2011, p. 1.

[52] A. Papoulis*, Probability, Random Variables and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[53] K.-H. Ng, N. Jamal, and L. DeWerd, "Global quality control perspective for the physical and technical aspects of screen-film. Mammography image quality and radiation dose," *Radiation Protection Dosimetry*, vol. 121, no. 4, pp. 445–451, 2006.

[54] C. Cerjan, "The Zernike-Bessel representation and its application to Hankel transforms," *J. Opt. Soc. Amer. A*, vol. 24, no. 6, p. 1609, 2007.

[55] N. Jamal, K. H. Ng, L. M. Looi, D. McLean, A. Zulfiqar, S. P. Tan, W. F. Liew, A. Shantini, and S. Ranganathan, "Quantitative assessment of breast density from digitized mammograms into Tabar's patterns," *Phys. Med. Biol.*, vol. 51, pp. 1–15, 2006.

[56] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[57] N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels, "Gradient and texture analysis for the classification of mammographic masses," *IEEE Trans. Med. Imag.*, vol. 19, no. 10, pp. 1032–1043, Oct. 2000, 10.1109/42.887618.

[58] R. Mousa, Q. Munib, and A. Moussa, "Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural," *Expert Systems With Applicat.*, vol. 28, pp. 713–723, 2005.

[59] R. Sarunas and P. Vitalijus, "On dimensionality, sample size, classification error, and complexity of classification algorithm in pattern recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 3, pp. 242–252, May 1980, 10.1109/TPAMI.1980.4767011.

[60] A. K. Jain, R. P. W. Duin, and J. C. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 1, pp. 4–37, Jan 2000, 10.1109/34.824819.

[61] S. Kulback and R. A. Liebler, "On information and sufficiency," *Ann. Math. Statistics*, vol. 22, pp. 79–86, 1951.

[62] S. Watanabe*, Pattern Recognition: Human and Mechanical*. New York, NY, USA: Wiley, 1995.

[63] R. Webb*, Statistical Pattern Recognition*, 2nd ed. New York, NY, USA: Wiley, 2002, 0470845147.

[64] M. C. Thomas, "The best two independent measurements are not the two best," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-4, no. 1, pp. 116–117, Jan. 1974, 10.1109/TSMC.1974.5408535.

[65] Koutroumbas*, Pattern Recognition*, 3rd ed. New York, NY, USA: Elsevier, 2006, 0-12-369531-7.

[66] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms," *Computerized Med. Imag. Graphics*, vol. 26, no. 5, pp. 309–319, Sep.–Oct. 2002, 10.1016/S0895-6111(02)00031-9, 0895-6111.

[67] O. Tsujii, M. T. Freedman, and S. K. Mun, "Classification of microcalcifications in digital mammograms using trendoriented radial basis function neural network," *Pattern Recog.*, vol. 32, pp. 891–903, 1999.

[68] A. P. Dhawan, Y. Chitre, and C. Kaiser-Bonasso, "Analysis of mammographic microcalcifications using gray-level image structure features," *IEEE Trans. Med. Imag.*, vol. 15, no. 3, pp. 246–259, Jun. 1996, 10.1109/42.500063.

[69] A. Papoulis, Probability, random variables and stochastic processes.
[70] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory*, vol. 8, no. 2, pp. 179–187, 1962.
[71] L. Devroye, L. Gyorfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. New York, NY, USA: Springer-Verlag, 1996.
[72] L. M. Bruce and R. R. Adhami, "Classifying mammographic mass shapes using the wavelet transform modulus-maxima method," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 1170–1177, Aug. 1999.
[73] M. M. Eltoukhy, I. Faye, and B. B. Samir, "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation," *Computers Biol. Medicine*, vol. 42, no. 1, pp. 123–128, Jan. 2012, 10.1016/j.compbiomed.2011.10.016, 0010-4825.
[74] M. M. Eltoukhy, I. Faye, and B. B. Samir, "A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram," *Computers Biol. Medicine*, vol. 40, no. 4, pp. 384–391, Apr. 2010, 10.1016/j.compbiomed.2010.02.002, 0010-4825.
[75] C. Heang-Ping, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. Sanjay-Gopal, "Improvement of Radiologists' characterization of mammographic masses by using computer-aided diagnosis: An ROC study," *Radiology*, vol. 212, pp. 817–827, Sep. 1999.
[76] B. Sahiner, C. Heang-Ping, N. Petrick, W. Datong, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images," *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 598–610, Oct. 1996, 10.1109/42.538937.
[77] S. Watanabe, *Pattern Recognition: Human and Mechanical*. New York, NY, USA: Wiley, 1995.
[78] L. Hadjiiski, B. Sahiner, and H. P. Chan, "Classification of malignant and benign masses based on hybrid ART2LDA approach," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 1178–1187, Aug. 1999.
[79] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proc. 12th IAPR Int. Conf. Pattern Recognition Vol. 2—Conference B: Computer Vision Image Processing*, Oct. 9–13, 1994, vol. 2, pp. 279–283, 10.1109/ICPR.1994.576920.
[80] Z. Huo, M. L. Giger, C. J. Vyborny, and C. E. Metz, "Breast cancer: Effectiveness of computer-aided diagnosis—Observer study with independent database of mammograms," *Radiology*, vol. 224, pp. 560–568, Aug. 2002, 10.1148/radiol.2242010703.
[81] Sanjay-Gopal, "Improvement of Radiologists' characterization of mammographic masses by using computer-aided diagnosis: An ROC study," *Radiology*, vol. 212, pp. 817–827, Sep. 1999.
[82] M. A. Alolfe, W. A. Mohamed, A.-B. M. Youssef, A. S. Mohamed, and Y. M. Kadah, "Computer aided diagnosis in digital mammography using support vector machine and linear discriminant analysis classification," in *Proc. 16th IEEE Int. Conf. Image Processing (ICIP)*, Nov. 7–10, 2009, pp. 2609–2612, 10.1109/ICIP.2009.5413992.
[83] J. A. Anderson, "Logistic discrimination," in *Handbook of Statistics*. North Holland, Amsterdam: Springer-Verlag, 1982, vol. 2, pp. 169–191.
[84] *American College of Radiology, ACR BI-RADS- Mammography, Ultrasound & Magnetic Resonance Imaging*, 4th ed. Reston, VA: American College of Radiology, 2003.
[85] C. Marrocco, M. Molinara, C. D'Elia, and F. Tortorella, "A computer-aided detection system for clustered microcalcifications," *Artificial Intell. Medicine*, vol. 50, no. 1, pp. 23–32, Sep. 2010, 10.1016/j.artmed.2010.04.007, 0933-3657.
[86] Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," *JNCI J. Nat. Cancer Inst.*, vol. 96, no. 3, pp. 185–190, 2004, 10.1093/jnci/djh067.
[87] B. Sahiner, C. Heang-Ping, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," *Med. Phys.*, vol. 25, p. 516, 1998.
[88] A. Retico, P. Delogu, M. E. Fantacci, A. P. Martinez, A. Stefanini, and A. Tata, "A scalable computer-aided detection system for microcalcification cluster identification in a pan-European distributed database of mammograms," *Nuclear Instrum. Methods Phys. Res. Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 569, no. 2, pp. 601–605, Dec. 20, 2006, 10.1016/j.nima.2006.08.094, 0168-9002.
[89] Y. Sung-Nien and H. Yu-Kun, "Detection of microcalcifications in digital mammograms using combined model-based and statistical textural features," *Expert Syst. Applicat.*, vol. 37, no. 7, pp. 5461–5469, Jul. 2010, 10.1016/j.eswa.2010.02.066, 0957-4174.
[90] M. De Santo, M. Molinara, and F. Tortorella, "Automatic classification of clustered microcalcifications by a multiple expert system," *Patt. Recog.*, vol. 36, pp. 1467–1477, 2003.
[91] Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a Community Breast Center," *Radiology*, vol. 220, pp. 781–786, Sep. 2001, 10.1148/radiol.2203001282.
[92] L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Trans. Med. Imag.*, vol. 24, no. 3, pp. 371–380, Mar. 2005, 10.1109/TMI.2004.842457.
[93] S. Pohlman, K. A. Powell, and N. A. Obuchowski, "Quantitative classification of breast tumors in digitized mammograms," *Med. Phys.*, vol. 23, pp. 1337–1345, 1996.
[94] Castellino, Potential contribution of computer-aided detection to the sensitivity of screening mammography radiology May 2000, vol. 215, no. 2, pp. 554–562.
[95] Castellino, Invasive lobular carcinoma of the breast: Mammographic characteristics and computer-aided detection radiology Oct. 2002, vol. 225, no. 1, pp. 182–189, 10.1148/radiol.2251011029.
[96] S. D. Tzikopoulos, M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, and S. Theodoridis, "A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry," *Computer Methods Programs Biomedicine*, vol. 102, no. 1, pp. 47–63, Apr. 2011, 10.1016/j.cmpb.2010.11.016, 0169-2607.
[97] Bovik, "Computer-Aided detection and diagnosis in mammography," in *Handbook of Image and Video Processing*. New York, NY, USA: Elsevier, 2003.
[98] Indovina, "Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography," *Eur. J. Radiol.*, vol. 45, no. 2, pp. 135–138, Feb. 2003, 10.1016/S0720-048X(02)00011-6, 0720-048X.
[99] W. Chia-Hung, S. Y. Chen, and X. Liu, "Mammogram retrieval on similar mass lesions," *Computer Methods Programs Biomedicine*, Oct. 8, 2010, 10.1016/j.cmpb.2010.09.002, 0169-2607.
[100] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer Verlag, 1995.
[101] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory and Methods*. New York, NY, USA: Wiley, 1998.
[102] R. Courant and D. Hilbert, *Methods of Mathematical Physics*. New York, NY, USA: Interscience, 1953.
[103] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[104] "Computer-Aided detection (CAD) in screening mammography: Sensitivity of commercial CAD systems for detecting architectural distortion," *AJR*, vol. 181, pp. 1083–1088, October 2003.
[105] L. Chen-Chung, T. Chung-Yen, T. Ta-Shan, and Y. Shyr-Shen, "An improved GVF snake based breast region extrapolation scheme for digital mammograms," *Expert Systems Applicat.*, vol. 39, no. 4, pp. 4505–4510, Mar. 2012, 10.1016/j.eswa.2011.09.136, 0957-4174.
[106] K. Doi, H. MacMahon, S. Katsuragawa, R. M. Nishikawa, and Y. Jiang, "Computer-aided diagnosis in radiology: Potential and pitfalls," *Eur. J. Radiol.*, vol. 31, no. 2, pp. 97–109, Aug. 1999, 10.1016/S0720-048X(99)00016-9, 0720-048X.
[107] T. Nan-Chyuan, C. Hong-Wei, and H. Sheng-Liang, "Computer-aided diagnosis for early-stage breast cancer by using wavelet transform," *Computerized Med. Imag. Graphics*, vol. 35, no. 1, pp. 1–8, Jan. 2011, 10.1016/j.compmedimag.2010.08.005, 0895-6111.
[108] K. Doi, "Overview on research and development of computer-aided diagnostic schemes," *Seminars in Ultrasound, CT, and MRI*, vol. 25, no. 5, pp. 404–410, Oct. 2004, 10.1053/j.sult.2004.02.006, 0887-2171.
[109] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Med. Imag. Graphics*, vol. 31, no. 4–5, pp. 198–211, Jun.–Jul. 2007, 10.1016/j.compmedimag.2007.02.002, 0895-6111.
[110] H. D. Cheng and M. Cui, "Mass lesion detection with a fuzzy neural network," *Pattern Recognition*, vol. 37, no. 6, pp. 1189–1200, 2004.
[111] D. B. Fogel, E. C. Wasson, III, E. M. Boughton, and V. W. Porto, "Evolving artificial neural networks for screening features from mammograms," *Artif. Intell. Med.*, vol. 14, no. 3, pp. 317–326, Nov. 1998.
[112] R. M. Rangayyan, F. J. Ayres, and J. E. L. Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *J. Franklin Inst.*, vol. 344, no. 3–4, pp. 312–348, May–Jul. 2007, 10.1016/j.jfranklin.2006.09.003, 0016-0032.
[113] S. Katsuragawa and K. Doi, "Computer-aided diagnosis in chest radiography," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4–5, pp. 212–223, Jun.–Jul. 2007, 10.1016/j.compmedimag.2007.02.003, 0895-6111.

[114] B. Verma and P. Zhang, "A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms," *Applied Computing*, vol. 7, no. 2, pp. 612–625, Mar. 2007, 10.1016/j.asoc.2005.02.008, 1568-4946.

[115] S. Timp and N. Karssemeijer, "Interval change analysis to improve computer aided detection in mammography," *Med. Image Anal.*, vol. 10, no. 1, pp. 82–95, Feb. 2006, 10.1016/j.media.2005.03.007, 1361-8415.

[116] S. Dua, H. Singh, and H. W. Thompson, "Associative classification of mammograms using weighted rules," *Expert Systems Applicat.*, vol. 36, no. 5, pp. 9250–9259, Jul. 2009, 10.1016/j.eswa.2008.12.050, 0957-4174.

[117] A. Papadopoulos, D. I. Fotiadis, and A. Likas, "An automatic microcalcification detection system based on a hybrid neural network classifier," *Artificial Intell. Medicine*, vol. 25, no. 2, pp. 149–167, Jun. 2002, 10.1016/S0933-3657(02)00013-1, 0933-3657.

[118] A. Keleş, A. Keleş, and U. Yavuz, "Expert system based on neuro-fuzzy rules for diagnosis breast cancer," *Expert Syst. Applicat.*, vol. 38, no. 5, pp. 5719–5726, May 2011, 10.1016/j.eswa.2010.10.061, 0957-4174.

[119] P. Zhang, B. Verma, and K. Kumar, "Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection," *Pattern Recognition Lett.*, vol. 26, no. 7, pp. 909–919, May 15, 2005, 10.1016/j.patrec.2004.09.053, 0167-8655.

[120] C. Varela, P. G. Tahoces, A. J. Méndez, M. Souto, and J. J. Vidal, "Computerized detection of breast masses in digitized mammograms," *Computers Biol. Medicine*, vol. 37, no. 2, pp. 214–226, Feb. 2007, 10.1016/j.compbiomed.2005.12.006, 0010-4825.

[121] R. Mousa, Q. Munib, and A. Moussa, "Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural," *Expert Systems Applicat.*, vol. 28, no. 4, pp. 713–723, May 2005, 10.1016/j.eswa.2004.12.028, 0957-4174.

[122] J. J. James, "The current status of digital mammography," *Clinical Radiol.*, vol. 59, no. 1, pp. 1–10, Jan. 2004, 10.1016/j.crad.2003.08.011, 0009-9260.

[123] S. Timp, C. Varela, and N. Karssemeijer, "Computer-aided diagnosis with temporal analysis to improve radiologists' interpretation of mammographic mass lesions," *IEEE Trans. Inform. Technol. Biomedicine*, vol. 14, no. 3, pp. 803–808, May 2010, 10.1109/TITB.2010.2043296.

[124] Pizer, "Image processing algorithms for digital mammography: A pictorial essay," *Radiographics*, vol. 20, pp. 1479–1491, Sep. 2000.

[125] C. E. Floyd, Jr, J. Y. Lo, A. J. Yun, D. C. Sullivan, and P. J. Kornguth, "Prediction of breast cancer malignancy using an artificial neural network," *Cancer*, vol. 74, no. 11, pp. 2944–2948, 1994.

[126] P. McLeod, B. Verma, and M. Park, "Soft clustering and support vector machine based technique for the classification of abnormalities in digital mammograms," in *Proc. 5th Int. Conf. Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Dec. 7–10, 2009, pp. 185–189, 10.1109/ISSNIP.2009.5416794.

[127] M. L. Giger, "Computer-aided diagnosis of breast lesions in medical images," *Computing Sci. Eng.*, vol. 2, no. 5, pp. 39–45, Sep./Oct. 2000, 10.1109/5992.877391.

[128] E. Song, S. Xu, X. Xu, J. Zeng, Y. Lan, S. Zhang, and H. Chih-Cheng, "Hybrid segmentation of mass in mammograms using template matching and dynamic programming," *Academic Radiol.*, vol. 17, no. 11, pp. 1414–1424, Nov. 2010, 10.1016/j.acra.2010.07.008, 1076-6332.

[129] Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer Verlag, 1995.

[130] L. Devroye, L. Gyorfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. New York, NY, USA: Springer-Verlag, 1996.

[131] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," *Academic Radiol.*, vol. 6, no. 1, pp. 22–33, Jan. 1999, 10.1016/S1076-6332(99)80058-0, 1076-6332.

[132] Y. Yuan, M. L. Giger, H. Li, N. Bhooshan, and C. A. Sennett, "Multimodality computer-aided breast cancer diagnosis with FFDM and DCE-MRI," *Academic Radiol.*, vol. 17, no. 9, pp. 1158–1167, Sep. 2010, 10.1016/j.acra.2010.04.015, 1076-6332.

[133] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," *Academic Radiol.*, vol. 6, no. 1, pp. 22–33, Jan. 1999, 10.1016/S1076-6332(99)80058-0, 1076-6332.

[134] P. McLeod, B. Verma, and M. Park, "Soft clustering and support vector machine based technique for the classification of abnormalities in digital mammograms," in *Proc. 5th Int. Conf. Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Dec. 7–10, 2009, pp. 185–189, 10.1109/ISSNIP.2009.5416794.

[135] J. Tang, R. M. Rangayyan, J. Xu, E. I. Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances," *IEEE Trans. Inform. Technol. Biomedicine*, vol. 13, no. 2, pp. 236–251, Mar. 2009, 10.1109/TITB.2008.2009441.

[136] S. T. Juliette, K. J. Schilling, J. W. Hoffmeister, E. Friedmann, R. McGinnis, and R. G. Holcomb, "Detection of breast cancer with full-field digital mammography and computer-aided detection," *AJR*, vol. 192, pp. 337–340, Feb. 2009, 10.2214/AJR.07.3884.

[137] A. Malich, S. Schmidt, D. R. Fischer, M. Facius, and W. A. Kaiser, "The performance of computer-aided detection when analyzing prior mammograms of newly detected breast cancers with special focus on the time interval from initial imaging to detection," *Eur. J. Radiol.*, vol. 69, no. 3, pp. 574–578, Mar. 2009, 10.1016/j.ejrad.2007.11.038, 0720-048X.

[138] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representation by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press, 1986, vol. 1, pp. 318–362.

[139] Helenon, Computer-aided detection (CAD) in mammography: Does it help the junior or the senior radiologist?.

[140] Y. Jiang, R. M. Nishikawa, E. E. Wolverton, C. E. Metz, M. L. Giger, R. A. Schmidt, and C. J. Vyborny, "Malignant and benign clustered microcalcifications: Automated feature analysis and classification," *Radiology*, vol. 198, pp. 671–678, 1996.

[141] M. Bhattacharya, "Fuzzy logic based segmentation of microcalcification in breast using digital mammograms considering multiresolution," in *Proc. Machine Vision and Image Processing Conf.*, 2007, IMVIP 2007.

[142] N. Petrick, B. Sahiner, C. Heang-Ping, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evaluation of a mass-detection algorithm for computer-aided diagnosis—Experience in 263 patients," *Radiology*, vol. 224, pp. 217–224, Jul. 2002, 10.1148/radiol.2241011062.

[143] Elmore, "Influence of computer-aided detection on performance of screening mammography," *N. Engl. J. Med.*, vol. 356, pp. 1399–1409, 2007.

[144] M. J. Collins, J. Hoffmeister, and S. W. Worrell, "Computer-aided detection and diagnosis of breast cancer," *Seminars in Ultrasound, CT, and MRI*, vol. 27, no. 4, pp. 351–355, Aug. 2006, 10.1053/j.sult.2006.05.009, 0887-2171.

[145] Lopez, "Impact of mammographic breast density on computer-assisted detection (CAD) in a breast imaging department," *Radiology (English Edition)*, vol. 53, no. 5, pp. 456–461, Sep.–Oct. 2011, 10.1016/j.rxeng.2010.08.001, 2173-5107.

[146] K. Boris and T. Evangelos, "Fuzzy logic in computer-aided breast cancer diagnosis: Analysis of lobulation," *Artificial Intell. Medicine*, vol. 11, pp. 75–85, 1997.

[147] B. Zheng, C. Yuan-Hsiang, X. H. Wang, W. F. Good, and D. Gur, "Application of a Bayesian belief network in a computer-assisted diagnosis scheme for mass detection," in *Proc. SPIE*, 1999, vol. 3661, p. 1553.

[148] J. L. Viton, M. Rasigni, G. Rasigni, and A. Llebaria, "Method for characterizing masses in digital mammograms," *Opt. Eng.*, vol. 35, p. 3453, 1996.

[149] L. Zhen and A. K. Chan, "An artificial intelligent algorithm for tumor detection in screening mammogram," *IEEE Trans. Med. Imag.*, vol. 20, no. 7, pp. 559–567, Jul. 2001, 10.1109/42.932741.

[150] L. Lihua, W. Qian, P. C. Laurence, A. C. Robert, and A. J. Thomas, "Improving mass detection by adaptive and multiscale processing in digitized mammograms," in *Proc. SPIE*, vol. 3661, pp. 490–498.

[151] B. C. Patel and G. R. Sinha, "Article: An adaptive K-means clustering algorithm for breast image segmentation," *Int. J. Computer Applicat.*, vol. 10, no. 4, pp. 35–38, Nov. 2010.

[152] H. Zhao, L. Li, W. Xu, and J. Zhang, "A novel clustering method based on K-MEANS with region growing for micro-calcifications in mammographic images," in *Proc. Int. Conf. Computer Information Application (ICCIA)*, Dec. 3–5, 2010, pp. 1–4, 10.1109/ICCIA.2010.6141521.

[153] R. M. Rangayyan, F. J. Ayres, and J. E. L. Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *J. Franklin Inst.*, vol. 344, no. 3–4, pp. 312–348, May–Jul. 2007, 10.1016/j.jfranklin.2006.09.003, 0016-0032.

[154] R2 Technologu [Online]. Available: www.r2tech.com/

[155] iCAD [Online]. Available: www.icadmed.com/

[156] T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology*, vol. 220, pp. 781–786, Sep. 2001, 10.1148/radiol.2203001282.

[157] S. Ciatto, M. R. D. Turco, G. Risso, S. Catarzi, R. Bonardi, V. Viterbo, P. Gnutti, B. Guglielmoni, L. Pinelli, A. Pandiscia, F. Navarra, A. Lauria, R. Palmiero, and P. L. Indovina, "Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography," *Eur. J. Radiol.*, vol. 45, no. 2, pp. 135–138, Feb. 2003, 10.1016/S0720-048X(02)00011-6, 0720-048X.

[158] L. J. W. Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology*, vol. 215, no. 2, pp. 554–562, May 2000.

[159] W. P. Evans, L. J. W. Burhenne, L. Laurie, K. F. O'Shaughnessy, and R. A. Castellino, "Invasive lobular carcinoma of the breast: Mammographic characteristics and computer-aided detection," *Radiology*, vol. 225, no. 1, pp. 182–189, Oct. 2002, 10.1148/radiol.2251011029.

[160] J. A. Baker, E. L. Rosen, J. Y. Lo, E. I. Gimenez, R. Walsh, and M. S. Soo, "Computer-Aided detection (CAD) in screening mammography: Sensitivity of commercial CAD systems for detecting architectural distortion," *AJR*, vol. 181, pp. 1083–1088, Oct. 2003.

[161] D. Gur, J. H. Sumkin, H. E. Rockette, M. Ganott, C. Hakim, L. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," *JNCI J. Nat. Cancer Inst.*, vol. 96, no. 3, pp. 185–190, 2004, 10.1093/jnci/djh067.

[162] J. J. Fenton, S. H. Taplin, P. A. Carney, L. Abraham, E. A. Sickles, C. D'Orsi, E. A. Berns, G. Cutter, R. E. Hendrick, W. E. Barlow, and J. G. Elmore, "Influence of computer-aided detection on performance of screening mammography," *N. Engl. J. Med.*, vol. 356, pp. 1399–1409, 2007.

[163] C. Balleyguier, K. Kinkel, J. Fermanian, S. Malan, G. Djen, P. Taourel, and O. Helenon, "Computer-aided detection (CAD) in mammography: Does it help the junior or the senior radiologist?," *Eur. J. Radiol.*, vol. 54, no. 1, pp. 90–96, Apr. 2005, 10.1016/j.ejrad.2004.11.021, 0720-048X.

[164] C. R. Castellano, C. V. Nuñez, R. C. Boy, A. A. Gil, J. M. P. Varela, and M. B. Lopez, "Impact of mammographic breast density on computer-assisted detection (CAD) in a breast imaging department," *Radiology (English Edition)*, vol. 53, no. 5, pp. 456–461, Sep.–Oct. 2011, 10.1016/j.rxeng.2010.08.001, 2173-5107.

[165] C. Kimme, B. J. O'Loughlin, and J. Sklansky, "Automatic detection of suspicious abnormalities in breast radiographs," in *Data Structures, Computer Graphics, and Pattern Recognition*, A. Klinger, K. S. Fu, and T. L. Kunii, Eds. New York, NY, USA: Academic, 1977, pp. 427–447.

[166] A. Petrosian, H. P. Chan, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computer-aided diagnosis in mammography: Classification of mass and normal tissue by texture analysis," *Phys. Med. Biol.*, vol. 39, pp. 2273–2288, 1994.

[167] S. K. Kinoshita, P. M. A. Marques, A. F. F. Slates, H. R. C. Marana, R. J. Ferrari, and R. L. Villela, "Detection and characterization of mammographic masses by artificial neural network," in *Proc. 4th Int. Workshop Digital Mammography*, pp. 489–490.

[168] W. E. Polakowski, D. A. Cournoyer, S. K. Rogers, M. P. DeSimio, D. W. Ruck, J. W. Hoffmeister, and R. A. Raines, "Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency," *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 811–819, Dec. 1997, 10.1109/42.650877.

[169] C. E. Priebe, R. A. Lorey, D. J. Marchette, J. L. Solka, and G. W. Rogers, "Nonparametric spatio-temporal change point analysis for early detection in mammography," in *Proc. 2nd Int. Workshop Digital Mammography*, Jul. 10–12, 1994, pp. 111–120.

[170] M. Sameti, J. Morgan-Parkes, R. K. Ward, and B. Palcic, "Classifying image features in the last screening mammograms prior to detection of a malignant mass," in *Proc. 4th Int. Workshop Digital Mammography*, Nijmegen, The Netherlands, Jun. 1998, pp. 127–134.

[171] Y. Chitre, A. P. Dhawan, and M. Moskowitz, "Artificial neural network based classification of mammographic microcalcifications using image structure features," *Int. J. Pattern Recogn. Artif. Intell.*, vol. 7, no. 12, pp. 1377–1402, 1993.

[172] B. Verma and J. Zakos, "A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques," *IEEE Trans. Inform. Technol. Biomedicine*, vol. 5, no. 1, pp. 46–54, Mar. 2001, 10.1109/4233.908389.

[173] H. Yoshida, R. Nishikawa, G. Maryellen, and K. Doi, "Computer-aided diagnosis in mammography: Detection of clustered microcalcifications based on multiscale edge representation," in *Computer Assisted Radiology*. Amsterdam, The Netherlands: Elsevier, 1996.

[174] A. Oliver, J. Marti, R. Marti, A. Bosch, and J. Freixenet, "A new approach to the classification of mammographic masses and normal breast tissue," in *Proc. 18th Int. Conf. Pattern Recognition*, vol. 4, pp. 707–710, 10.1109/ICPR.2006.113, 0-0 0.

[175] N. Szekely, N. Toth, and B. Pataki, "A hybrid system for detecting masses in mammographic images," *IEEE Trans. Instrumentation Measurement*, vol. 55, no. 3, pp. 944–952, Jun. 2006, 10.1109/TIM.2006.870104.

[176] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. Syst., Man Cybern.*, vol. 22, no. 3, pp. 418–435, May/Jun. 1992, 10.1109/21.155943.

[177] L. I. Kuncheva, C. J. Whitaker, and R. P. W. Duin, "Limits on the majority vote accuracy in classifier fusion," *Pattern Anal. Applicat.*, vol. 6, pp. 22–31, 2003.

[178] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[179] D. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, pp. 241–259, 1992.

[180] A. S. Constantinidis, M. C. Fairhurst, and A. F. R. Rahman, "A new multiexpert decision combination algorithm and its application to the detection of circumscribed masses in digital mammograms," *Pattern Recognition*, vol. 34, no. 8, pp. 1527–1537, 2001.

[181] A. S. Constantinidis, M. C. Fairhurst, and A. F. R. Rahman, "Detection of circumscribed masses in digital mammograms using behavior—Knowledge space method," *Electron. Lett.*, vol. 36, no. 4, 2000.

[182] N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," *Med. Phys.*, vol. 23, no. 10, pp. 1685–1696, 1996.

[183] A. Osareh and B. Shadgar, "Machine learning techniques to diagnose breast cancer," in *Proc. 5th Int. Symp. Health Informatics Bioinformatics (HIBIT)*, Apr. 20–22, 2010, pp. 114–120, 10.1109/HIBIT.2010.5478895.

Karthikeyan Ganesan received the B.E. and M.Sc. degrees.

He is a Research Engineer with the Electrical and Computer Engineering Department, Ngee Ann Polytechnic, Singapore. He specializes in applied physics, computational biology and biomedical engineering. His expertise includes medical imaging and physics, image reconstruction, statistical analysis, pattern recognition, machine learning and data visualization.

U. Rajendra Acharya received the Ph.D. degree from National Institute of Technology Karnataka, Surathkal, India, and the D.Engg. degree from Chiba University, Japan.

He is a Visiting Faculty Member at Ngee Ann Polytechnic, Singapore. He is also: Adjunct Professor at the University of Malaya, Malaysia; adjunct faculty at Singapore Institute of Technology–University of Glasgow, Singapore; Associate faculty member at SIM University, Singapore; and Adjunct faculty at Manipal Institute of Technology, Manipal, India. He has published more than 275 papers, in refereed international SCI-IF journals (176), international conference proceedings (48), textbook chapters (62), books (13 including in press) with h-index of 22 (h index = 17 without self-citations) and 26 (Google Scholar, more than 2400 citations). His major research interests include biomedical signal processing, bio-imaging, data mining, visualization and biophysics for better healthcare design, delivery and therapy.

Dr. Acharya is on the editorial board of many journals and served as Guest Editor for many journals.

Chua Kuang Chua received the B.Eng. degree from the National University of Singapore, in 1989, M.Sc. degree from the University of Essex, in 1996, and Ph.D. degree from Queensland University of Technology, in 2010. His Ph.D. topic was analysis of cardiac and epileptic signals using higher order spectra.

He is a Senior Lecturer at Ngee Ann Polytechnic, Singapore. He has published more than 20 papers in international journals. His research interests include digital signal processing, nonlinear biosignal processing and artificial intelligence. Currently, he is involved in projects such as early stage of breast cancer diagnosis using mammograms, patient-sensing automated systems to speed up lower-limb rehabilitation, and a noninvasive tool to evaluate tear evaporation rate for dry eye application.

Lim Choo Min received the B.Eng., M. Eng., and Ph.D. degrees from the National University of Singapore in 1976, 1981, and 1986, respectively.

He is a Senior Director in the school of Engineering, Ngee Ann Polytechnic, Singapore. He has published more than 80 international journal papers and edited a book. His research interests include neural network-based control, fuzzy logic control, bioinformatics, optimization and the development of hardware (training kit sets) and software for supporting mobile computing. Currently, he is working on two external funded projects: exoskeleton to rehabilitate paralyzed arm based on patient healthy arm guidance and early stage of breast cancer diagnosis using mammograms.

K. Thomas Abraham received the B.Sc. Pharm. degree from the National University of Singapore, in 1981, and the MBA degree in health services management from the University of Dallas, TX, USA, in 1998, and the Ph.D. degree in business and management from the University of South Australia, in 2007.

He practised pharmacy until 1993 before embarking on a hospital management career at East Shore Hospital (now known as Parkway East Hospital) and later at Mount Elizabeth Hospital in Singapore. In 2008, he joined the Parkway College as Director of Allied Health to develop allied health programs. Since 2009, he has been the CEO of SATA CommHealth, a not-for-profit organization focusing on community health. In his role as CEO, he is responsible for the management and development of SATA CommHealth's business and community services. He has also delivered many talks and training programs both locally and overseas. He is currently a board member of the International Union Against Tuberculosis and Lung Diseases Asia Pacific. His research interests include branding, corporate reputation, and community health services.

Kwan-Hoong Ng received the M.Sc. degree in medical physics from University of Aberdeen and Ph.D. degree in medical physics from the University of Malaya, Malaysia. He is also a certified by the American Board of Medical Physicists.

He is a Senior Professor at the Department of Biomedical Imaging and a Senior Consultant of the University of Malaya Medical Center, Kuala Lumpur, Malaysia. He has authored/coauthored more than 200 papers in peer-reviewed journals and 15 book chapters. He has presented over 450 scientific papers, more than 200 of which were invited lectures. He has directed several workshops on radiology quality assurance, digital imaging, and scientific writing. His main research contribution has been in the biophysical characterization of breast diseases and developing computer methods as tools to improve the diagnostic capability of mammography. He has also been directing research initiatives in digital imaging, radiological protection, and radiation dosimetry.

Dr. Ng is the Cofounder and Coeditor-in-Chief of the *Biomedical Imaging and Intervention Journal*. He has been serving as an IAEA expert, a member of International Advisory Committee of the World Health Organization and a consulting expert for the International Commission on Non-Ionizing Radiation Protection (ICNIRP).