NORMAL MAMMOGRAM ANALYSIS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Yajie Sun

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2004

This dissertation is dedicated to my grandmother

ACKNOWLEDGMENTS

I wish to express my sincere thanks to my Major Professor Edward J. Delp for his guidance and supervision. This research would not have been completed without his care and dedication in constructively criticizing my work. I also want to express my sincere thanks to Professor Charles F. Babbs for his constructive ideas and suggestions which manifests itself throughout my work. I have truly enjoyed and benefited from working with them.

I also wish to express my sincere thanks to the other members of my committee, Professor Charles A. Bouman and Professor Zygmunt Pizlo, for their support.

I also want to thank CADx Systems for supporting this research.

Finally, I would like to thank my parents for their unconditional support and love. They inspired me and gave me the strength to finish this work.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

SYMBOLS

A_z	The area under the ROC curve
μCa	Microcalcification

ABBREVIATIONS

CAD	computer-aided diagnosis
TDLU	terminal ductal lobular unit
ROI	Region of Interest
ROC	Receiver Operating Characteristics
FROC	free-response ROC
FP	False Positive
TP	True Positive
FPF	False Positive Fraction
TPF	True Positive Fraction
TNF	True Negative Fraction
FNF	False Negative Fraction
DDSM	Digital Database for Screening Mammography
LBP	Local Binary Pattern
GLCM	Gray Level Co-occurrence Matrix

ABSTRACT

Sun, Yajie. Ph.D., Purdue University, August, 2004. Normal Mammogram Analysis. Major Professor: Edward J. Delp.

Breast cancer is the leading cause of cancer death among women. Screening mammography is the only method currently available for the reliable detection of early and potentially curable breast cancer. Research indicates that the mortality rate could decrease by 30% if women age 50 and older have regular mammograms.

The detection rate can be increased 5-15% by providing the radiologist with results from a computer-aided diagnosis (CAD) system acting as a "second opinion." However, among screening mammograms routinely interpreted by radiologists, very few (approximately 0.5%) cases actually have breast cancer. It would be beneficial if an accurate CAD system existed to identify normal mammograms and thus allowing the radiologist to focus on suspicious cases. This strategy could reduce the radiologist's workload and improve screening performance.

In this dissertation, we propose a new full-field mammogram analysis method focusing on characterizing and identifying normal mammograms. A mammogram is analyzed region by region and is classified as normal or abnormal. We present methods for extracting features that can be used to distinguish normal and abnormal regions of a mammogram. A set of 86 features from four different types of characterization is extracted from each region. We implement a method to select a nearly optimal feature subset for the classification.

We propose a unique multi-stage cascading classification method to boost the classification performance. The classifier performs better than a single classifier in that it significantly reduces the false positive rate, the misclassification rate of normal

mammograms as abnormal. We have tested this technique on a set of ground-truth full-field mammograms. The results are comparable to human readers.

This approach is independent of the type of abnormalities and may complement computer-aided detection based on the recognition of specific types of abnormal structures.

1. INTRODUCTION

Breast cancer is the most common cancer of American women and is the leading cause of cancer-related death among women aged 15-54 [2]. In 1996, the American Cancer Society estimated that 184,300 women will be diagnosed with breast cancer and that 44,300 women will die from it [3]. Another study showed approximately 720,000 new cases will be diagnosed world-wide per year, this accounts for about 20% of all malignant tumor cases [4]. The World Health Organization's International Agency for Research on Cancer estimates that more than 150,000 women worldwide die of breast cancer each year [5]. Since breast cancer is a progressive disease, evolving through stages of cellular dedifferentiation and growth, the time at which breast cancer is detected is crucial. The earlier breast cancer is detected, the higher is the chance of survival [6–8]. Screening Mammography is the only method currently available for the reliable detection of early and potentially curable breast cancer.

Mammography is high-resolution x-ray imaging of the compressed breast. This involves radiation transmission through the tissue and the projection of anatomical structures on a film screen or image sensor. Associated with the x-ray imaging projection is a reduction in anatomical information from a 3D organ to a 2D film/image. Two imaging projections of each breast, craniocaudal (CC) and mediolateral oblique (MLO) views, as shown in Figure 1.1, are routinely obtained. This permits some indication of three dimensions and an understanding of overlapping structures. High-quality mammogram with high spatial resolution and adequate contrast separation allows radiologists to observe fine structures. Studies have shown that the mortality rate could decrease by 30% if all women age 50 and older have regular mammograms [9].

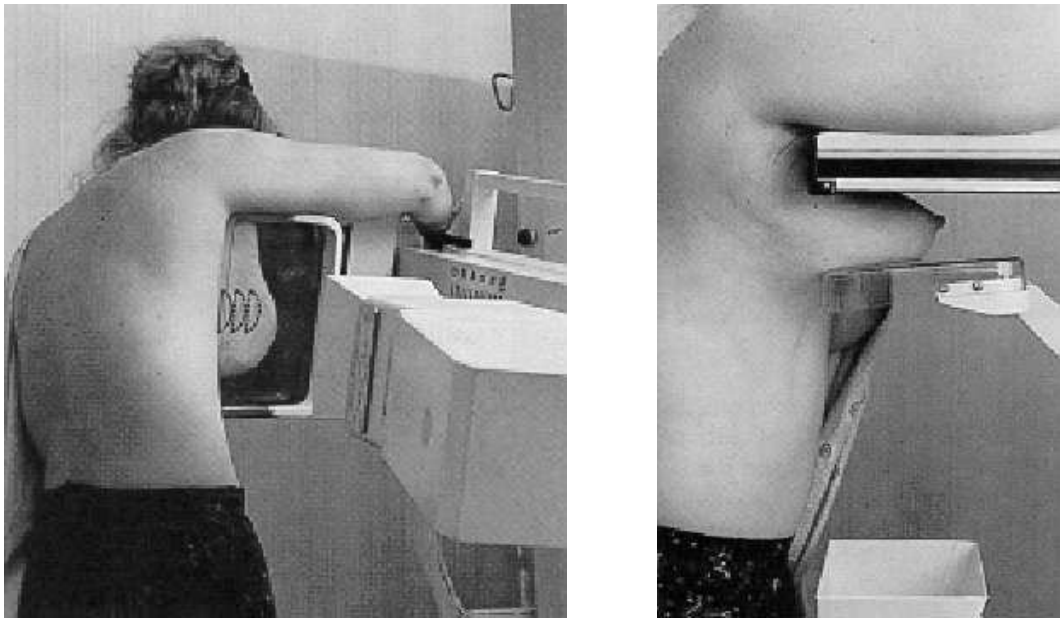


Fig. 1.1. Illustration of two views taken in screening mammography

With the wide spread development of screening programs in the USA, radiologists have had to read a large number of mammograms. Reading mammograms is difficult and requires a great deal of experience. Several studies have shown retrospectively that 20% to 40% of breast cancer fails to be detected at screening [10–12] due to radiologist fatigue, the complex image structure of the breast tissue, and the subtlety of the cancer. Even the most experienced mammographic readers only have a correct detection rate of 85-91% [13–15]. Moreover, a study found that there is about 2.6% to 15.9% false positive reading of negative or benign mammograms by radiologists [16]. Several studies showed that double reading by two radiologists can improve detection sensitivity up to 15% [17]. However, implementing double reading can be very costly, time consuming and logistically difficult. It has been proposed that a computer-aided diagnostic (CAD) system be used as a second reader to assist the radiologist, leaving the final decision to the human [18–20]. CAD can increase the diagnostic accuracy and efficiency with high reproducibility. It has shown that the performance of a radiologist can be increased 5-15% by providing the radiologist with results from a CAD system as a "second opinion" [17, 21, 22]. It has also been shown that a CAD system can detect approximately 50% of the lesions which are missed at screening [23, 24]. The first CAD system for screening mammography approved by USA Food and Drug Administration in 1998 was the **Image Checker** system from R2 Technology [25]. The **Second Look** system of CADx Medical Systems is a competitor of R2. We expect the increasing use of CAD systems with the wide spread implementation of screening programs worldwide.

1.1 A Brief Introduction to Mammography

The human breast has 15-16 major lobes with each lobe having a major duct opening at the nipple. Originating from the nipple, each major duct extends back into the breast in a branching network of smaller ducts that defines a lobe. A major duct branches until the distal portion ends in the terminal lobules. The final

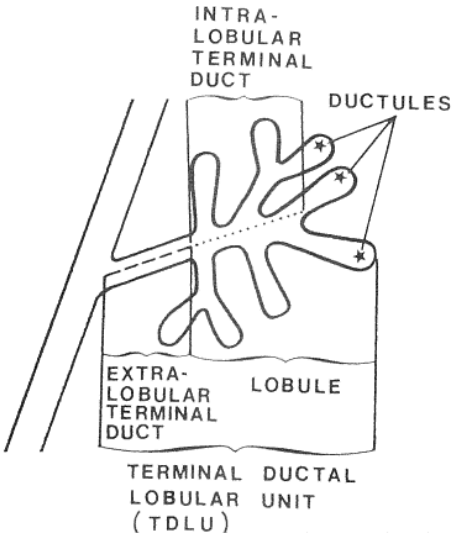


Fig. 1.2. Illustration of a TDLU(from [1])

branch from the branching duct is known as the extralobular terminal duct and is continuous with ductules that extends like fingers in the lobule. The terminal ductal lobular unit (TDLU) is the extralobular terminal duct and the lobule. The TDLU is the place where most cancers arise, such as ductal carcinoma in situ and lobular carcinoma in situ. Cooper's ligaments bridge the superficial and deep pectoral fascia (a sheet of connective tissue) and semi-compartmentalize the lobular structures of the breast, along which the breast tissue is loosely supported. Within a honeycomb-like pattern, fatty and glandular tissues are distributed along the Cooper's ligaments, giving breast tissue a characteristically scalloped appearance in mammograms. The disturbed bulging or convex contours seen on mammograms could be a possible sign of an underlying abnormality [1, 7, 26]. Figure 1.2 illustrates the general structure of a TDLU.

Normal breasts have a wide variation in mammographic appearance. The pattern exhibited by a breast which are predominantly composed of fat can often be called normal if no disturbing pattern is found. Figure 1.3 shows two normal mammograms.

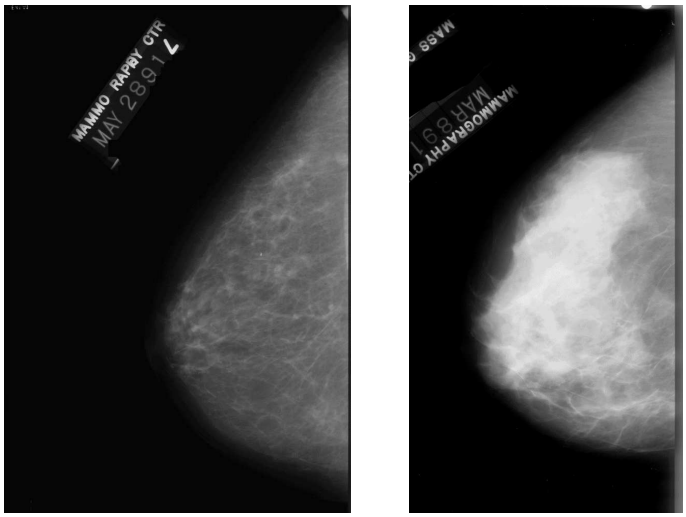


Fig. 1.3. (a) predominantly fatty normal mammogram, (b) dense normal mammogram

There is not a qualitative or quantitative definition of what is normal, though most normal mammograms appear with regular and undisturbed ductal patterns.

Breast cancers usually appear with disturbed ductal structures. There are three major types of breast cancer: circumscribed/oval masses, spiculated lesions and microcalcifications. Malignant lesions generally have a more irregular shape than benign lesions. Circumscribed masses are compact and roughly elliptical. Radiolucent lesions with a halo or encapsule are usually benign. High radiopaque lesion with irregular or ill-defined boundary should be considered with a high degree of suspicion. Figure 1.4 shows a benign and a malignant mammogram. Spiculated lesions have a central tumor mass that is surrounded by a radiating pattern of linear spicules. Most spiculated lesions are malignant. Figure 1.5 shows a mammogram with a spiculated lesion. Microcalcifications appear as bright dot-spots on screening mammograms, usually in the form of clusters. These are calcium deposits from cell secretion and necrotic cellular debris. The shape and distribution of breast calcifications indicate malignancy. Benign microcalcifications are often smooth and sharply outlined and have high uniform density. Malignant microcalcifications usually appear in irregular shape and variably distributed [1, 7, 27]. Figure 1.6 shows a benign and a malignant calcification cluster. There is no deterministic boundary between benign and malignant types.

1.2 Our View of Normal Detection

Since the majority of the screening mammograms are normal, it could help reduce radiologists' workload and improve the efficacy of screening programs if a detection system could readily identify normal mammograms in the clinical environment. Our view of normal detection is as a "first look" reader. Figure 1.7 shows the diagram of our view of normal detection as a "first look" pre-screening system. It is a novel concept since conventional Computer-Aided Diagnosis (CAD) system focuses on detecting individual cancers to assist a radiologist as a "second opinion" reader. We

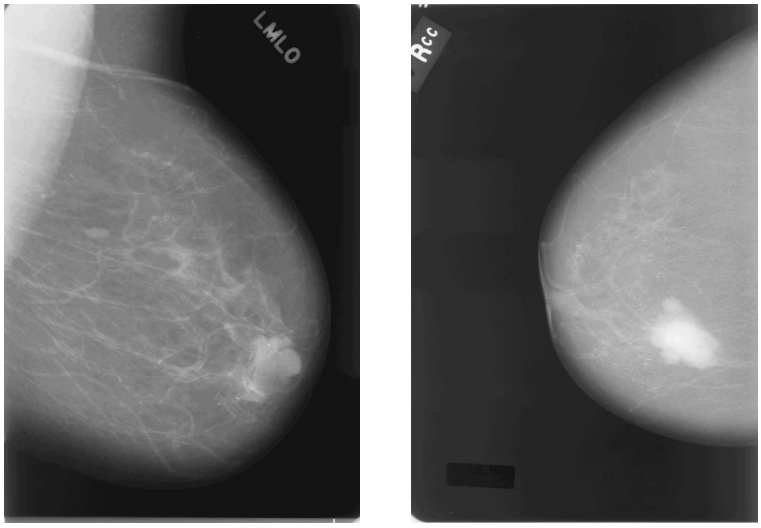


Fig. 1.4. (a) benign mass mammogram, (b) malignant mass mammogram

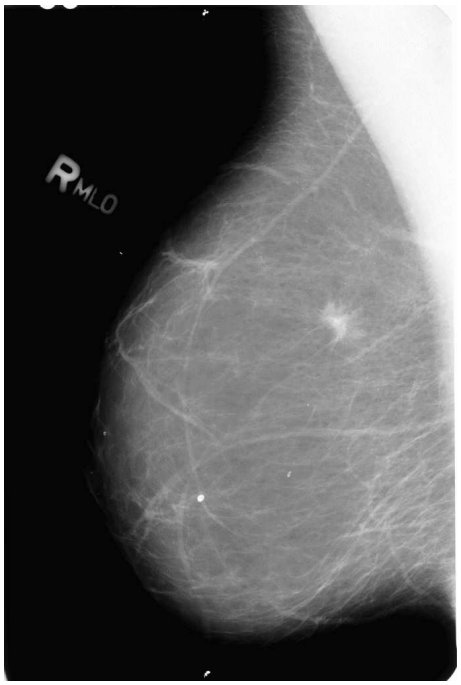
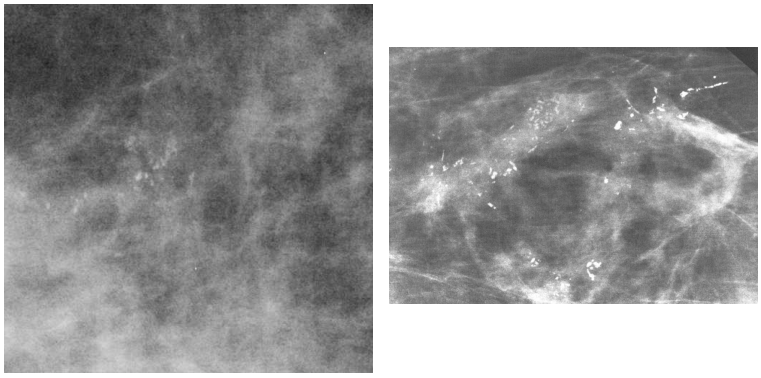


Fig. 1.5. a spiculated lesion mammogram



(a) (b)

Fig. 1.6. (a) a benign calcification, (b) a malignant calcification

propose our normal detection system as a pre-screening method. The patient with unambiguous normal mammogram will be sent home, and the patient with mammogram identified as abnormal during pre-screening normal detection will be sent to a physician for the further processing.

A tumor detection system is characterized by the goal of a high detection rate of cancer mammograms as abnormal. Meanwhile, it seeks to lower false detection rate of normal mammograms as abnormal, though it is not critical in the tumor detection system. A normal detection system is characterized by the goal of a high detection rate of normal mammograms as normal, and a very low rate of abnormal as normal. A very low rate of abnormal as normal is very important in the normal detection since it is critical and risky to classify an abnormal mammogram as normal. Therefore, a normal detection is much more challenging than a tumor detection since normal detection focuses on both improving correct normal detection rate and reducing false detection rate of abnormal as normal. Theoretically, a normal detection system could be used to detection all normal and abnormal mammograms as a super CAD tool.

1.3 Performance Definitions of Computer Aided Detection

The Receiver Operating Characteristic (ROC) is often used to evaluate computer aided detection performance. The Receiver Operating Characteristic (ROC) or free-response ROC (FROC) provides the most comprehensive description of detection or diagnostic accuracy [22, 28–34].

In computer aided detection, there are two classes: one class is cancer or the abnormal class, and the other is the normal class. The following definitions are used to describe correct classification or misclassification for each class:

- True Positive (TP)
  TP = An abnormal classified as abnormal

- True Negative (TN)
  TN = A normal classified as normal

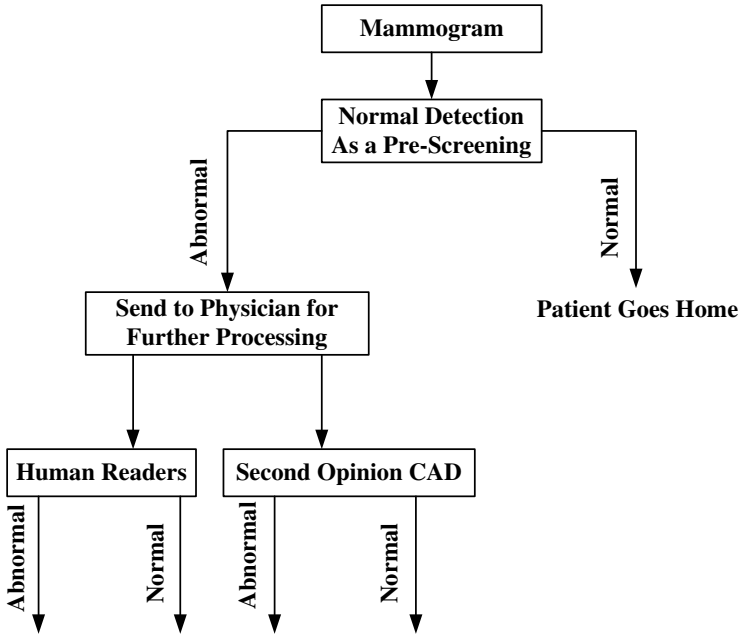


Fig. 1.7. Our view of normal detection as a "first look" reader

- False Positive (FP) or False Alarm

  FP = A normal classified as abnormal

- False Negative (FN)

  FN = An abnormal classified as normal

The classification performance is described using the percentages of correct or incorrect classification of normal or abnormal data:

- True Positive Fraction (TPF): Sensitivity

$$TPF = \frac{\text{Number of abnormal classified as abnormal}}{\text{Total number of abnormal}} \tag{1.1}$$

- True Negative Fraction (TNF): Specificity

$$TNF = \frac{\text{Number of normal classified as normal}}{\text{Total number of normal}} \tag{1.2}$$

- False Positive Fraction (FPF): 1-Specificity

$$FPF = \frac{\text{Number of normal classified as abnormal}}{\text{Total number of normal}} \tag{1.3}$$

- False Negative Fraction (FNF): 1-Sensitivity

$$FNF = \frac{\text{Number of abnormal classified as normal}}{\text{Total number of abnormal}} \tag{1.4}$$

These four fractions are not independent: $TPF + FNF = 1$, and $FPF + TNF = 1$. The goal of normal mammogram identification is to maximize TNF with very low FNF, which is the same as minimizing FPF with very high TPF. The basic argument is that in a clinical situation, women who are classified as normal are sent home whereas women classified as abnormal are sent to a physician to read their mammograms.

Figure 1.8 illustrates how a ROC is generated from the distributions of the normal class and the abnormal class. The classification threshold is changed to obtain the different pairs of TPF and FPF. The ROC is a plot of TPF's versus FPF's. Figure 1.9

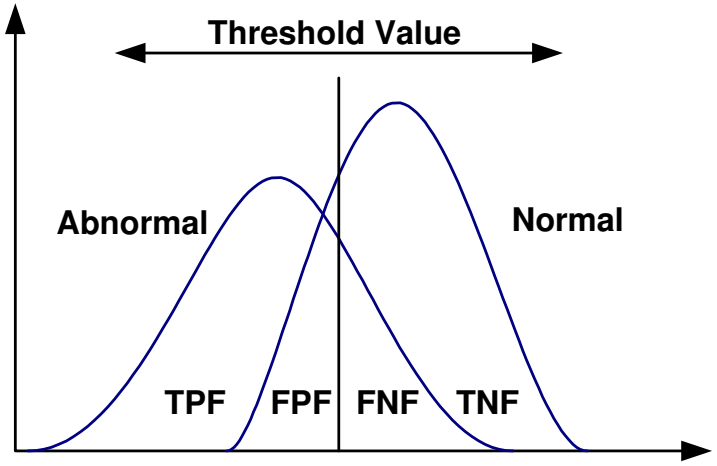


Fig. 1.8. Illustration of how to obtain different pairs of TPF and FPF for a ROC curve

shows a Receiver Operating Characteristic (ROC). The area under a ROC is known as A_z, and is used to evaluate the overall discriminatory power of the classification method. A_z is positive, and is bounded from 0 to 1.0. The higher A_z is, the higher the overall performance of the classification method is. A ROC is often used to select an operating point that provides the trade-off between TPF and FPF.

In this thesis, we will use the above definitions to evaluate our work on normal mammogram analysis.

1.4 Current Status of Computer Aided Cancer Detection

Over the last decade, a great deal of work has been reported on the detection of individual abnormalities, such as microcalcifications, masses and spiculated lesions. We will review the current status of computer-aided diagnosis (CAD) of breast cancer.

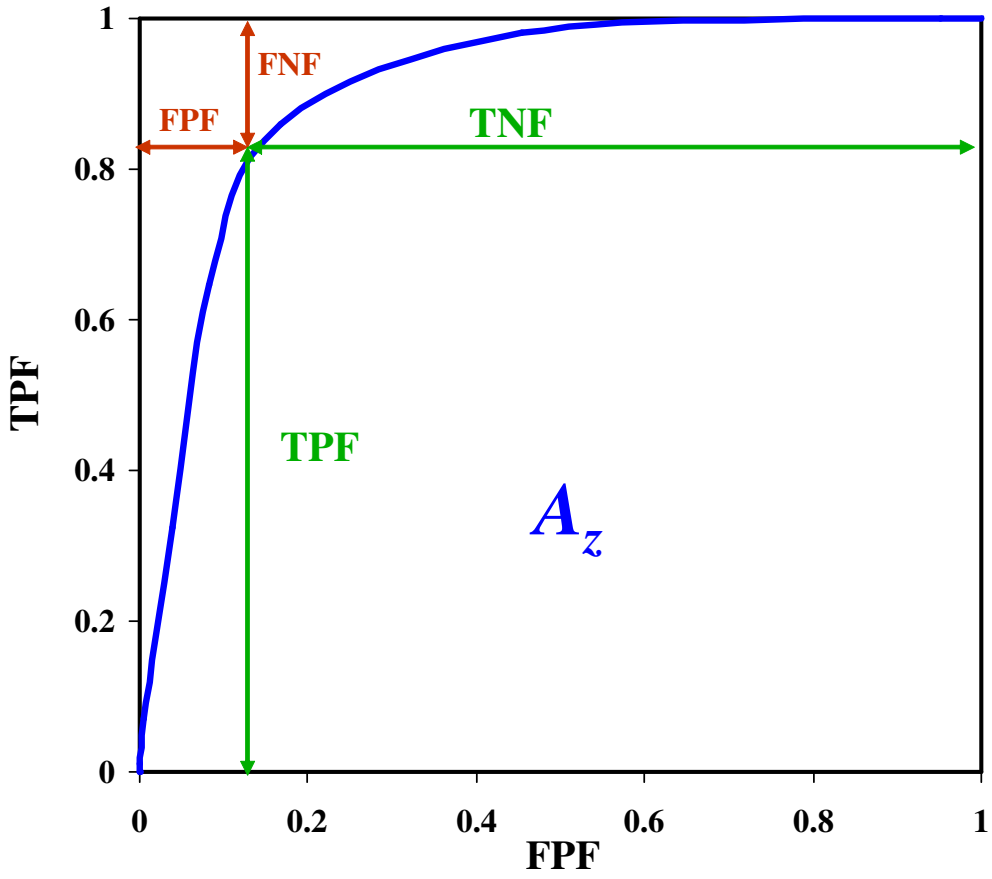


Fig. 1.9. A ROC Curve with TPF, FPF, TNF, and FNF labelled.

1.4.1 Review of Microcalcification Detection

Microcalcification detection is a difficult problem due to the small size of calcifications. This is especially true for the detection of tiny and faint individual calcifications that are not in clusters. In most microcalcification detection, microcalcification clusters can be detected with high sensitivity, but not all of the individual calcifications in the cluster. The research work described in [35] is interesting in that it introduced shape features to differentiate true calcifications and similar normal structures.

Statistical and shape features are generally used to classify mircocalcifications since calcifications are usually brighter than the background and have a dot or small disc shape. In [35,36], a preprocessing step transformed the input mammogram into an image with a homogenous noise level. The detection scheme used three features: a shape parameter using the Hough Transform for detecting lines and the local contrast at two different spatial resolutions. A distinction between faint microcalcifications and connective tissue can only be made using shape. The main contribution of the work is that shape features are used to differentiate true positives from false positives. In [37], a microcalcification detection scheme was developed by using a shape parameter and deterministic relaxing labelling. A shape parameter is necessary to distinguish curvilinear connective tissue from microcalcifications. Results from the segmentation of 15 mammograms showed all true microcalcification were detected correctly. Both Hough and phase congruency shape parameters allow the detection of the linear connective tissue better and produced a low false positive fraction (FPF) when compared to not using the shape parameter. In [38], each mammogram was decomposed into subband images, the statistics skewness and kurtosis were then extracted from each subimage. The test data consisted of 40 mammograms from the Nijmegen database [39]. All of the 105 microcalcification clusters in the test data were detected with an average of 3.3 false alarms per image. Fourth order statistics from an adaptive wavelet decomposition and local maxima were combined to detect

microcalcifications [40]. The adaptive wavelet method performed better than the conventional wavelet transform.

Different approaches have been used to detect microcalcifications, such as neural networks and support vector machines. In [41], a convolution neural network (CNN) selected by simulated annealing improved the performance of microcalcification detection. A training dataset consisted of 472 mammograms (which contained 253 biopsy-proven malignant microcalcification clusters). At a false positive rate of 0.7 per image, the sensitivity of image-based performance was 84.6% with the optimized CNN, in comparison with 77.2% with manually selected CNN. In [42,43], a Support Vector Machine (SVM) was developed to detect microcalcifications. The SVM was trained and tested on more than 76 mammograms with 1120 microcalcifications (38 mammograms for the training and 38 for testing). At 1 false positive cluster per image, the SVM achieved a 0.89 true positive fraction (TPF), compared with 0.76 TPF of the best of four other approaches. In [44], a technique for optimizing the weights of individual scales in the wavelet transform to improve the performance of microcalcification detection was proposed. In the study, a total of 165 microcalcification and a total of 132 normal Region of Interests (ROIs) (size 128×128) were obtained from 39 mammograms. The weights for scales 1 to 5 decomposed with an 8-tap Daubechies' Symmlet were optimized with all 297 ROIs as a training set by a jackknife method repeated 20 times. The performance of microcalcification detection from a reconstructed image of weighted wavelet coefficients was evaluated by A_z. It yielded an average A_z of 0.92, and outperformed the lateral-image-subtraction technique (A_z of 0.86). In [45], a new technique that fused two-view information from the craniocaudal (CC) and mediolateral-oblique (MLO) views was shown to improve microcalcification detection. At a sensitivity of 70%, the false positive fractions were 0.17 and 0.27 with the fusion and single-view detection methods, respectively. This indicates lateral information provides valuable information that can reduce the false positive rate.

Algorithms have also been developed for separating malignant and benign microcalcifications. In [46], microcalcifications were initially detected by using a statistical method. A total of 16 features were then extracted to determine the malignancy of a cluster. The k Nearest-Neighbor (kNN) method was used in a leave-one-out classification. The method's best performance was $A_z = 0.83$, using 9 automatically selected features. The result was also compared with the performance of radiologists. Ten radiologists read and assessed the malignancy for each patient. The result showed that the performance of this method ($A_z = 0.83$) was considerably higher than that of the radiologists ($A_z = 0.63$). In [47], shape and texture features were used to differentiate benign from malignant microcalcifications. The shape-based method consisted of 17 shape features of individual calcifications or clusters, and texture-based method consisted of 44 texture features extracted from the gray level co-occurrence matrix [48]. A genetic algorithm is used for feature selection, and a k Nearest-Neighbor (kNN) classifier was designed for the classification of malignant microcalcifications. The test data contained 74 malignant and 29 benign microcalcification clusters. For shape features, the best performance was $A_z = 0.82$ and for texture features $A_z = 0.72$. Co-occurrence texture features were also used in [49, 50] for classification of malignant and benign microcalcifications. In [51], BI-RADS [52] features from radiologists were used to improve the performance of separating benign and malignant microcalcifications. From 292 test cases, with features from image analysis method, and the use of a linear classifier resulted in $A_z = 0.59$; with the BI-RADS features included, the performance was improved to $A_z = 0.69$. In [53], a classification scheme using the temporal change of mammographic features was developed to improve the differentiation of malignant and benign calcifications. Twenty statistical texture features and 21 morphological features were extracted. The difference features were obtained by subtracting a prior feature from the corresponding current feature. The linear discriminant classifier achieved an average A_z of 0.98 on the training data and A_z of 0.87 on the test data, comparing with the classifier

without temporal features, which yielded an average A_z of 0.88 on the training data and A_z of 0.81 on the test data.

1.4.2 Review of Mass Detection

Circumscribed masses usually have variable sizes with normal dense tissue causing difficulties for mass detection. Many approaches use image segmentation to locate the masses after initial region growing, resulting in poor boundary description. The sensitivity of mass detection is generally lower than that of microcalcification detection because masses are not usually brighter than the background.

In [54], an adaptive region growing technique was used to segment masses from normal background. It achieved a 97% detection sensitivity for a set of 51 mammograms. A fuzzy region growing method for segmenting breast masses was proposed in [55, 56]. In [57, 58], a two-stage adaptive density-weighted contrast enhancement filter was used in combination with a Difference-of-Guassian(DoG) edge detector for the detection of masses. Using a set of morphological features they reported a 96% detection accuracy at 4.5 false positives per image on 25 mammograms. The later study on a larger set of 168 mammograms yielded a 80% detection sensitivity at 2.3 false positives per image using a set of texture features extracted from the gray-level co-occurrence matrices (GLCM) [48]. In [59], breast mass portions were segmented by the intensity links from the central body into the surrounding regions. Features based on the flow orientation in an adaptive ribbon of pixels around the mass boundaries were used to separate mass regions from normal breast regions. A performance of $A_z = 0.87$ was achieved for a dataset of 56 mammograms from the Mammographic Image Analysis Society (MIAS) database [60] consisting of 30 benign, 13 malignant and 13 normal cases. A sensitivity of 81% at 2.2 false positives per image was obtained. The detected mass regions (13 malignant and 19 benign) were further classified as benign and malignant by using five features based on gray-level co-occurrence matrices. The classification of benign vs. malignant yielded a

performance of $A_z = 0.79$. In [61], the initial mass was segmented from a difference of Gaussian (DoG) filtered images through multi-level thresholding. Features including shape, fractal dimension, the output from a Laguerre-Gauss(LG) Channelized Hotelling observer were used to reduce the false positive rate. It achieved a sensitivity of 88%. Using the selected features, the false positives per image were reduced from 20 to 5 with no loss in sensitivity.

Statistical modelling has been used effectively for the detection of mammographic masses. In [62], a statistical segmentation method was developed based on Markov Random Field (MRF) model. The pixel classes were estimated by minimizing the expected value of the number of misclassified pixels. This is known as the "maximizer of the posterior marginals" (MPM) estimate. Unlike other MPM algorithms, this method does not require that the values of all parameters of the marginal conditional probability mass functions of the pixel classes be known *a priori*. It combines the expectation-maximization (EM) algorithm for parameter estimation with the MPM algorithm for segmentation. This EM/MPM algorithm achieved 100% detection of circumscribed mass over testing mammograms. A method based on a multiresolution markov random field was developed in [63]. The segmented regions were classified by a fuzzy binary decision tree as normal tissues or masses. The method achieved 90% sensitivity with two false positive per image. In [64], a wavelet decomposed image of a mammogram was modelled with a Gaussian Markov Random Field (GMRF) and adaptive features based on GMRF were defined for each pixel. Masses were segmented via the fuzzy c-means algorithm using the localized features. The segmentation results were further enhanced by using the expectation-maximization (EM) algorithm. In [65], an approach for knowledge-based mass detection was presented. A suspicious region is determined to be "positive" for a mass, depending on comparing not only positive masses but also negative regions. A classification measure was obtained from the integration of two likelihood measures using two features: circularity and compactness. The performance improved from $A_z = 0.83$ to 0.87 in the training set and from $A_z = 0.80$ to 0.83 in the independent test set

with known negative regions included. Other techniques have been investigated for the detection of masses in screening mammograms, such as neural networks [66,67], genetic algorithm [68,69], support vector machines [70], wavelet packets [71], texture analysis [72], and graph techniques [73].

Shape and texture features from the boundary of the mass have been prominent indicators of malignancy. In [74], an iris filter was used for the detection of rounded convex regions, features from the gray-level co-occurrence matrix of the iris filter was used to isolate malignant tumors from normal tissue. The method achieved a detection rate of 90.4% at 1.3 false positive per image with a dataset of 1214 mammograms containing 208 malignant masses. In [75], a system was designed for clinical use to differentiate malignant and benign masses. The features used in the experiment included shape descriptors and additional collected personal data. A decision tree was implemented for the classification of malignant masses from benign masses. A total of 25 biopsy-proven masses (including 10 malignant and 15 benign cases) were tested with only one case in each group misclassified. In [76], an approach for separation of malignant and benign masses was developed. First, the masses were segmented using pixel aggregation with likelihood analysis. Then a set of features were used with a neural network to separate the malignant and benign masses. An A_z of 0.71 was achieved using 51 mammograms (28 malignant and 23 benign cases). In [77], temporal changes were used to classify masses as malignant or benign. 126 temporal pairs, including 73 malignant and 53 benign, were studied. A linear classifier was used to classify malignant masses from benign using extracted morphological, texture and spiculation features. The average A_z of malignant detection by 5 radiologists was 0.79. With computer-aided results, the performance of the radiologists was improved to $A_z = 0.87$. The improvement was statistically significant. In [78], a mammogram is segmented to obtain initial candidates, shape features were then extracted from the segmented lesions to separate malignant masses from benign. A kNN classifier was used to obtain image-based performance of $A_z = 0.79$, and cased-based performance of $A_z = 0.84$. In [79], ten texture and shape features were used to train a

three-layer back-propagation neural network (BNN) for classification of malignant masses. The average A_z for the radiologists was 0.846, while BNN achieved an A_z of 0.923.

1.4.3 Review of Spiculated Lesion Detection

Among the three major breast cancers, spiculated lesions are the most difficult for computer-aided diagnosis systems to detect. The results from [80, 81] are one of the best detection methods for spiculated lesions.

In [80, 81], a multiresolution representation of a screening mammogram was obtained from a 2D nonseparable wavelet transformation. Four features were extracted from each decomposed image. Three classifier steps were used, with the detection results propagated from the coarsest resolution to the original resolution. Decision trees were used for the classification. A filtering step was also used to obtain a local consensus. This approach was tested on a dataset consisting of 19 spiculated and 19 normal mammograms from the Mammographic Image Analysis Society (MIAS) database. It achieved 84.2% true positive detection rate at less than 1 false positive per image and 100% true positive rate at 2.2 false positive per image. In [82], a spiculated lesion detection based on the skeleton analysis using the iris filter was developed. A modified Hough Transform was used to extract the radiating lines from the center of a mass region to differentiate the stellate lesions from benign masses. It achieved an accuracy of 74% for the detection of stellate lesions from a dataset of 34 mammograms including 14 spiculated cases. In [83], a spiculation detection was based on the statistical analysis of the edge orientation map for detecting spicules. This technique was tested on a total of 50 mammograms (31 normal, 9 malignant spiculated and 10 architectural distortion) from the Mammographic Image Analysis Society (MIAS) database. A 90% true positive fraction was achieved at 1 false positive per image. A statistical representation of line pattern structure was used to detect stellate lesions in [84]. A vector of line-strength, orientation and scale at

each pixel was obtained from a multi-scale line detector. The posterior probability of stellate lesion was estimated from the statistical representation of the line patterns. This technique achieved a sensitivity of 89% at 0.23 false positives per image. A multi-scale directional recursive median filter was used to detect the central masses of stellate lesions, this was combined with the factor analysis of line-strength pattern structures to improve the performance of stellate lesion detection [85,86]. In [87–89], a spiculated lesion detection technique was developed. Features were extracted from the neighborhoods of every pixels of the screening mammogram. The feature vectors were then used to train a binary decision tree; this tree classifier was used to label each pixel with the probability of abnormality. Finally, spatial filtering was used to force a local consensus on the presence or absence of a spiculated lesion. In one experiment with 50 normal and 12 spiculated lesion mammograms selected from Digital Database for Screening Mammography (DDSM) [90], this method achieved either perfect detection with a false positive fraction of 0.27, or zero false positive fraction at a true positive detection of 0.92.

1.4.4 Non X-ray Methods of Breast Cancer Detection

There are other imaging modalities used for breast cancer detection. These are ultrasound [91–93], microwave [94,95], and infra-red [96]. Even though the resolution is much lower than X-ray mammography, one major attraction is that ultrasound, microwave, and infra-red do little harm to the body.

1.5 Normal Mammogram Analysis

Almost all of the previous research in computer-aided diagnosis (CAD) focused on the detection of individual cancer types, such as microcalcification. However, among high volume x-ray screening mammogram programs, only a few (approximately 0.5%) cases actually present with breast cancer [19]. There are 2.6% to 15.9% false positive readings of negative or benign mammograms by radiologists [16]. These

women will be unnecessarily called back and referred for further biopsies. The study in [97] has shown that these costly and invasive procedures are only 15%-34% likely to actually show malignancy at histological examination. We feel that a system for the detection of normal mammograms with very high specificity could allow a radiologist to focus more on suspicious cases. This strategy could reduce the radiologist's workload, and improve screening efficiency and the true positive rate. A possible further scenario would be one where the normal detection system is used for pre-screening normal mammograms. This system would detect unequivocally normal mammograms and label the rest of mammograms as suspicious. These "suspicious" mammograms would then be read by a radiologist. The pre-screening system would have to have a large true negative fraction and a small false negative fraction.

In [98–100], a statistical method was developed for identifying normal areas from abnormal regions. Each mammogram was decomposed into a sum of independent difference images with a wavelet expansion. Based on the histograms of the mammogram regions, each difference image could be modelled as a Laplace distribution. The normal region could be identified from microcalcification regions using a Neyman-Pearson test. However, the emphasis of this modelling is with the separation of calcifications and normal regions. It is a microcalcification detector in reality. In [101–103], normal ductal structure and texture background separation was studied. In [81, 104], a line model was developed to extract and analyze the curvilinear structure of normal tissues. After the linear structure removal, normal mammograms appear "featureless" and relatively uniform, while abnormal mammograms have their abnormalities highlighted in a "featureless" background. The work presented in this thesis is initially based on these concepts [81, 104].

Similarly, curvilinear structures were detected using statistical modelling and linear classification based on scale-orientation signatures [105–109]. In [26, 110–112], methods for extracting curvilinear structures of normal mammograms were developed.

1.6 Overview of the Dissertation

In this dissertation, we will propose a normal mammogram analysis system focusing on achieving a high detection rate for normal mammograms, while keep the misclassification rate of labelling abnormal breast regions as normal very small. This method is fundamentally different from other approaches, which identify normal mammograms by detecting cancers. Our approach is independent of the type of abnormalities, and may complement computer-aided detection based on the recognition of specific types of abnormal structures.

The thesis is organized as follows. This chapter introduced computer-aided detection and described why normal detection could be beneficial. Chapter 2 presents the formulation of the normal mammogram detection problem and key procedures. Chapter 3 presents our methods of image preprocessing and feature extraction. We will present four types of features for characterizing normal mammogram regions. In Chapter 4, we propose a unique multi-stage cascading classification system for improving classification performance using complex data sets. In Chapter 5, the results of our normal analysis of full-field mammograms are presented. Finally, Chapter 6 is the conclusion of this work and a discussion of future work.

2. MODELLING NORMAL DETECTION

In this chapter, we will model the normal mammogram detection problem as a two-class pattern classification task. A theoretical scheme will be derived and formulated as a foundation for the implementation of our normal mammogram analysis.

2.1 The Normal Detection Concept

Normal breast tissue consists of fatty, glandular and fibrous tissues. Normal areas are usually characterized by curvilinear structures that are slightly brighter than their surroundings. These curvilinear structures are the milk ducts or ductules. The ducts spread out radially from the nipple in the form of a tree-like pattern of ductal structures, ending in the milk production lobules. Abnormal (suspicious in a weaker sense) breast tissue appears with disturbed curvilinear structures.

It should be emphasized that normal detection is not equal to labelling breast tissue as normal after removing a combination of individual abnormalities. There are two very important aspects in our normal detection philosophy. First, normal breast tissue is identified according to its characteristics. Second, we require the very important constraint of a low misclassification rate of labelling abnormal breast tissue as normal. Without this constraint, normal detection will fail in mammographic screening. It would be risky and dangerous if a high percentage of abnormal breasts fails to be uncovered in clinical screening situations.

There are two levels of normal detection. One is region based detection, i.e. labeling a region as either normal or abnormal. The other is global detection, i.e. classify a full-field mammogram as normal if there is no region in the mammogram classified as abnormal. Our goal is to develop a full-field normal mammogram identification system based on regional analysis.

2.2 Modelling Normal Detection

The goal of our research is to design a detection system that associates a label with a mammogram. For a mammogram, the system will label it either as normal or abnormal. In the following, normal detection will be formulated as a two-class pattern classification problem.

Generally, a digitized mammogram is approximately 5000×3000 pixels with a pixel resolution of about 50 microns. Each pixel is represented by 12 bits. Hence, a digitized mammogram is nearly 30MB. The first step is to map the image space into the characteristic feature space. The feature space (or domain) is partitioned into two disjunctive classes Ω_{normal} and $\Omega_{abnormal}$ (they overlap in reality due to non-perfect feature representation). After a mammogram or a region is represented as a feature vector, the second step is to design a classifier based on a training set of normal and abnormal cases. A good classification system means a high normal detection rate with a very low misclassification rate.

Figure 2.1 shows the two steps of normal mammogram identification.

2.2.1 Feature Extraction

The first step of normal detection is to map a mammogram into a set of features. Features can be a set of real numbers which characterize a normal mammogram. The question is how to find the characteristic features which can be used to discriminate the class of normal mammograms from the class of abnormal mammograms. Some simple statistical features could be the mean and standard deviation of pixel intensities of a region. However, the process of feature extraction should be problem-dependent. Non-mammogram-specific statistical features can not completely separate normal and abnormal breast tissue due to the complex structures of normal breast tissue and the subtleties of breast cancers. Mammogram-specified features must be explored, including curvilinear and texture features of the normal breast tissue and shape features of abnormal regions [35, 37, 47, 113, 114]. To ex-

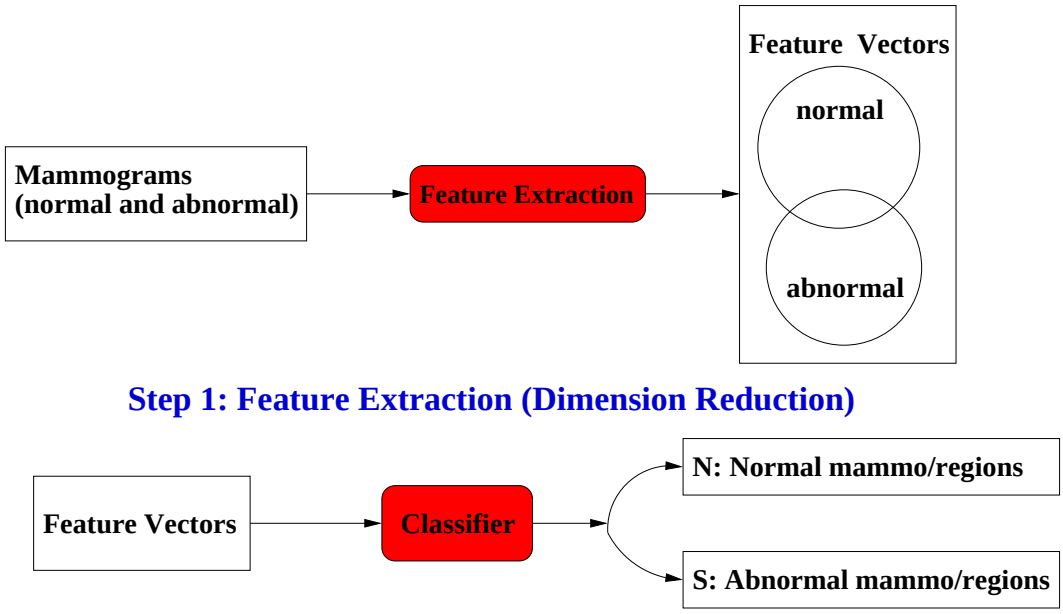


Fig. 2.1. Two steps in identifying normal mammograms

tract "good" features is not only an engineering problem but also an art. One needs to understand the breast anatomy and x-ray mammography to be able to extract characteristic features for separating normal mammograms from abnormal.

It is intuitive to believe that an increase in the number of features can achieve more accurate classification results. Hence, the more features, the better. However, this is incorrect due to the "curse-of-dimensionality" [115, 116]. In high dimensional space it is impossible to obtain large training sets equally spaced, i.e. with increasing dimension an exponentially increasing number of training data is required to guarantee a densely sampled Euclidean space. It can be shown that with increasing dimensionality of the feature vector the discriminating power decreases [117].

We will look at feature extraction and selection in more details in the later chapters.

2.2.2 Two-Class Classification

Assume each (normal or abnormal) mammogram can be represented as a k dimension feature vector after feature extraction. Denote the feature vector as $X, X \in \Re^k$ (a k-dimension random vector in the feature space). X has an associated ground-truth label $Y, Y \in \{N, S\}$, where N represents the class of normal mammograms and S is the class of abnormal mammograms. $\Re^k$ can be partitioned into two disjunctive sub-domain Ω_N and Ω_S with $\Re^k = \Omega_N \cup \Omega_S$, where Ω_N is the domain of normal feature vectors and Ω_S is the domain of abnormal feature vectors. The random vector (X, Y) has the joint probability density function $p(X, Y)$ characterizing the distribution of normal and abnormal mammograms. Hence, we can formulate our normal detection problem as a two-class classification problem.

$$\text{Given } X, Y,\ X \in \Re^k, \text{ and } Y \in \{N, S\}, (X, Y) \sim p(X, Y) \tag{2.1}$$

Find a mapping/classifier g such that

$$g : \Re^k \rightarrow \{N, S\}, \min_g \varepsilon(g) = C_{S,N} P\{g(X) \neq Y | Y = N\} + C_{N,S} P\{g(X) \neq Y | Y = S\}, \tag{2.2}$$

where $\varepsilon(g)$ is the total misclassification error, $C_{S,N}$ is the cost of deciding $X \in \Omega_S$ when $X \in \Omega_N$ and $C_{N,S}$ is the cost of deciding $X \in \Omega_N$ when $X \in \Omega_S$ with $C_{N,S} \gg C_{S,N}$.

Since more than 95% mammograms are normal, a 95% detection rate can be achieved by classifying all mammograms as normal. This is why $\min_g \varepsilon(g) = P\{g(X) \neq Y\}$ will not work for the normal detection problem. The high cost $C_{N,S}$ has to be used to penalize the misclassification rate of abnormal breast tissue as normal tissue.

The best mapping is the Bayes classifier:

$$g_{Bayes} = \{g : \inf_{g:\Re^k \rightarrow \{N,S\}} C_{S,N}\, P\{g(X) \neq Y | Y = N\} + C_{N,S}\, P\{g(X) \neq Y | Y = S\}\} \tag{2.3}$$

Below we derive the theoretical Bayes classifier based on minimizing the minimum cost [115].

Assume we have *a priori* probability $P(Y = N)$ and $P(Y = S)$, and prior density function $p(X|Y = N)$ and $p(X|Y = S)$. To simplify the notation, let

$$P(Y = N) \equiv P_N, \text{ and } P(Y = S) \equiv P_S \tag{2.4}$$

$$p(X|Y = N) \equiv p_N(X), \text{ and } p(X|Y = S) \equiv p_S(X) \tag{2.5}$$

Then we have

$$\begin{aligned} p(X) &= \sum_Y p(X, Y) \\ &= P(Y = N)p(X|Y = N) + P(Y = S)p(X|Y = S) \\ &= P_N\, p_N(X) + P_S\, p_S(X) \end{aligned} \tag{2.6}$$

Then the *a posteriori* probability can be obtained according to the *Bayes Theorem* as

$$P(Y = N|X) = \frac{P_N\, p_N(X)}{p(X)} \tag{2.7}$$

$$P(Y = S|X) = \frac{P_S\, p_S(X)}{p(X)} \tag{2.8}$$

Generally, $P(Y = N|X)$ and $P(Y = S|X)$ are functions of X, i.e. they are also random vectors. Hence, denote

$$P(Y = N|X) \equiv q_N(X), \text{and } P(Y = S|X) \equiv q_S(X) \tag{2.9}$$

Re-assign the cost to the decision given X, $C_{i,j}$ = cost of deciding $X \in \Omega_i$ when $X \in \Omega_j$, where $i, j \in \{N, S\}$. The conditional cost of classifying $X \in \Omega_i$ given $X, r_i(X)$, is

$$r_i(X) = C_{i,N}\, q_N(X) + C_{i,S}\, q_S(X), i = N \text{ or } S \tag{2.10}$$

The Bayes Classifier is

$$r_N(X) \underset{\Omega_S}{\overset{\Omega_N}{\lessgtr}} r_S(X) \tag{2.11}$$

Let

$$r(X) = \min[r_N(X), r_S(X)] \tag{2.12}$$

The total cost is

$$\begin{aligned} r &= E\{r(X)\} = \int \min[r_N(X), r_S(X)]p(X)\,dX \\ &= \int \min[C_{N,N}P_N\,p_N(X) + C_{N,S}P_S\,p_S(X), C_{S,N}P_N\,p_N(X) + C_{S,S}P_S\,p_S(X)]\,dX \\ &= \int_{L_N} [C_{N,N}P_N\,p_N(X) + C_{N,S}P_S\,p_S(X)]\,dX \\ &\quad + \int_{L_S} [C_{S,N}P_N\,p_N(X) + C_{S,S}P_S\,p_S(X)]\,dX, \end{aligned} \tag{2.13}$$

where L_N and L_S are the classification domains of normal feature vectors and abnormal feature vectors determined by equation 2.11.

Replacing $\int_{L_S} p_S(X)dX$ with $1 - \int_{L_N} p_N(X)dX$, we have

$$r = (C_{S,N}P_N + C_{S,S}P_S) + \int_{L_N} [(C_{N,N} - C_{S,N})P_N\,p_N(X) + (C_{N,S} - C_{S,S})P_S\,p_S(X)]dX \tag{2.14}$$

L_N is chosen to minimize the total cost. Thus the minimum cost classifier is to categorize all and only the X's to L_N, which make the integrand of 2.14. Hence,

$$(C_{N,N} - C_{S,N})P_N\,p_N(X) + (C_{N,S} - C_{S,S})P_S\,p_S(X) \underset{\Omega_S}{\overset{\Omega_N}{\lessgtr}} 0 \tag{2.15}$$

Therefore, we obtain

$$\frac{p_N(X)}{p_S(X)} \underset{\Omega_S}{\overset{\Omega_N}{\gtrless}} \frac{(C_{N,S} - C_{S,S})P_S}{(C_{S,N} - C_{N,N})P_N} \tag{2.16}$$

or

$$-\ln \frac{p_N(X)}{p_S(X)} \underset{\Omega_S}{\overset{\Omega_N}{\lessgtr}} \ln \frac{(C_{S,N} - C_{N,N})P_N}{(C_{N,S} - C_{S,S})P_S} \tag{2.17}$$

Unfortunately, the underlying probability density functions $p_N(X)$ and $p_S(X)$ can not be easily determined. Generally, only a finite set of sample data is available. These are often known as training samples. Here we assume there are M feature vectors X_i with the corresponding ground truth class label Y_i: M_1 from Ω_N and M_2 from Ω_S with $M_1 + M_2 = M$.

$$\Gamma_M = \{(X_1, Y_1), \cdots, (X_M, Y_M)\} \tag{2.18}$$

Now, the classifier or mapping g_M is

$$g_M = \min_{g'} [C_{S,N}P\{g'(X_1^N, \cdots, X_{M_1}^N) \neq (Y_1^N, \cdots, Y_{M_1}^N)\}$$

$$
\begin{aligned}
& + C_{N,S} P\{g'(X_1^S, \cdots, X_{M_2}^S) \neq (Y_1^S, \cdots, Y_{M_2}^S)\}] \\
= & \min_{g'} [C_{S,N} \sum_{i=1}^{M_1} U(g'(X_i^N), Y_i^N) + C_{N,S} \sum_{i=1}^{M_2} U(g'(X_i^S), Y_i^S)], \qquad (2.19)
\end{aligned}
$$

where $(X_i^N, Y_i^N) \in \Gamma_M$ and $(X_i^S, Y_i^S) \in \Gamma_M$, cost $C_{N,S} \gg C_{S,N}$, and function $U(\cdot)$ is as follows:

$$
U(g(X), Y) = \begin{cases} 1 & g(X) \neq Y \\ 0 & g(X) = Y \end{cases} \qquad (2.20)
$$

- Parametric Classification

  The mean vector and Covariance matrix are important parameters used to characterize the distribution of each class. They can be estimated from the samples. The sample mean and covariance matrix are defined by

  $$
  \hat{M} = \frac{1}{N} \sum_{i=1}^{N} X_i \qquad (2.21)
  $$

  and

  $$
  \hat{\Sigma} = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \hat{M})(X_i - \hat{M})^T \qquad (2.22)
  $$

  The mean vector and covariance matrix can be estimated from the samples for the feature vectors of normal breast tissue, $\hat{M}_N$ and $\hat{\Sigma}_N$; $\hat{M}_S$ and $\hat{\Sigma}_S$ for the feature vectors of abnormal breast tissue. A model of the classifier type is then assumed. For example, a generic quadratic classifier can be expressed as

  $$
  h(X) = X^T Q X + V^T X + v_0 \underset{\Omega_S}{\overset{\Omega_N}{\lessgtr}} 0, \qquad (2.23)
  $$

  where Q, V, and v_0 are a matrix, vector and scalar, respectively. We can find Q, V, and v_0 by optimizing a metric function $f(\hat{M}_N, \hat{\Sigma}_N, \hat{M}_S, \hat{\Sigma}_S)$. A linear classifier can be considered as a special case of a generic quadratic classifier with Q as a zero matrix.

- Nonparametric Classification

  A nonparametric classifier does not assume the structures of the underlying probability density functions. This is one of the attractions of nonparametric classification. The nonparametric classifier converges to the *Bayes classifier*

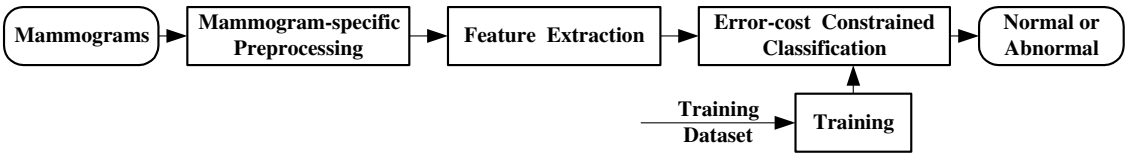


Fig. 2.2. Normal mammogram classification

when an infinite number of samples are used. For the general classification problem with the distribution unknown, nonparametric classification methods achieves comparable or better performance. Neural networks, support vector machines, genetic algorithms, and decision trees are widely used supervised nonparametric methods. We will discuss the binary decision tree classifier in the later chapters and its combination with other classifiers to achieve enhanced classification performance.

2.2.3 Full-Field Analysis

Figure 2.2 shows the diagram of normal detection. The pre-processing step is used to "normalize" the mammogram/region for more consistent feature extraction, which will be described in detail in Chapter 3.

This classification scheme will be used to analyze full-field mammograms and classify each mammogram as either normal or abnormal.

2.3 Challenges

Normal breast tissue have many variations from person to person. It is a challenging problem to characterize normal mammograms. There are two key steps in this problem:

- How to extract a characteristic feature vector X?

  It is obvious that this is the crucial step in solving the normal detection prob-

lem. Mammogram-specific features have to be explored in order to address this problem successfully. One type of feature will be curvilinear patterns of normal breast tissue. Since the complex structures of breast tissue and the different sizes of breast cancers, multi-scale features are intuitively interesting.

- How to design the classifier g_M?

  A good classification system should have a high detection rate of normal breast tissue, but a very low misclassification rate of abnormal as normal. The later is critical in clinical use. The risk is too high if a breast cancer fails to be detected in a clinical screening situation. We will present a unique multi-stage classification system to improve the overall performance of normal detection.

3. PREPROCESSING AND FEATURE EXTRACTION

In this chapter, we will present mammogram-specific preprocessing to "normalize" a mammogram. Due to scattered and extra-focal radiation of x-ray imaging, a preprocessing and enhancement step is necessary to extract consistent and comparable features across different mammograms. A power transform is then used to "normalize" the features.

Figure 3.1 shows the block diagram of our overall system, including preprocessing, feature extraction, and classification. The details of the classification will be given in Chapter 4.

3.1 Mammogram Database

All of the mammograms used in our work were obtained from the Digital Database for Screening Mammography (DDSM) distributed by University of South Florida [90]. The DDSM is a database of digitized mammograms with associated ground truth and other information. The purpose of this database is to provide a large set of mammograms that are free and can be used by researchers to evaluate and compare the performance of computer-aided detection (CAD) algorithms. The database contains 2620 cases available in 43 volumes with each case having four views (mediolatral oblique and cranio caudal views of left and right breasts). Cancer cases are provided with ground-truth information concerning each cancer, such as its location and boundary, along with BI-RADS [52] values provided by the radiologist.

It is necessary to normalize each mammogram to the optical density according to the calibration characteristics of the digitizer used, which removes the gray level intensity variation due to the digitization processing. There is a calibration equation for each digitizer available in the DDSM [90] that describes how to convert gray levels

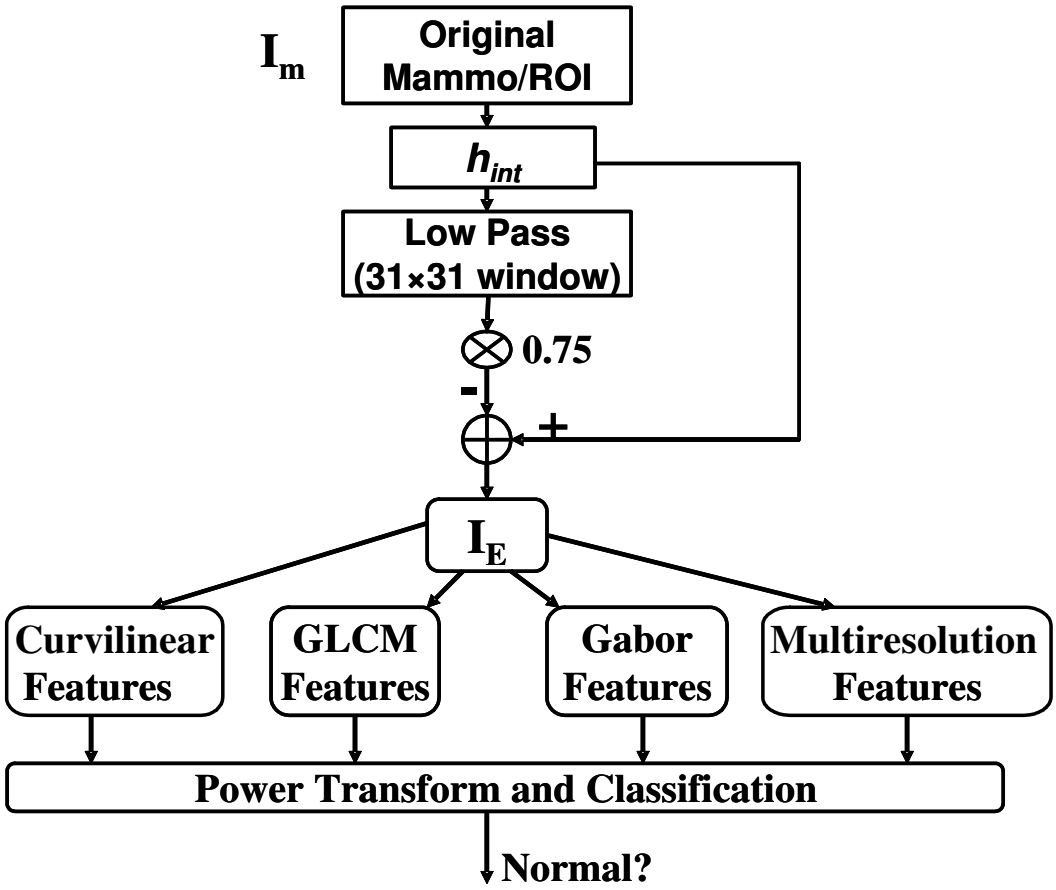


Fig. 3.1. System for identifying normal breast regions

to optical density. For example, for a mammogram scanned from a Howtek digitizer, the following equation converts a gray level (GL) to an optical density (OD): OD = 3.789 - 0.00094568 * GL. The normalized mammogram are then linearly mapped to 8 bits of gray level.

All of our training regions and independent testing full-field mammograms are randomly selected from the DDSM. Region analysis is a very natural processing approach. First, many mammography-specific features are characterized by the surrounding structures, such as the halo or encapsule around benign tumors (e.g. cysts or fibroadenoma), or the extensive ductal appearance of normal tissue. Second, it is computationally much more efficient than pixel-level processing since a typical

digitized mammogram is approximately 5000 by 3000 pixels. For our work, each training region is 512×512 pixels, which is an approximate size that can contain most breast tumors. This is also very desirable for multi-resolution analysis. The training regions are manually extracted from mammograms and are different from the full-field mammograms used for classification. All normal regions were extracted from normal mammograms. According to the "ground truth" available from the DDSM, cancer regions were extracted from cancer cases with the known cancer at the center of the region. In this work, we used 296 normal and 164 abnormal regions for training the classifier.

3.2 Enhancement Based on h_{int} Representation

A standardized mammogram representation can be obtained based on modelling the X-ray physics of the image formation process. We used the techniques described in [26] that models the complete imaging process and compensates for degrading factors such as scattering [26, 118–120]. This approach models the image creation process starting with the X-ray tube voltage and ending with the film or sensor characteristics. The resulting image, known as the h_{int} representation, estimates the height of non-fatty tissue in the breast for each pixel in the image. This representation is intrinsic to the breast with the contribution of the imaging system removed. It estimates the height of "interesting" tissue such as fibrous/gradular tissue and cancerous tissue.

The h_{int} representation allows for more accurate analysis of the mammogram. For example, masses are represented by large changes in the amount of interesting tissue and hence they can be analyzed more reliably than using general "enhanced" images.

In our work, we used a simplified transform based on a mono-energetic h_{int} model and an enhancement step to further remove the background. We denote the enhanced

image as I_E: $I_E = h_{int} - 0.75 * LP(h_{int})$, where $LP(h_{int})$ is a low-pass filtered version of h_{int} using a 31×31 window averaging filter.

The following shows how the mono-energetic h_{int} is obtained: The film photo energy E_p can be obtained from the gray-level (GL) value.

$$D(x,y) = \frac{1}{\alpha}(GL(x,y) - \eta) \tag{3.1}$$

$$E_p(x,y) = \frac{1}{\beta}10^{D(x,y)/\gamma} \tag{3.2}$$

where $GL(x,y)$ and $D(x,y)$ are the gray level value and film density at pixel coordinate(x,y), α, η, β, and γ are the parameters.

$E_p(x,y)$ can then be used to derive h_{int} with the following expression [26].

$$E_p(x,y) = \phi(x,y)A_p t_s ES(E)G(E)e^{-\mu_l h_l}e^{-h_{int}(\mu_{int}-\mu_{fat})-H\mu_{fat}} \tag{3.3}$$

where ϕ is the photon flux, A_p is the pixel area, t_s is the film exposure time, E is the energy, $S(E)$ is the absorption ratio of the screen to primary photons, G(E) is the transmission ratio of the grid for primary photons, μ_l is the linear attenuation of lucite, h_l is the thickness of a lucite compression plate, μ_{fat} and μ_{int} are the attenuation rate of fat tissue and interesting tissue separately, and H is the thickness of compressed breast between two lucite plates. We then have

$$h_{int}(x,y) = \frac{H\mu_{fat}}{\mu_{fat}-\mu_{int}} + \frac{\ln E_p(x,y) - \ln(\phi(x,y)A_p t_s ES(E)G(E)e^{-\mu_l h_l})}{\mu_{fat}-\mu_{int}} \tag{3.4}$$

$$= \frac{H\mu_{fat}}{\mu_{fat}-\mu_{int}} + \frac{\ln E_p(x,y) + C_{phys}}{\mu_{fat}-\mu_{int}} \tag{3.5}$$

C_{phys} represents the contribution of physical parameters of image formation, which could be compensated with H. We used the following parameters in our work:

$$\alpha = -196, \eta = 490, \tag{3.6}$$

$$\beta = 1.648 \times 10^{11}, \gamma = 3.0, \tag{3.7}$$

$$\mu_{fat} = 0.322, \mu_{int} = 0.506, \tag{3.8}$$

$$H = 75 \tag{3.9}$$

All of features will be extracted from I_E. Figure 3.2 shows two mammograms and their enhanced images based on the h_{int} representation.

3.3 Feature Extraction

The first step in normal analysis is to extract characteristic features from I_E. From the perspective of pattern classification, feature extraction is a very important step in that the ultimate performance of the system is not determined by optimal parameters of the classifier, but by the intrinsic separability of the feature vectors [115].

The characterization of normal tissue poses a real challenge due to the complexity of normal tissues and the fact that a normal mammogram is not well defined [7, 27]. Subtle breast cancers may not be easily distinguished from the surrounding normal tissue. The heterogenous nature of different breast cancers of different sizes also poses real challenges for feature extraction.

In this section, we will describe several feature sets that we believe can be used to separate normal and abnormal regions. All of our features are extracted from 512×512 regions of I_E. There are four types of features extracted from each region as shown in Figure 3.1.

3.3.1 Curvilinear Features

Though normal breast tissue may have very different appearance, unequivocally normal breast areas are characterized by curvilinear markings. These curvilinear structures are the ductal structures of the breast tissue, which are lactation pathways [7]. The curvilinear markings are not randomly oriented, but rather tend to radiate from the nipple toward the chest wall. If a tumor or a scar appears in an area, the surrounding curvilinear structure is disturbed, usually appearing as a random pattern or is partially absent. Curvilinear structures have been extensively studied for characterizing normal breast tissues [26,81,104,110–112]. Because the curvilinear structures are not straight lines, and are "noisy" along the line directions, general line or edge detection algorithms, such as the Hough Transform [121], will not be adequate to capture the curvilinear structures. We used a line detection algorithm

we previously developed [81, 104] to extract curvilinear structures characteristic of normal breast tissue. The algorithm is based on a model of a line as a string of pixels having similar values, but having significantly different contrast to the surrounding pixels. The standard deviation is a good measure of gray level similarity among pixels. Let $f(i,j)$ be the pixel graylevel at spatial location (i,j); and $L(\theta,l)$ be a string of pixels in the direction θ and of length l; and $N_{L(\theta,l)}$ be the number of pixels within $L(\theta,l)$. Then the standard deviation of the pixel gravlevel in $L(\theta,l)$ is

$$\sigma(\theta,l)=\sqrt{\frac{1}{N_{L(\theta,l)}-1}\sum_{(m,n)\in L(\theta,l)}(f(m,n)-\bar{f}_{L(\theta,l)})^2} \tag{3.10}$$

where $\bar{f}_{L(\theta,l)}$ is the average gray level within $L(\theta,l)$

$$\bar{f}_{L(\theta,l)}=\frac{1}{N_{L(\theta,l)}}\sum_{(m,n)\in L(\theta,l)}f(m,n) \tag{3.11}$$

Let $\sigma_{i,j}(\theta,l)=\min_{(i,j)\in L(\theta,l)}\sigma(\theta,l)$, and $\sigma_{i,j}(l)=\min_{\theta}\sigma_{i,j}(\theta,l)$. If pixel (i,j) belongs to a line in direction θ_* and of length greater than l, then $\sigma_{i,j}(\theta_*,l)$ is small. Hence, the smaller $\sigma_{i,j}(l)$ is, the larger the probability that pixel (i,j) is on a line.

The measure of surrounding pixel difference can be obtained from the standard deviation of $\sigma_{i,j}(\theta,l)$ with regard to θ

$$\sigma_{\sigma,(i,j)}(l)=\sqrt{\frac{1}{N_\theta-1}\sum_{\theta}(\sigma_{i,j}(\theta,l)-\bar{\sigma}_{i,j}(\theta,l))^2} \tag{3.12}$$

where N_θ is the total number of directions, and $\bar{\sigma}_{i,j}(\theta,l)$ is the average

$$\bar{\sigma}_{i,j}(\theta,l)=\frac{1}{N_\theta}\sum_{\theta}\sigma_{i,j}(\theta,l) \tag{3.13}$$

The larger $\sigma_{\sigma,(i,j)}(l)$, the greater the surrounding pixel difference. Finally, each pixel (i,j) is determined to be as a line pixel or not according to the following rule:

$$CL_{bin}(i,j)=\begin{cases}1(\text{line}) & \text{if } \sigma_{i,j}(l)<T_\sigma \text{ and } \sigma_{\sigma,(i,j)}(l)>T_{\sigma_\sigma}\\ 0(\text{not}) & \text{otherwise}\end{cases} \tag{3.14}$$

where T_σ and T_{σ_σ} are thresholds determined experimentally. The algorithm is robust to noise and is capable of extracting quasi-linear curves with different widths

and angles. It is expected that abnormal regions have less curvilinear structures, especially circumscribed mass regions. The following parameters were used in our work: line length $l = 20$, $N_\theta = 16$, $T_\sigma = 23$ and $T_{\sigma_\sigma} = 3.5$.

With our line detection algorithm, we obtained a binary image CL_{bin} indicating whether a pixel in the region belongs to a curvilinear structure, and an angle map Ang_{map} indicating the line direction of a curvilinear pixel. Figure 3.3 shows an example of a CL_{bin} and Ang_{map}. A set of features were extracted from the detected curvilinear structure. These features captured the statistical summary of the curvilinear pixels in the region. A total of 18 curvilinear features was extracted:

- *LinePixelCount*
- *A: Upper right half line pixel count*
- *B: Lower left half line pixel count*
- *C: Upper left half line pixel count*
- *D: Lower right half line pixel count*
- *HalfRatio*
- *HalfRatio2*
- *AngMean*
- *AngStd*
- *LocalLineMean*
- *LocalLineStd*
- *LocaLineEntropy*
- *LocalAngAve*
- *LocalAngStd*

- *LocalAngEntropy*
- *LBPMean*
- *LBPStd*
- *LBPEntropy*

We will describe each feature below. From the binary curvilinear structure of a region (size 512×512), Feature *LinePixelCount* is defined as the total number of curvilinear pixels in the region. Feature A is line pixel count in the upper right diagonal half of a region, $A = \sum\sum_{i<j} CL_{bin}(i,j)$; feature B is line pixel count in the lower left diagonal half of a region, $B = \sum\sum_{i\geq j} CL_{bin}(i,j)$; feature C is line pixel count in the upper left diagonal half of a region, $C = \sum\sum_{i+j<512} CL_{bin}(i,j)$; and feature D is line pixel count in the lower right diagonal half of a region, $D = \sum\sum_{i+j\geq 512} CL_{bin}(i,j)$.

Feature *HalfRatio* is defined as $HalfRatio = \frac{A}{B}$, and feature *HalfRatio2* is defined as $HalfRatio2 = \frac{C}{D}$.

Feature *AngMean* is the average line angle of the curvilinear pixels in the region; feature *AngStd* is the standard deviation of the angle of the curvilinear pixels in the region. Both are obtained from the angle map Ang_{map}.

There are 6 localized Features: *LocalLineMean, LocalLineStd, LocalLineEntropy, LocalAngAve, LocalAngStd, LocalAngEntropy.* The binary curvilinear image CL_{bin} and the angle map Ang_{map} of a region are dissected into 8×8 disjoint sub-blocks, with 4096 sub-blocks since the region size is 512×512. We define l_i as the line pixel count of each sub-block of CL_{bin}, and a_i as the average angle of each sub-block of Ang_{map}, $i = 1, 2, ..., 4096$. Then we can define 6 localized features from the histogram of l_i and a_i. Figure 3.4 shows the diagram of the sub-block scheme used for determining the 6 localized features.

Histograms of 12 bins are then obtained for l_i and a_i. For value l_i's, we obtain the relative frequency ${p_j}^l$ and bin value ${x_j}^l$ at bin j, $j = 1, 2, ..., 12$. Similarly, we

have the relative frequency ${p_j}^a$ and bin value ${x_j}^a$ at bin j, $j = 1, 2, ...12$ for a_i's. Feature *LocalLineMean* is the average of l_i's:

$$LocalLineMean = \sum_{j=1}^{12} {p_j}^l \, {x_j}^l \tag{3.15}$$

Feature *LocalLineStd* is defined as the standard deviation of l_i's:

$$LocalLineStd = \sum_{j=1}^{12} {p_j}^l \, ({x_j}^l \, - \, LocalLineMean)^2 \tag{3.16}$$

Feature *LocalLineEntropy* is defined as the entropy of l_i's:

$$LocalLineEntropy = \sum_{j=1}^{12} - {p_j}^l \, \log {p_j}^l \tag{3.17}$$

Similarly, we define 3 localized features from a_i:

$$LocalAngAve = \sum_{j=1}^{12} {p_j}^a \, {x_j}^a \tag{3.18}$$

$$LocalAngStd = \sum_{j=1}^{12} {p_j}^a \, ({x_j}^a \, - \, LocalAngAve)^2 \tag{3.19}$$

$$LocalAngEntropy = \sum_{j=1}^{12} - {p_j}^a \, \log {p_j}^a \tag{3.20}$$

The last 3 curvilinear features are extracted from the Local Binary Pattern (LBP) [122,123]. A LBP is obtained from the dot product of a rectangular template and a binary region of the same size. Figure 3.5 illustrates how a LBP is obtained from a 3×3 rectangular template on a 3×3 binary region. The LBP is obtained from the dot product of two 3×3 matrices, $1 * 1 + 2 * 1 + 4 * 0 + 128 * 0 + 8 * 0 + 64 * 1 + 32 * 1 + 16 * 1 = 115$. Moving the center of the rectangular template over each pixel a binary image, we can obtain a LBP for each pixel. A histogram of LBP could be used to represent and analyze the binary image. Using 3×3 rectangular template, it can generate a histogram of 256 LBP.

The Local Binary Pattern (LBP) is computationally simple, and provides highly discriminative texture information, such as edges. One of advantage of the LBP is that different templates could be designed to capture binary patterns. Using the

template shown in Figure 3.5, we obtained LBP values for disjoint 3×3 neighborhoods. A histogram of 12 bins is generated, with relative frequency ${p_j}^{lbp}$ of bin value ${x_j}^{lbp}$ at bin j, $j = 1, 2, ...12$. Feature *LBPMean, LBPStd, LBPEntropy* are defined as:

$$LBPMean = \sum_{j=1}^{12} {p_j}^{lbp} {x_j}^{lbp} \tag{3.21}$$

$$LBPStd = \sum_{j=1}^{12} {p_j}^{lbp} ({x_j}^{lbp} - LBPMean)^2 \tag{3.22}$$

$$LBPEntropy = \sum_{j=1}^{12} - {p_j}^{lbp} \log {p_j}^{lbp} \tag{3.23}$$

Figure 3.6 shows the histograms of feature *LinePixelCount, Angstd* for the data set of 296 normal and 164 abnormal regions. Figure 3.7 shows the scatter plots of two pairs of curvilinear features. We have found that features from line angles play an equal role as the features from line densities in classification. This is understandable since ductal structure has a direction: from nipple toward the chest wall.

3.3.2 Gray Level Co-occurrence Features

Texture represents the spatial arrangement of the pixels in a region. Characterization of spatial patterns can be adequately specified by a 2D spatial dependence matrix known as the Gray Level Co-occurrence Matrix (GLCM) [48]. Each entry (i, j) of the matrix at row i and column j is the relative frequency of occurrence of pairwise gray levels i, j separated by a distance d and a direction α. The distance d can be adjusted to match the size of basic texture elements in the image. The direction parameter α can be specified to the direction of the spatial repetition period of the basic texture element.

Let $GL(m, n)$ be the gray level of pixel (m, n), $P(i, j, d, \alpha)$ be the number of occurrence of pairwise gray levels i, j separated by a distance d and at direction α,

and $p(i,j,d,\alpha)$ be the relative frequency corresponding to $P(i,j,d,\alpha)$. The following shows how to obtain each entry $p(i,j,d,\alpha)$ of the GLCM for an image of size $H \times W$.

$$\begin{aligned} P(i,j,d,\alpha) &= \#\{((k,l),(m,n)) \in (H \times W) \times (H \times W) | \\ &\quad k - m = \lceil d \sin\alpha \rceil, l - n = \lceil d \cos\alpha \rceil, \\ &\quad GL(k,l) = i, GL(m,n) = j\} \\ p(i,j,d,\alpha) &= \frac{P(i,j,d,\alpha)}{N_{d,\alpha}}, \text{ where } N_{d,\alpha} = \sum_i \sum_j P(i,j,d,\alpha) \end{aligned}$$

$\#$ denotes the number of elements in the set. An isotropic GLCM with $d = 1$ was obtained as an average matrix from four matrices, at $\alpha = 0°, 45°, 90°$, and $135°$, i.e.

$$p(i,j) = \frac{1}{4}(p(i,j,1,0°) + p(i,j,1,45°) + p(i,j,1,90°) + p(i,j,1,135°)) \tag{3.24}$$

We can then obtain the estimated marginal probabilities from $p(i,j)$,

$$p_x(i) = \sum_{j=0}^{N-1} p(i,j) \tag{3.25}$$

$$p_y(j) = \sum_{i=0}^{N-1} p(i,j) \tag{3.26}$$

where N is the number of distinct gray levels. From the isotropic GLCM $p(i,j)$ of $d = 1$, we extracted 16 features, including features defined by Haralick [48] and additional cluster features defined in [124].

- *Energy*
- *Entropy*
- *MaxProb*
- *Correlation*
- *DiagCorr*
- H_{xy1}
- H_{xy2}

- *DEnergy*
- *DEntropy*
- *Inertia*
- *Homogeneity*
- *SEnergy*
- *SEntropy*
- *SVar*
- *SShade*
- *SProm*

Energy, Entropy, hxy1, hxy2, MaxProb are defined as

$$Engergy = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)^2 \tag{3.27}$$

$$Entropy = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} -p(i,j)\log\, p(i,j) \tag{3.28}$$

$$H_{xy1} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} -p(i,j)\log(p_x(i)p_y(j)) \tag{3.29}$$

$$H_{xy2} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} -p_x(i)p_y(j)\log(p_x(i)p_y(j)) \tag{3.30}$$

$$MaxProb = \max p(i,j) \tag{3.31}$$

Correlation, DiagCorr are defined as

$$Correlation = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{p(i,j)(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y} \tag{3.32}$$

$$DiagCorr = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)|i-j|(i+j-\mu_x-\mu_y) \tag{3.33}$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the means and standard deviations of $p_x(i), p_y(j)$, respectively.

The gray level difference histogram (GLDH) is defined as

$$D(k) = \sum_{i=0}^{N-1} \sum_{\substack{j=0 \\ |i-j|=k}}^{N-1} p(i,j), \quad k = 0, 1, ..., N-1 \tag{3.34}$$

The following four features *DEnergy, DEntropy, Inertia, and Homogeneity* are defined from GLDH:

$$DEnergy = \sum_{k=0}^{N-1} D(k)^2 \tag{3.35}$$

$$DEntropy = \sum_{k=0}^{N-1} -D(k) \log D(k) \tag{3.36}$$

$$Inertia = \sum_{k=0}^{N-1} k^2 D(k) \tag{3.37}$$

$$Homogeneity = \sum_{k=0}^{N-1} \frac{D(k)}{1+k^2} \tag{3.38}$$

The gray level sum histogram (GLSH) is defined as

$$S(k) = \sum_{i=0}^{N-1} \sum_{\substack{j=0 \\ |i+j|=k}}^{N-1} p(i,j), \quad k = 0, 1, ..., 2(N-1) \tag{3.39}$$

From GLSH, we can define 5 features:

$$SEnergy = \sum_{k=0}^{2(N-1)} S(k)^2 \tag{3.40}$$

$$SEntropy = \sum_{k=0}^{2(N-1)} -S(k) \log S(K) \tag{3.41}$$

$$SVar = \sum_{k=0}^{2(N-1)} (k-\mu)^2 S(k), \text{ where } \mu = \sum_{k=0}^{2(N-1)} kS(k) \tag{3.42}$$

$$SShade = \sum_{k=0}^{2(N-1)} \frac{(k-\mu_x-\mu_y)^3 S(k)}{(\sigma_x{}^2 + \sigma_y{}^2 + 2\rho\sigma_x\sigma_y)^{\frac{3}{2}}} \tag{3.43}$$

$$SProm = \sum_{k=0}^{2(N-1)} \frac{(k-\mu_x-\mu_y)^4 S(k)}{(\sigma_x{}^2 + \sigma_y{}^2 + 2\rho\sigma_x\sigma_y)^2} \tag{3.44}$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are means and standard deviations of $p_x(i), p_y(j)$, respectively, and ρ is the *Correlation* as defined in equation 3.32. The last two higher order features *SShade* and *SProm* were defined in [124], and they represent cluster shade and cluster prominence, respectively.

Figure 3.8 shows the histogram of the two features, *Homogeneity* and H_{xy1}. Figure 3.9 shows the scatter plots of some feature pairs. We can observer separation between normal and abnormal classes. We have tested different distances d, and $d = 1$ gave the best result in the experiments in terms of the classification separability. Figure 3.10 shows the classification performance and distance d relationship, where A_z is the area under the Receiver Operating Characteristic (ROC) curve.

3.3.3 Gabor Features

2D Gabor filters describe the 2D receptive-field profiles of simple cells found in the visual cortex of vertebrate animals. They are consistent with the human vision system (HSV) [125]. Gabor filters have been successfully used in describing texture information [126–128].

A Gabor filter has as its impulse response a Gaussian modulated sinusoidal plane wave:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \tag{3.45}$$

where W is the modulation frequency, x, y are coordinates in the spatial domain, and σ_x and σ_y are the standard deviations in the x and y direction. Its frequency response is:

$$G(u,v) = exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\} \tag{3.46}$$

where $\sigma_u = \frac{1}{2\pi\sigma_x}$, $\sigma_u = \frac{1}{2\pi\sigma_y}$, and u and v are coordinates in the frequency domain. A Gabor filter-bank consists of Gabor filters with Gaussians of different sizes modulated by sinusoidal plane waves of different orientation from the same mother Gabor filter as defined in equation (3.45):

$$g_{m,n}(x,y) = a^{-m} g(\tilde{x}, \tilde{y}), \quad a > 1 \tag{3.47}$$

where $\tilde{x} = a^{-m}(x\cos\theta + y\sin\theta)$, $\tilde{y} = a^{-m}(-x\sin\theta + y\cos\theta)$, $\theta = n\pi/K$ $(K = \text{total orientation}, \text{and}\, n = 0, 1, ..., K-1)$, and $g(\cdot,\cdot)$ is defined in equation (3.45). Given an image $I_E(r,c)$ of size $H \times W$, the discrete Gabor filtered output is given

by a 2D convolution, which is usually implemented in the frequency domain simply as the product:

$$I_{g_{m,n}}(r,c) = \sum_s \sum_t I_E(r-s, c-t) {g_{m,n}}^*(s,t),\ m = 0,1,...,S-1, n = 0,...,K-1 \tag{3.48}$$

where * indicates the complex conjugate. We can obtain the mean and standard deviation of the energy of the filtered image, which are often used as Gabor features.

$$\mu_{mn} = \frac{\sum_r \sum_c |I_{g_{m,n}}(r,c)|}{H \times W} \tag{3.49}$$

$$\sigma_{mn} = \frac{\sqrt{\sum_r \sum_c (|I_{g_{m,n}}(r,c)| - \mu_{mn})^2}}{H \times W} \tag{3.50}$$

Then we have a feature vector of size $S \times K$:

$$f = [\mu_{00} \quad \sigma_{00} \quad \cdots \quad \mu_{(S-1)(K-1)} \quad \sigma_{(S-1)(K-1)}]$$

The following design guarantees the adjacent half-peak contours touch each other [126], after choosing the number of orientations K, the number of scales S, and the upper and lower center frequencies U_h and U_l:

$$a = \left(\frac{U_h}{U_l}\right)^{\frac{1}{S-1}},\ \sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2\ln 2}}$$
$$\sigma_v = \tan\left(\frac{\pi}{2K}\right)\sqrt{\frac{U_h^2}{2\ln 2} - \sigma_u^2},\ W = U_h$$
$$m = 0,1,\cdots,S-1, \quad n = 0,1,\cdots,K-1$$

The advantage of a Gabor filter-bank is that it provides simultaneous localization in both the spatial and frequency domains. Figure 3.11 shows the real part of a Gabor filter-bank of four scales and four orientations with $a = 2$, with desirable characteristics of spatial locality and orientation selectivity. In our work, the highest and lowest frequencies of the Gabor filter-bank were experimentally chosen to suit mammogram region analysis. We chose 4 orientations and 4 scales for our Gabor filter-bank, i.e. 16 Gabor filters. The following parameters were used in our work: $S = 4$, $K = 4$, $U_h = \frac{\sqrt{2}}{16}$, and $U_l = \frac{\sqrt{2}}{128}$. We have $a = 2$.

We obtained the mean and standard deviation of the energy of each Gabor filtered image as features. Therefore, there were 32 Gabor features extracted from each region image:

$$f = [\mu_{00} \quad \sigma_{00} \quad \cdots \quad \mu_{33} \quad \sigma_{33}]$$

Figure 3.12 shows the histograms of two selected features, μ_{30} and σ_{31}. Figure 3.13 shows the scatter plots of two feature pairs.

3.3.4 Multiresolution Features

The last set of features were obtained from nonseparable wavelet decompositions of the mammogram. A nonseparable wavelet transform, the Quincunx Wavelet transform [129–134], is used in our research. A 2D quincunx wavelet transform is implemented with low and high pass filter banks similar to a 2D separable wavelet transform, the difference is that the low and high-pass kernels can not be separated into two one-dimensional kernels. Although separable wavelet transforms have a simple and well understood implementation, there are some considerations in using a non-separable wavelet decomposition: separable wavelet decompositions have vertical and horizontal cut-offs while the non-separable decomposition can have a cut-off at any angle, moreover, non-separable filter banks can be flexibly tailored for particular purposes, such as having linear phases.

The high and low pass filter banks used for 2D Quincunx wavelet decomposition in our research are

$$h_0(n_1, n_2) = \begin{pmatrix} & 1 & \\ 1 & 4 & 1 \\ & 1 & \end{pmatrix} \tag{3.51}$$

and

$$g_0(n_1, n_2) = \begin{pmatrix} & & 1 & & \\ & 2 & -4 & 2 & \\ 1 & -4 & -28 & -4 & 1 \\ & 2 & -4 & 2 & \\ & & 1 & & \end{pmatrix} \tag{3.52}$$

In quincunx down-sampling, each subsequent low-pass image is reduced by a factor of $\frac{1}{\sqrt{2}}$ in each dimension. The dilation matrix D_q for quincunx lattice is

$$D_q = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{3.53}$$

Down-sampling means that the pixels on the sampling lattice are kept while all others are discarded, which can be characterized by the following equation:

$$Y(\mathbf{n}) = X(D\mathbf{n}) \tag{3.54}$$

where D is the dilation matrix for the lattice. In the quincunx lattice, it becomes

$$Y(n_1, n_2) = X(n_1 + n_2, n_1 - n_2) \tag{3.55}$$

When performing the next level decomposition, we iterate the filter bank and the down-sampling by D_q. The overall down-sampling is characterized by an integer power of D_q. Since

$$D_q^2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 2I \tag{3.56}$$

Hence, the down-sampling of the odd level decomposition can be characterized by

$$D_q^{2k+1} = (D_q^2)^k D_q = (2I)^k D_q = 2^k D_q, \quad k = 0, 1, 2, \cdots \tag{3.57}$$

and the down-sampling of the even level decomposition can be characterized by

$$D_q^{2k} = (D_q^2)^k = (2I)^k = 2^k I, \quad k = 1, 2, \cdots \tag{3.58}$$

One of the significant difference from separable down-sampling is the change in basis which arises. The rotation of the image occurs due to a changes of basis. The

down-sampling rotates the image by $\frac{\pi}{4}$ and flips the image around the horizontal axis. The decomposition image appears in diamond-shape. According to equations 3.57 and 3.58, at the odd-level decompositions the orientation is rotated and the image is in diamond shape, however, the orientation is normal at even decomposition levels. Figure 3.14 shows the low pass decomposition sequences.

Only the first four even-level low-pass Quincunx wavelet decomposition images (images have normal orientation), i.e. images of spatial resolutions (256×256), (128×128), (64×64) and (32×32), are retained for feature extraction. Five features *MeaN, VariancE, SkewnesS, KurtosiS, and EntropY* are extracted from each decomposition image, hence there will be a total of 20 features extracted from first four even-level wavelet transform. These features are:

- $MeaN_2$
- $VariancE_2$
- $SkewnesS_2$
- $KurtosiS_2$
- $EntropY_2$
- $MeaN_4$
- $VariancE_4$
- $SkewnesS_4$
- $KurtosiS_4$
- $EntropY_4$
- $MeaN_8$
- $VariancE_8$
- $SkewnesS_8$

- $KurtosiS_8$
- $EntropY_8$
- $MeaN_{16}$
- $VariancE_{16}$
- $SkewnesS_{16}$
- $KurtosiS_{16}$
- $EntropY_{16}$

Feature $MeaN_k$ $(k = 2^L, L = 1, 2, 3, 4)$ is defined as the global average pixel value of the decomposition image at even level L.

$$MeaN_k = \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} X_L(i, j), \quad \text{where } M = 512/k \tag{3.59}$$

where M is the size (height or width) of the decomposition image at even level L. Since each mammogram is normalized to the optical density, the average pixel value can be used as a feature. Breast tumors appear brighter in a mammogram than the surrounding normal breast tissues, which makes the average gray level a valuable feature [76].

Feature $VariancE_k$ $(k = 2^L, L = 1, 2, 3, 4)$ is the standard deviation defined as:

$$VariancE_k = \sqrt{\frac{1}{M^2 - 1} \sum_{i=1}^{M} \sum_{j=1}^{M} (X_L(i, j) - MeaN_k)^2}, \quad \text{where } M = 512/k \tag{3.60}$$

$Variance_k$ is also a measure of pixel value variation. In a general sense, a normal region generally have less pixel variation than an abnormal region.

Feature $SkewnesS_k$ $(k = 2^L, L = 1, 2, 3, 4)$ is the third central moment, and defined as

$$SkewnesS_k = \frac{1}{M^2 - 1} \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} (X_L(i, j) - MeaN_k)^3}{{VariancE_k}^3}, \quad \text{where } M = 512/k \tag{3.61}$$

Skewness indicates the asymmetry of the tails of the distribution. A Gaussian distribution has a skewness of 0. A significant negative number indicates a long left-hand tail, and a large positive number indicates a long right-hand tail.

Feature $KurtosiS_k$ $(k = 2^L, L = 1, 2, 3, 4)$ is fourth central moment, and defined as

$$KurtosiS_k = \frac{1}{M^2 - 1} \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} (X(i,j) - MeaN_k)^4}{VariancE_k{}^4}, \quad \text{where } M = 512/k \quad (3.62)$$

Kurtosis indicates the flatness or sharpness of the distribution. A Guassian distribution has a kurtosis of 3. Values less than 3 indicates a relatively flat distribution, and greater than 3 indicates a spiked distribution. Kurtosis estimates the flatness of the histogram.

Feature $EntropY_k$ $(k = 2^L, L = 1, 2, 3, 4)$ has the same definition as in information theory. Obtain the histogram with 12 bins for a decomposition image at even level L, and the relative frequency ${p_i}^L, i = 1, 2, \cdots, 12$ for each bin. Then, the $EntropY_k$ is defined as:

$$EntropY_k = -\sum_{i=1}^{12} {p_i}^L \log {p_i}^L \quad (3.63)$$

Figure 3.15 shows histograms of feature $MeaN_{16}$ and $EntropY_8$ for the training data set of 296 normal and 164 abnormal regions. Figure 3.16 shows the scatter plot of some feature pairs.

3.3.5 Feature Summary

The above four sets of features combined into a 86-feature vector associated with each 512×512 region. The following is the complete list of our 86 features:

- **Curvilinear Features**: *LinePixelCount, A, B, C, D, HalfRatio, HalfRatio2, AngMean, AngStd, LocalLineMean, LocalLineStd, LocaLineEntropy, LocalAngAve, LocalAngStd, LocalAngEntropy, LBPMean, LBPStd, LBPEntropy*

- **GLCM Features**: *Energy, Entropy, MaxProb, Correlation, DiagCorr,* H_{xy1}, H_{xy2}, *DEnergy, DEntropy, Inertia, Homogeneity, SEnergy, SEntropy, SVar, SShade, SProm*

- **Gabor Features**: $\mu_{00} \quad \sigma_{00} \quad \cdots \quad \mu_{33} \quad \sigma_{33}$

- **Multiresolution Features**: $MeaN_2$, $VariancE_2$, $SkewnesS_2$, $KurtosiS_2$, $EntropY_2$, $MeaN_4$, $VariancE_4$, $SkewnesS_4$, $KurtosiS_4$, $EntropY_4$, $MeaN_8$, $VariancE_8$, $SkewnesS_8$, $KurtosiS_8$, $EntropY_8$, $MeaN_{16}$, $VariancE_{16}$, $SkewnesS_{16}$, $KurtosiS_{16}$, $EntropY_{16}$

Normal and abnormal classes are heterogeneous. A normal class consists of normal regions of different densities and complexities. An abnormal class is more diverse because it is a set of different breast cancers, such as microcalcifications and spiculated lesions. It is expected that the distribution of a feature may not be uni-modal and may not behave similar to a Gaussian due to the heterogeneous natures of normal and abnormal classes. From the histograms and scatter plots of some of the features, we observe that some distributions do appear as multi-modes.

3.4 Power Function Transform for Feature Normalization

Feature preprocessing is usually used to "enhance" features for the classification task. Features are usually transformed to normalize the scatter of the distribution and enhance the separation distance between the two classes. One of the well-known transforms is the "whitening" transform used to make the transformed features "independent."

The power function can be used to normalize two distributions to have similar variance, therefore reduce the effects of outliers. To simplify the transform, each feature in the feature vector is transformed independently. Let x_i, $i = N$ or S be one of the feature vectors, where N is the class of normal regions and S is the class

of abnormal regions, and y_i be the transformed feature vector, then we have the following power transform:

$$y_i = \alpha\, x_i^{\beta}, \quad i = N \text{ or } S, \quad 0.2 < \beta < 5 \tag{3.64}$$

α and β are not pre-set, and they are obtained from the mean and standard deviation of each feature in the training set:

$$\beta = \frac{\log(\frac{\sigma_S}{\sigma_N})}{\log(\frac{\mu_N}{\mu_S})} + 1, \quad \mu_N \neq \mu_S \tag{3.65}$$

$$\alpha = \exp\left\{\left[\left(\frac{\beta - 1}{2}\right)[\log \mu_N + \log \mu_S] + \frac{\log \sigma_N + \log \sigma_S}{2} + \log \beta\right]\right\} \tag{3.66}$$

where μ_N and σ_N are the mean and standard deviation of class N estimated from the training set. Similarly, μ_S and σ_S are the mean and standard deviation of class S. Note that α and β are the same for class N and S for a given feature, so an unknown unclassified sample can be transformed without knowledge of the class. After the power transformation, the distributions of the transformed feature for the two classes have approximately equal unit variance.

Figure 3.17 illustrates the power transformation. The power transform is used to normalize each feature so as to have similar standard deviations, which reduces the risk of over-fitting the outliers.

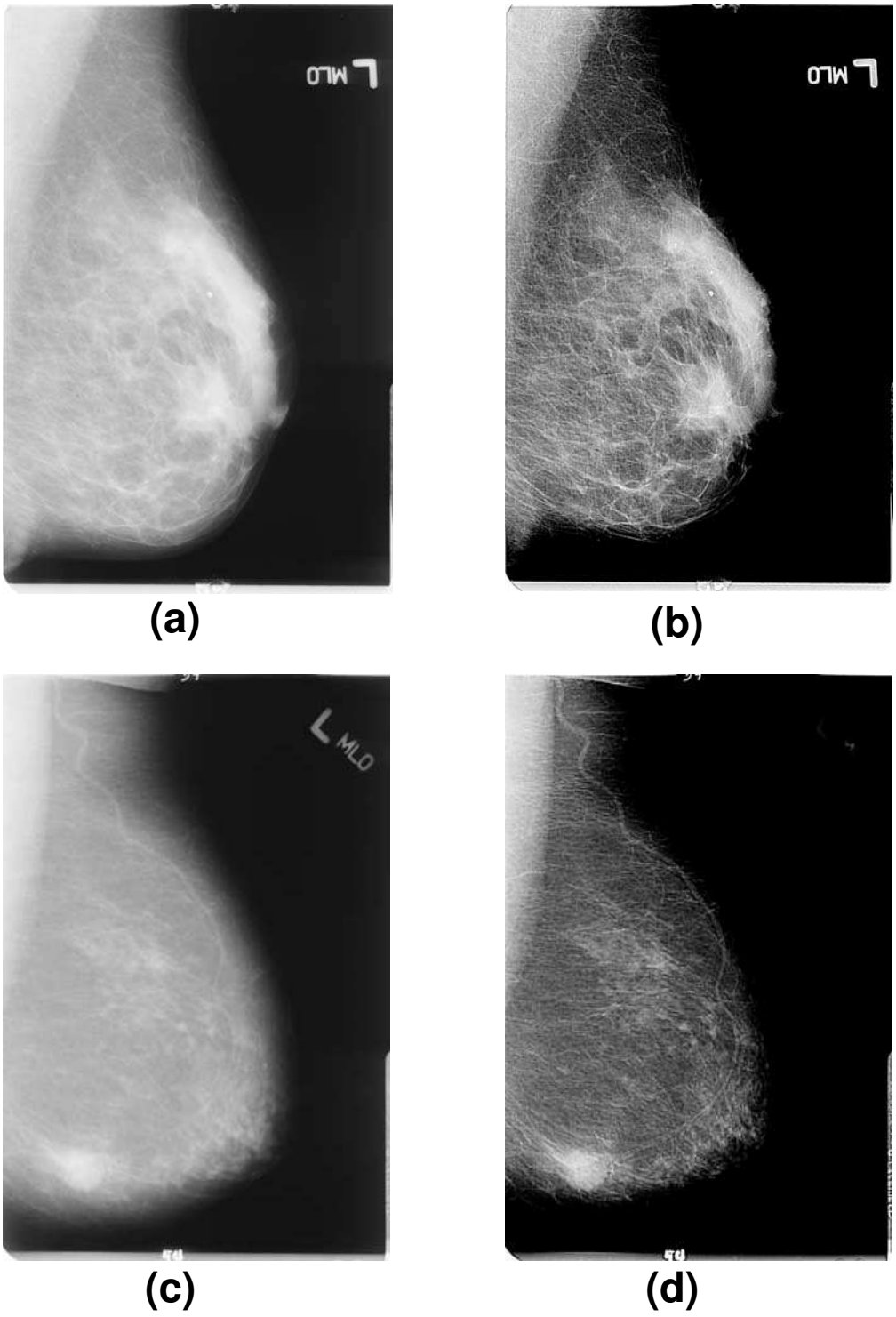


Fig. 3.2. (a)A screening mammogram; (b) I_E representation of (a); (c) Another screening mammogram; (d) I_E representation of (c).

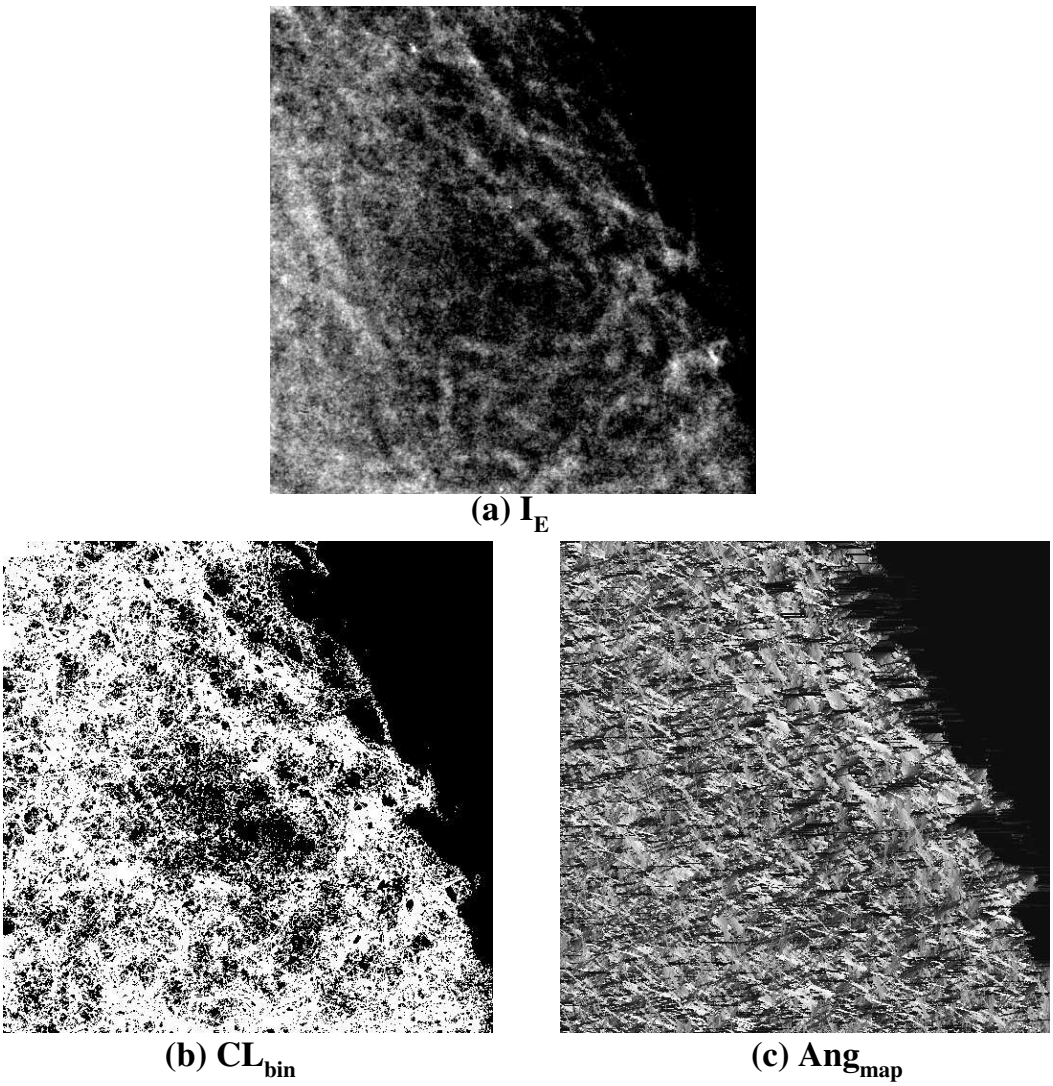


Fig. 3.3. (a) an I_E; (b) CL_{bin} of (a); (c) is the mapped 256 graylevel display of Ang_{map} of (a).

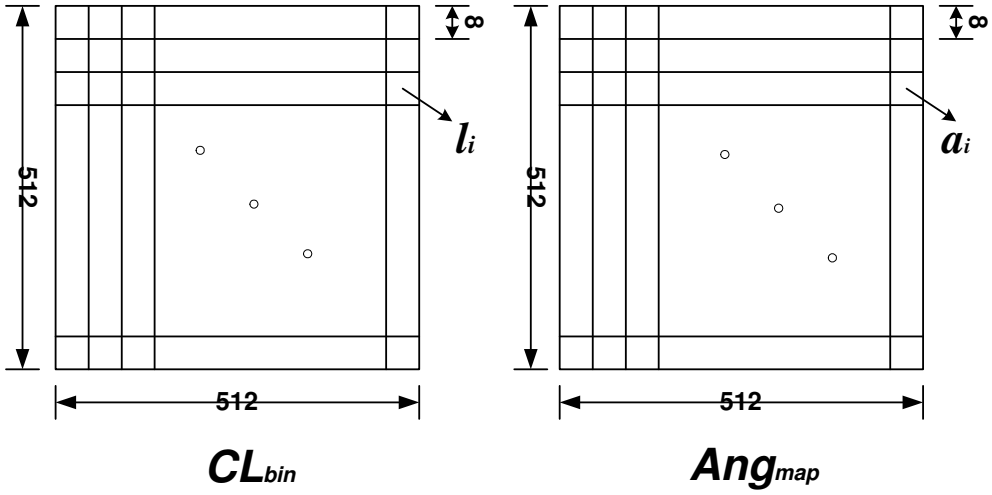


Fig. 3.4. 8×8 sub-blocks used for the 6 localized features.

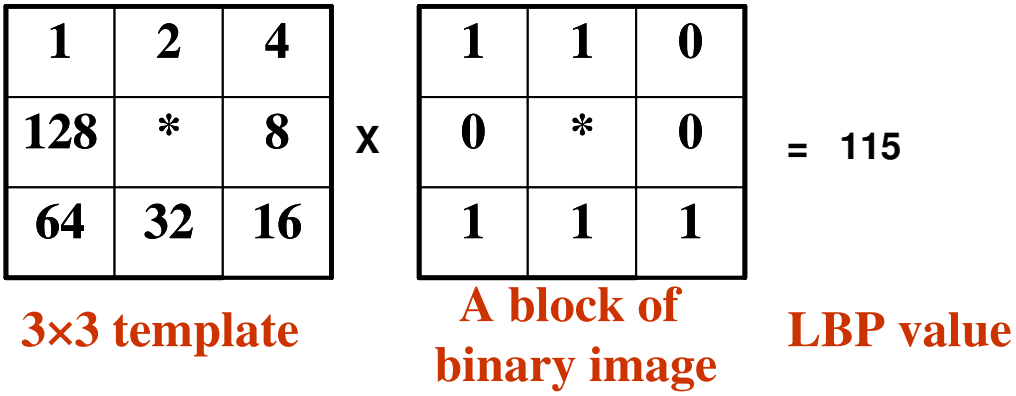


Fig. 3.5. Local Binary Pattern Illustration

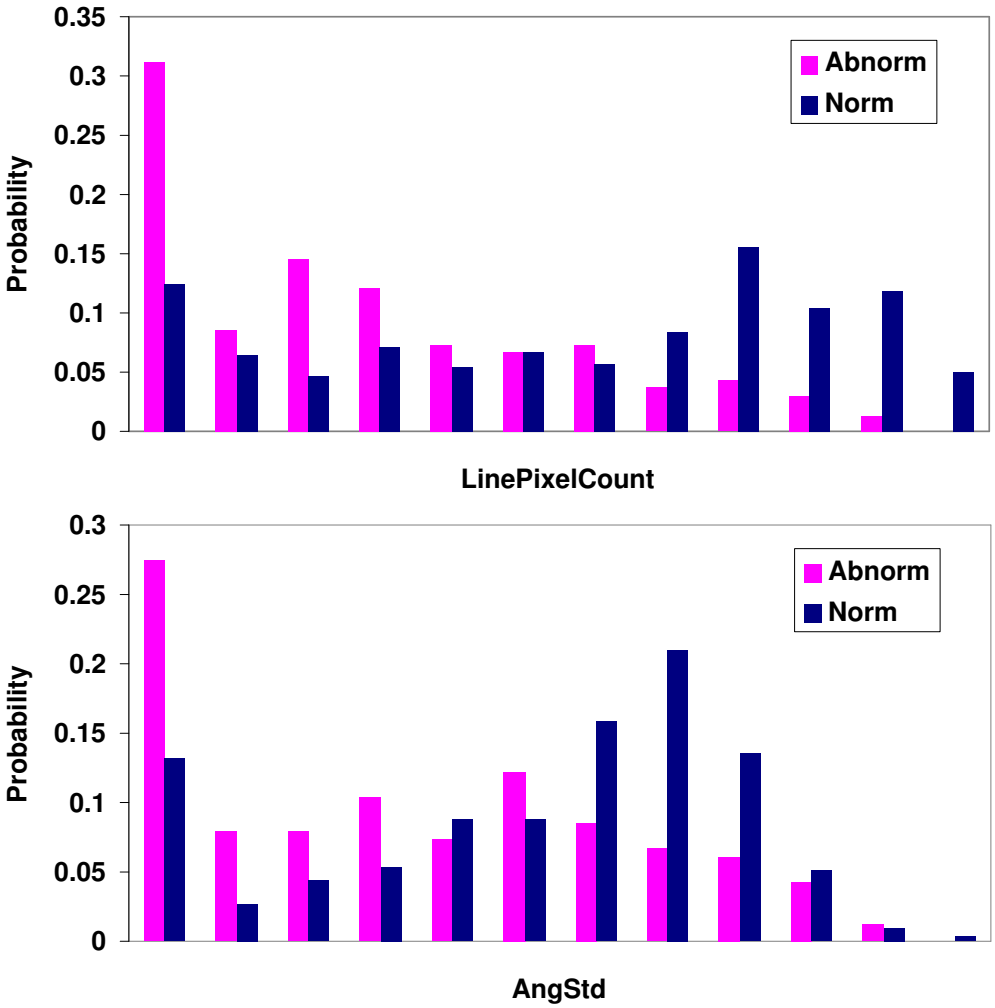


Fig. 3.6. Top: Histogram of feature *LinePixelCount* for a data set of 296 normal and 164 abnormal regions; Bottom: Histogram of feature *AngStd* for the same data.

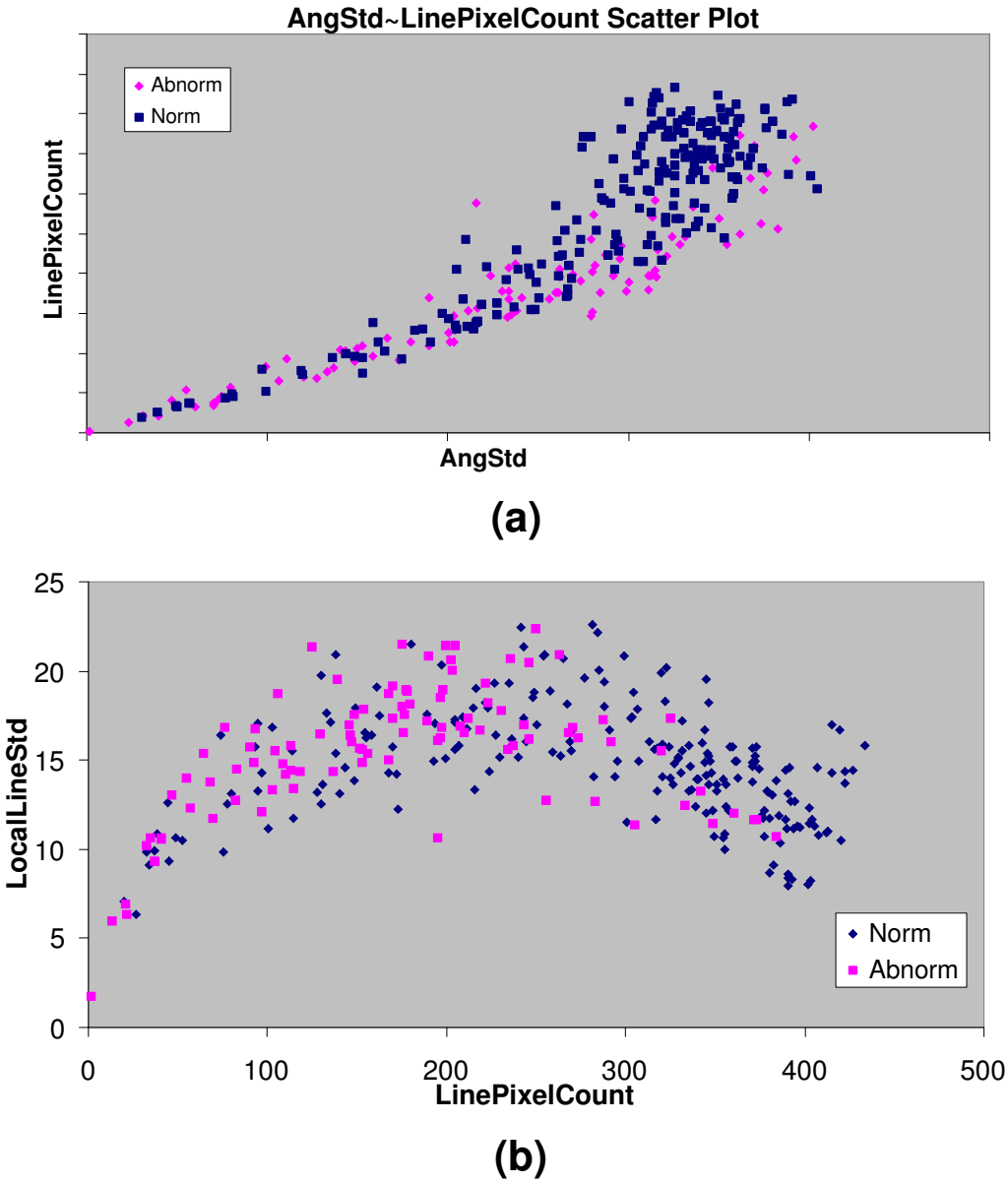


Fig. 3.7. (a) Scatter plot of feature *LinePixelCount* vs. feature *Angstd* for a data set of 296 normal and 164 abnormal regions; (b) Scatter plot of feature *LinePixelCount* vs. feature *LocalLineStd* for the same data.

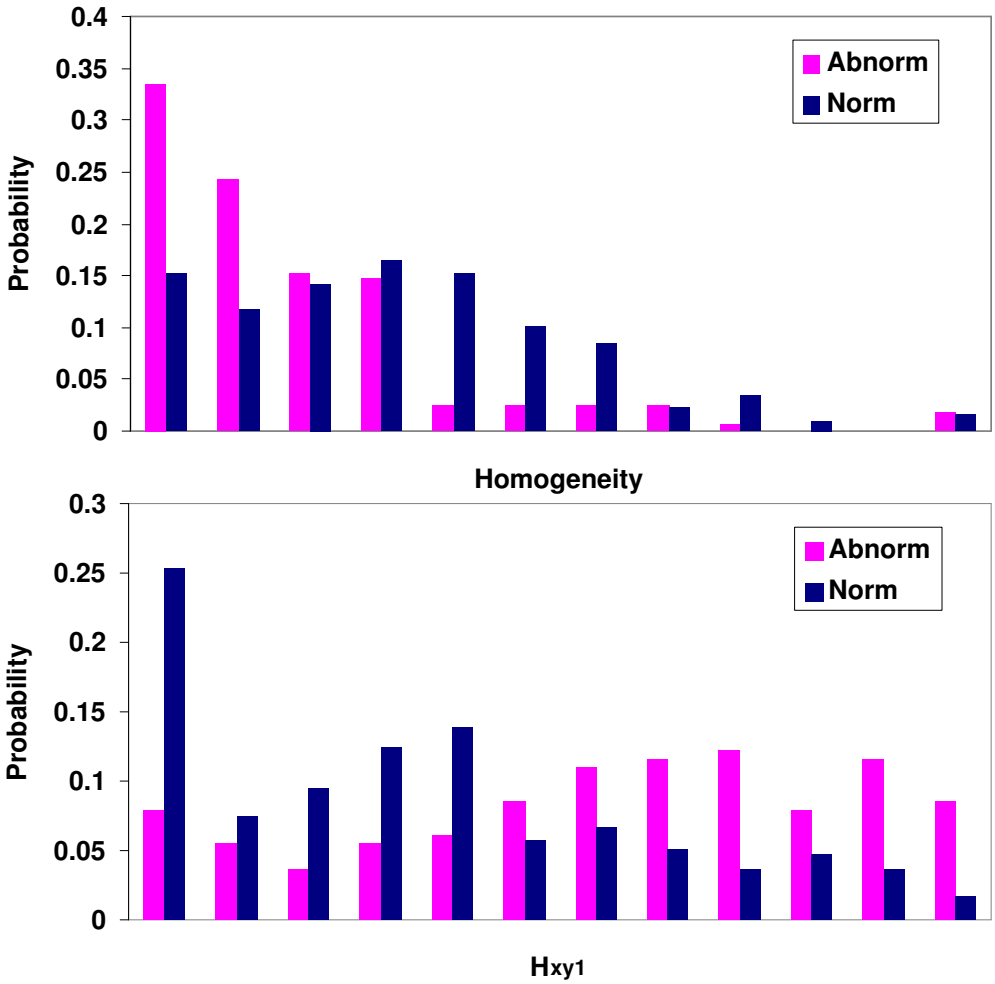


Fig. 3.8. Top: Histogram of feature *Homogeneity* for a data set of 296 normal and 164 abnormal regions; Bottom: Histogram of feature H_{xy1} for the same data.

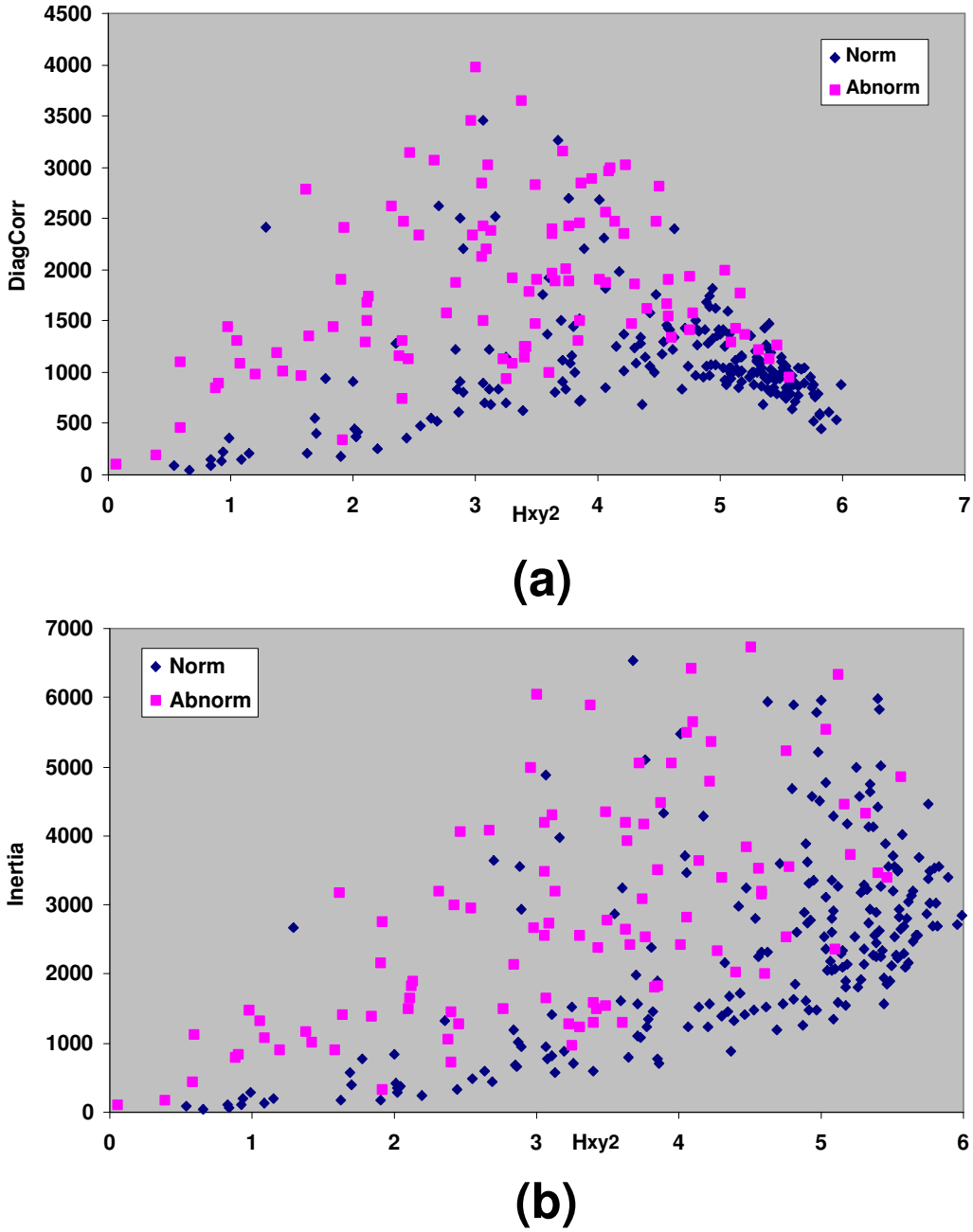


Fig. 3.9. (a) Scatter plot of feature *DiagCorr* vs. feature H_{xy2} for a data set of 296 normal and 164 abnormal regions; (b) Scatter plot of feature *Inertia* vs. feature H_{xy2} for the same data.

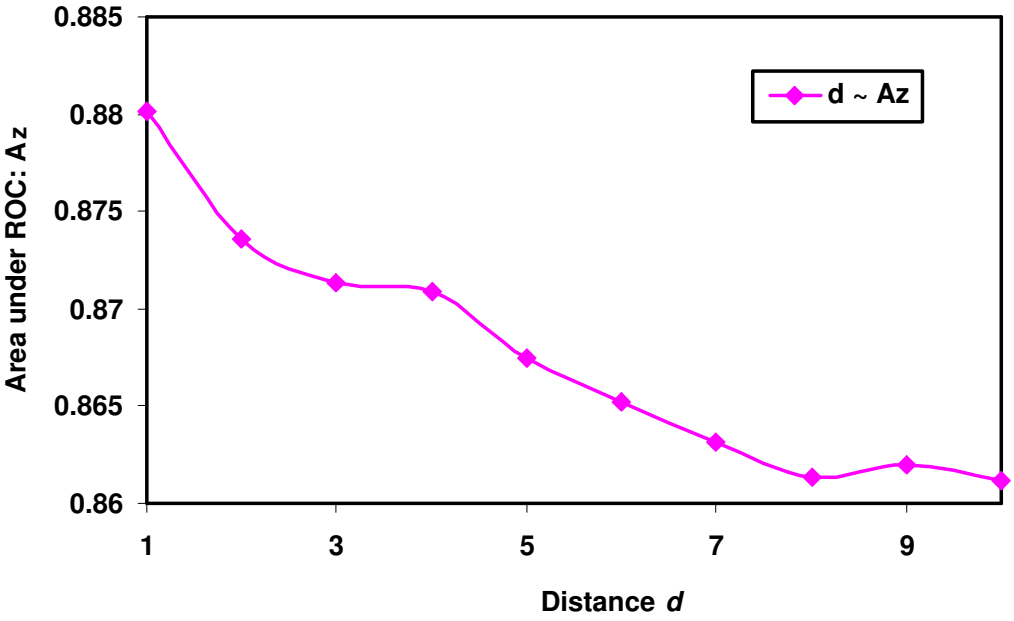


Fig. 3.10. Features extracted from the isotropic GLCM of $d = 1$ that give the best performance for our analysis

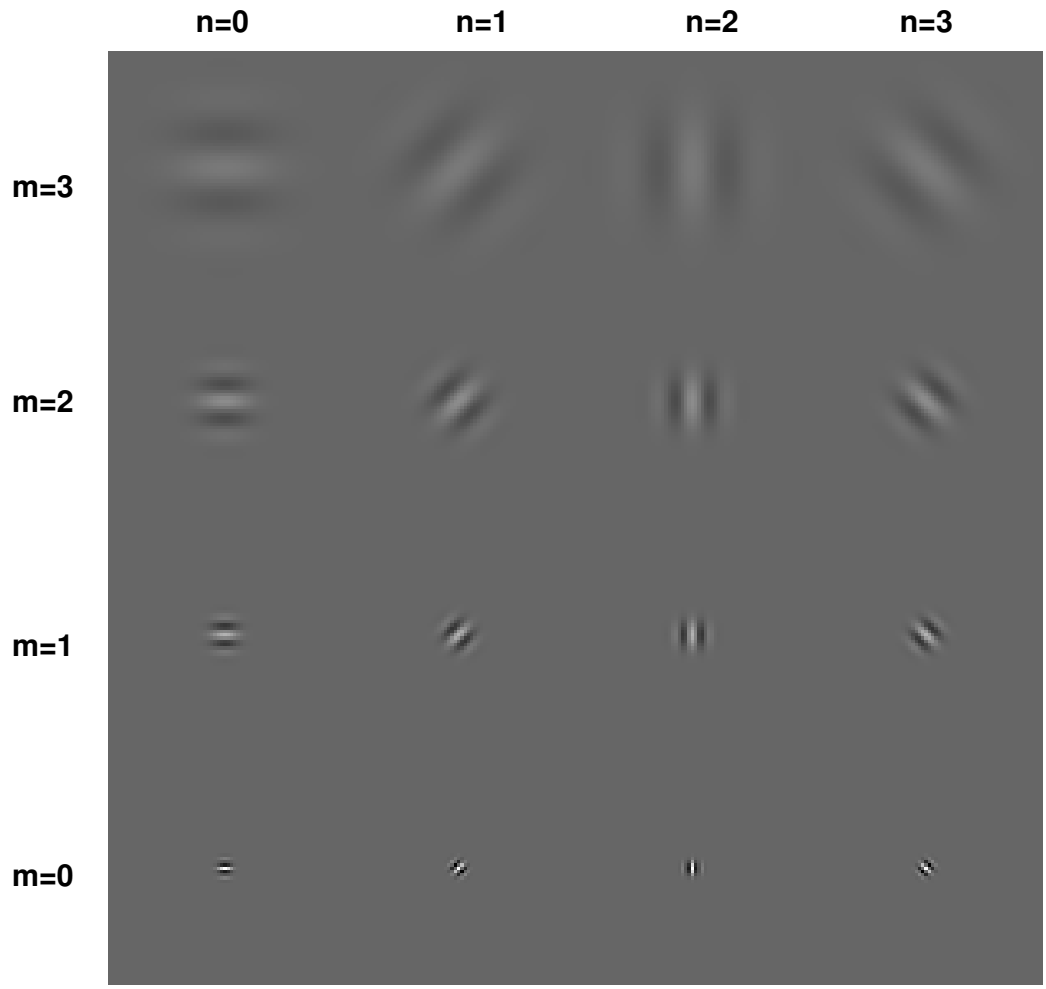


Fig. 3.11. The real parts of the Gabor filter-bank with 4 scales and 4 orientations at $a = 2$. This shows the desirable spatial locality and orientation selectivity.

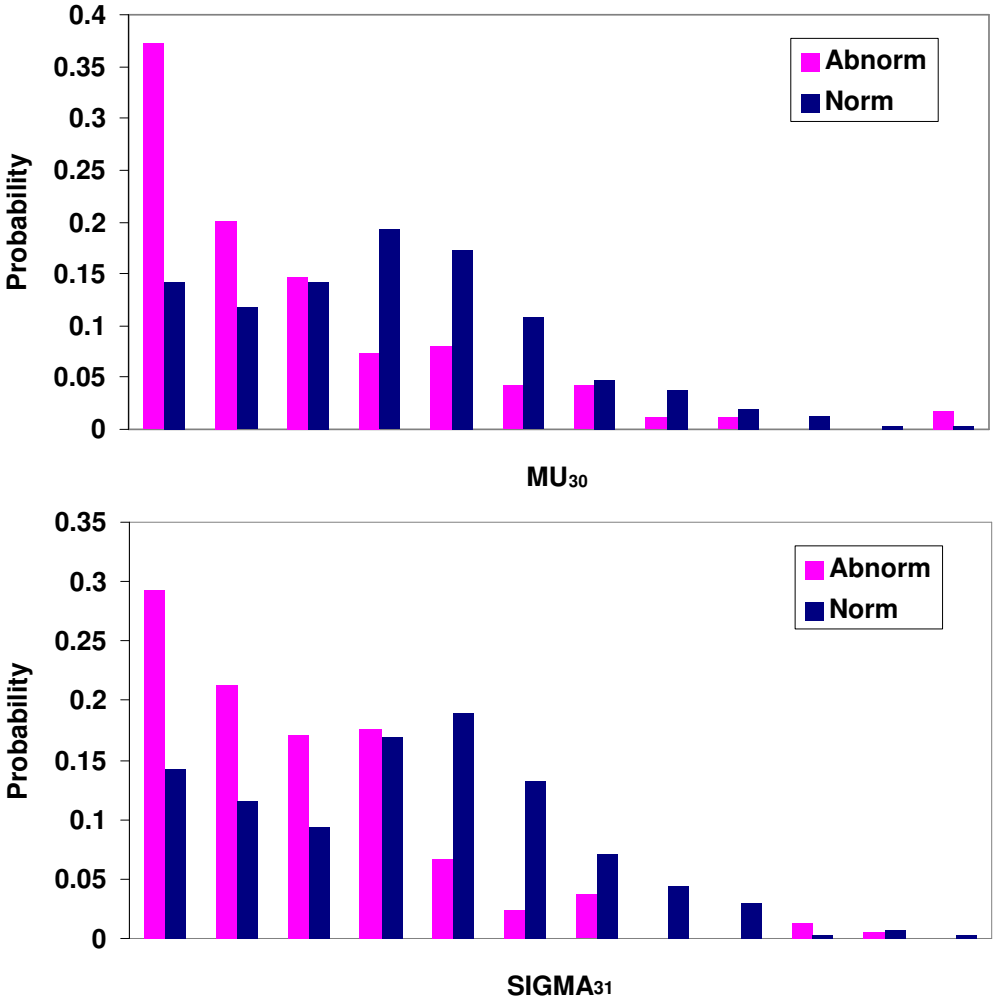


Fig. 3.12. Top: Histogram of feature μ_{30} for a data set of 296 normal and 164 abnormal regions; Bottom: Histogram of feature σ_{31} for the same data.

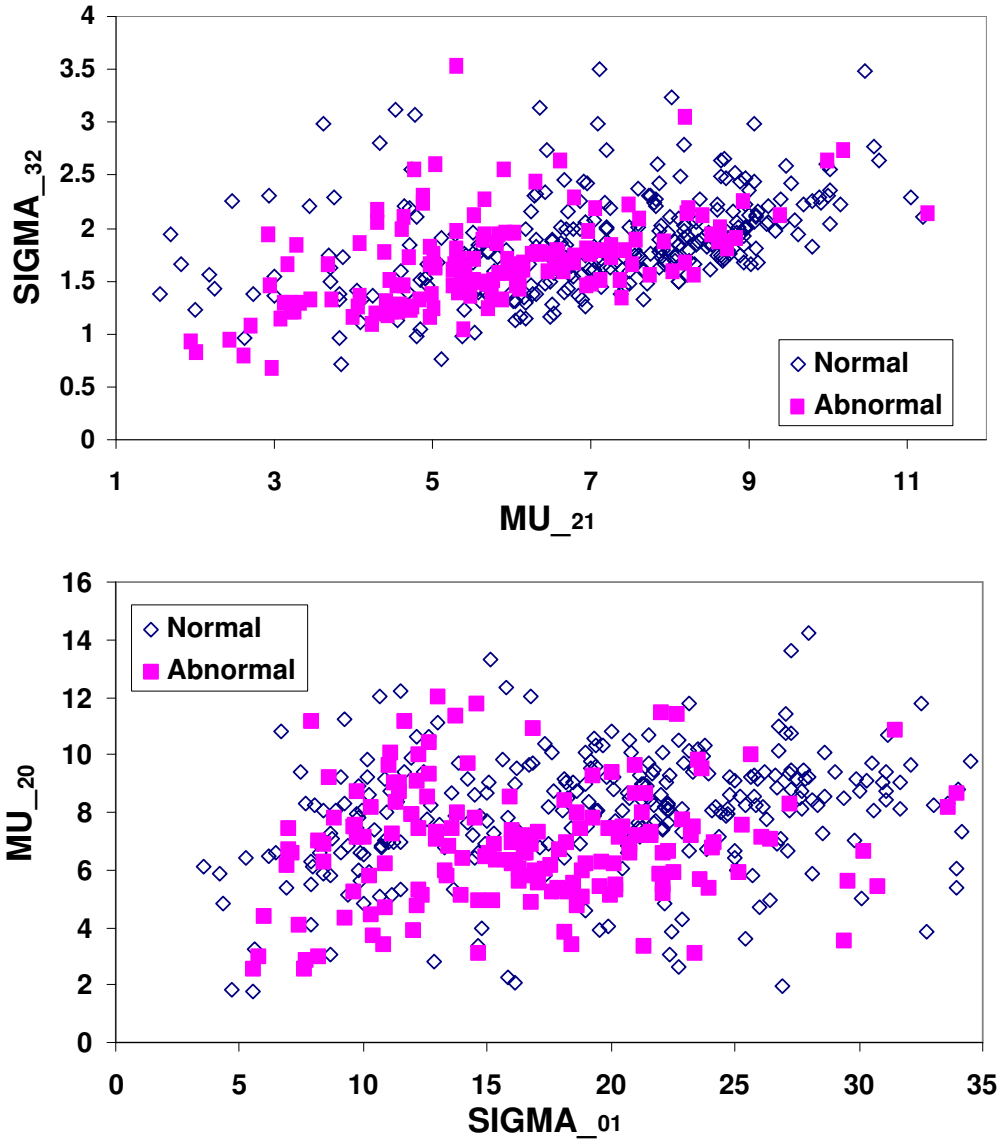


Fig. 3.13. Top: Scatter Plot of feature μ_{21} and σ_{32} for a data set of 296 normal and 164 abnormal regions; Bottom: Scatter Plot of feature σ_{01} and μ_{20} for the same data.

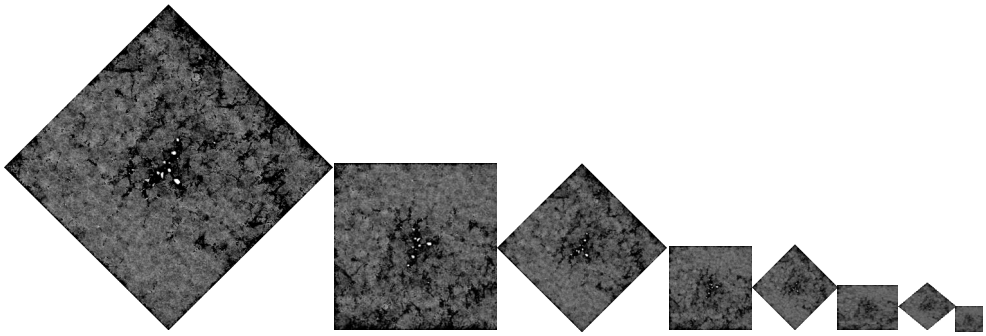


Fig. 3.14. The 2D Quincunx wavelet decomposition, only the first 4-even level (normal orientation) are used for feature extraction

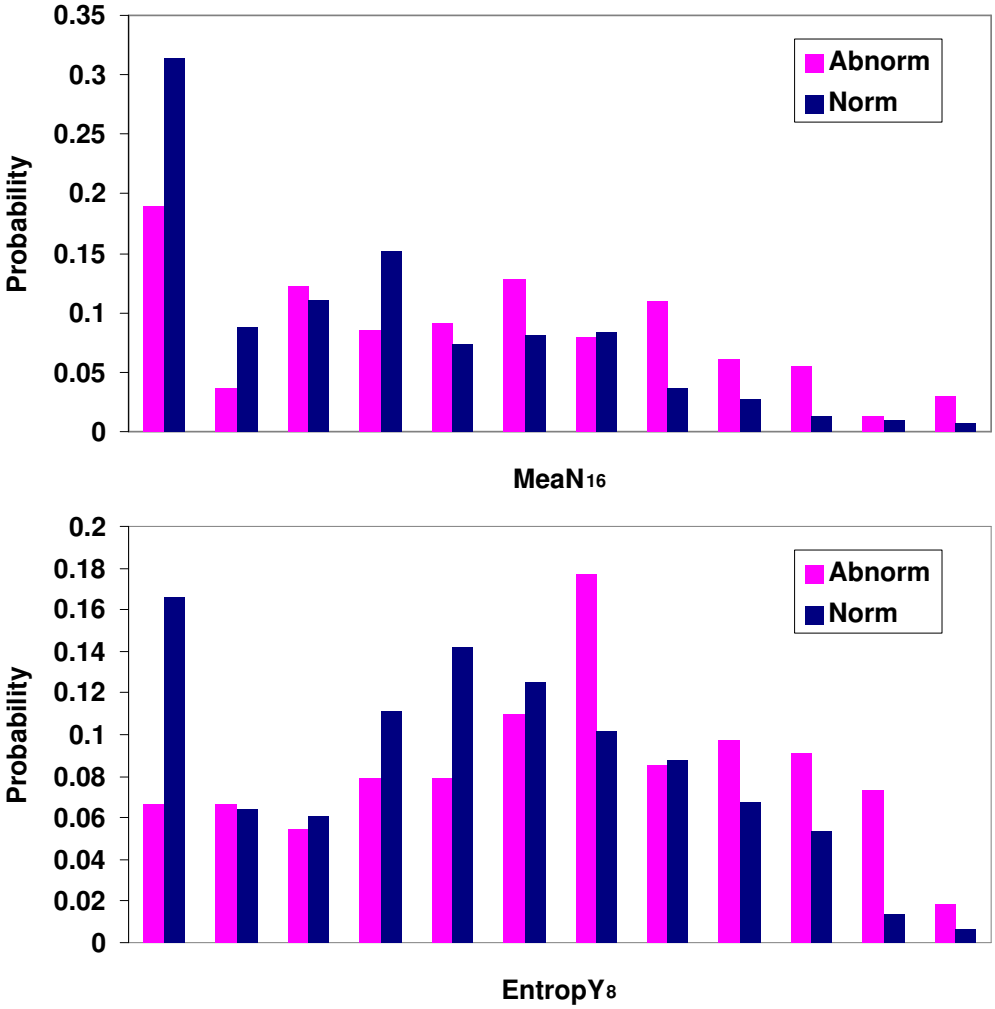


Fig. 3.15. Top: Histogram of feature $MeaN_{16}$ for a data set of 296 normal and 164 abnormal regions; Bottom: Histogram of feature $EntropY_8$ for the same data.

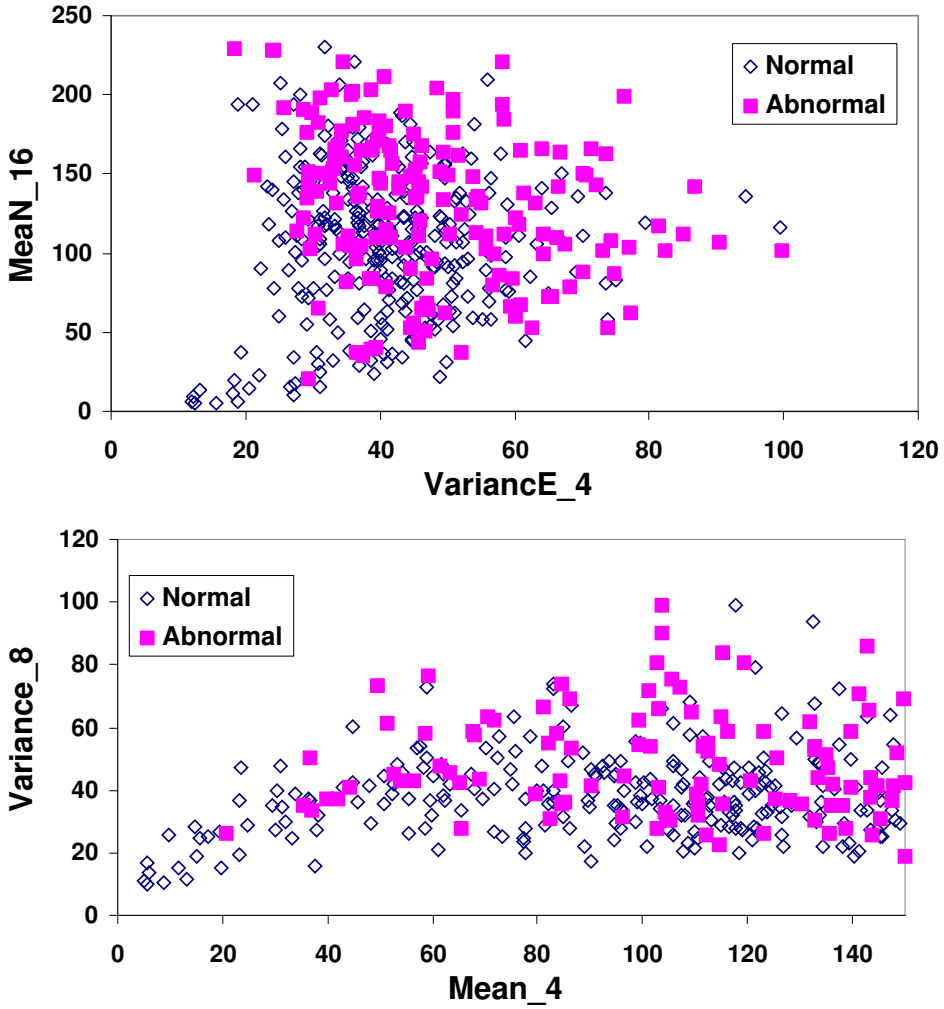


Fig. 3.16. Top: Scatter plot of feature $VariancE_4$ and $MeaN_{16}$ for a data of 296 normal and 164 abnormal regions; Bottom: Scatter plot of features $MeaN_4$ and $Variance_8$ for the same data.

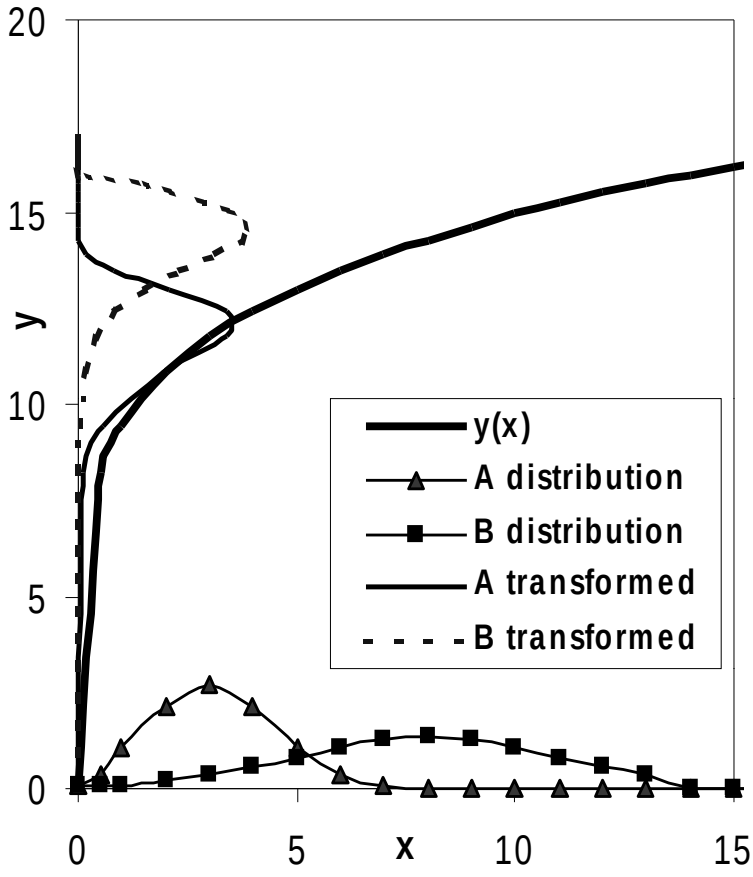


Fig. 3.17. Illustration of the power transform, where the two classes are labelled as class A and B.

4. MULTI-STAGE CASCADING CLASSIFICATION

In the previous chapter we defined features characterizing mammogram regions. Our next step will be to classify each mammogram region either as normal or as abnormal with a very high correct detection rate. Due to the heterogeneous natures of normal and abnormal classes, a general classification approach may not be adequate to capture each pattern. An efficient classification scheme has to be developed for this purpose. In this chapter, we will propose a unique multistage cascading classifier to improve the classification performance.

4.1 Classification and Classifier Combination

In general, we have a training set of feature vectors and the known *class labels* with the objective being to train the classifier such that the classifier can be used to classify new data not contained in the training data. This is known as supervised pattern classification. In some cases, training data having known class labels are not available. The goal is to unravel the underlying similarities and cluster (group) "similar" data together. This is known as unsupervised pattern classification or clustering [115, 116, 135]. Since class labels in the training set provide *a priori* information, supervised classification usually yields better performance than clustering. We will focus on supervised classification with a training data set having known class labels, i.e. the ground-truth information whether a region is normal or abnormal.

In practical supervised classification, N-dimensional feature vectors X_i's and the corresponding labels Y_i's are given as pair entries in the training database, (X_i, Y_i), $i = 1, 2, ..., M$, where M is the total number of samples, which is divided into the number of normal samples M_N and the number of abnormal samples M_S (M_S+M_N = M). The goal is to find a classification algorithm that can be trained to map X_i to

Y_i, and can be generalized for unknown test vectors ${X_i}^t$. The trained classifier g_M from a classification or mapping algorithm g' can be expressed as

$$\begin{aligned} g_M &= \min_{g'} [C_{S,N} P\{g'(X_1^N, \cdots, X_{M_1}^N) \neq (Y_1^N, \cdots, Y_{M_1}^N)\} \\ &\quad + C_{N,S} P\{g'(X_1^S, \cdots, X_{M_2}^S) \neq (Y_1^S, \cdots, Y_{M_2}^S)\}] \end{aligned} \tag{4.1}$$

where (X_i^N, Y_i^N) are normal training data, and (X_i^S, Y_i^S) are abnormal training data, and the misclassification costs $C_{N,S}$ and $C_{S,N}$ with $C_{N,S} \gg C_{S,N}$. There are many different supervised classification algorithms, from a simple linear classifier to nonlinear neural networks, such as linear discriminant analysis (LDA) [115, 116], decision trees [136–138], Bayesian networks [139–143], neural networks [144–146], support vector machines(SVMs) [147–149], and genetic algorithms [150–152].

4.1.1 Combination of Classifiers

Since training algorithms use different learning rules, which generate different class boundaries in the feature space, diverse classification results are obtained. This means that not all of the results from different training algorithms overlap and less-dominating patterns might be classified correctly in one algorithm, but might not be in another algorithm. A single classifier tends to have a high misclassification error in order to capture each pattern in the training data. The development of classifier combining systems has received increasing attention of late [153–159] in that it has been shown that such systems can be more robust and more accurate than a single classifier. A combined classifier system can be less sensitive to the tuning of internal parameters, noise and other defects in the data. The classifier combination can be a powerful solution to complex data classification problems because it allows simultaneous use of different classification algorithms to explore feature subspace. However, there is no guarantee that a combined classification system will be more robust and accurate in that it depends on the training data and the individual classifiers used.

Since the best combination of a set of classifiers depends on the application and the classifiers to be combined, there is no single best combination scheme nor any unequivocal relationship between the accuracy of a multiple classifier system and the individual constituent classifiers. Extensive studies have examined multiple classifier combination strategies with the most used method being majority voting [156]. The combination strategies generally can be categorized into two groups. One group is to find a rule to select the best classifier or a combination of classifiers, such as majority voting [156], sum or product rules [157], a statistical model [160], and the Dempster-Shafer theory of evidence [161]. The other group is to build a higher-level classifier based on the outputs of individual classifiers, such as a neural network combinator [155], Bayesian network integration [162], the stacked generalization [153], and the cascade generalization [158]. Strategies in the first group are simpler, but strategies in the second group can produce better results in complex data sets. There is no definitive distinction between the two groups. For example, a statistical model may be considered as a higher-level classifier; if the higher-level classifier is a decision tree, then it can be easily interpreted as a series of "if...then" rules. For a complex data set, instead of searching for a single classification method to fit the training data, methods combining different classification algorithms whose search spaces do not completely overlap yield better performance and less misclassification error [153, 157–159]. Since both normal and abnormal classes are complex heterogeneous mixtures, we are particularly interested in the stacked generalization [153] or the cascade generalization [158] for normal detection. We will propose and implement a unique multi-stage cascading classification system, which is in the family of stacked/cascade generalization.

4.2 Stacked and Cascade Generalization

The Stacked generalization [153, 163] is a classifier combination method. Instead of voting [156], it constructs a high-level generalizer to combine predictions from

lower-level classifiers. The Cascade Generalization [158] is a stacking classification algorithm used to merge classifiers. The cascade generalization relaxes the bias and reduces the error by using a set of classifiers sequentially. After each classification, the next classifier is trained on an extended data-set: the original data and the probability class distribution given by the current classifier. The cascade generalization [158] is described as follows:

Consider a two-class problem with training data $D = (X_i, Y_i)$ where X_i is an N-dimensional feature vector, $X_i = [X_{1,i}, X_{2,i}, \cdots, X_{N,i}]$, Y_i is a class label, $Y_i \in \{C_1, C_2\}$ where C_1 and C_2 are two class labels, and $i = 1, 2, \cdots, M$. A classifier $\Im$ is a function that maps the input feature vector to a class label, with parameters that can be adjusted for different problem settings. A specified classifier, $\Im(D)$, can be constructed from the training data D with fixed parameters. The trained classifier can be used as a predictor to assign a class label Y for the feature vector X, represented by $\Im(X, D)$, which outputs a vector of predicted class probabilities $[p_1, p_2]$ with $p_i = P(y = C_i|X)$. The constructive operator $\Phi(X, \Im(D))$ is defined as a predictive data extension operator, which concatenates the input feature vector X with the output class probabilities generated by $\Im(X, D)$. Hence, a cascade generalization is a sequential composition of classifiers, and the Φ operator extends the data at each classification step. Given the training data D, test data T, and the two classifiers $\Im_1$ and $\Im_2$, a cascade generalization is described as follows. Using Classifier $\Im_1$, generate the $Stage_1$ data

$$Stage_1 train = \Phi(D, \Lambda(\Im(D), D)) \tag{4.2}$$

$$Stage_1 test = \Phi(T, \Lambda(\Im(D), T)) \tag{4.3}$$

where the operator $\Lambda(\Im(D), H)$ is the output class probabilities generated by the classifier $\Im(D)$ on the data set H, H is the training or test data in the above equations. Classifier $\Im_2$ is trained on $Stage_1 train$ and tested on $Stage_1 test$:

$$\Lambda(\Im_2(Stage_1 train), Stage_1 test) \tag{4.4}$$

∇ is used to represent sequential combination of two classifiers, so we have:

$$\begin{aligned} \Im_2 \nabla \Im_1 &= \Lambda(\Im_2(Stage_1 train), Stage_1 test) \\ &= \Lambda(\Im_2(\Phi(D, \Lambda(\Im_1(D), D))), \Phi(T, \Lambda(\Im_1(D), T))) \end{aligned} \tag{4.5}$$

A cascade generalization of n classifiers is written as $\Im_n \nabla \Im_{n-1} \cdots \nabla \Im_1$.

With sequential classification, the data set is extended after each step. The output class probabilities from previous stages serve as "a prior" information for the last stage classifier, which reduces the bias and the training error, and the testing or generalization error.

4.3 A Unique Multi-Stage Cascading Classification System

In our classification task it is nontrivial to separate normal regions from all types of abnormal regions. First, the class of abnormal regions is a heterogeneous mixture of microcalcifications, circumscribed masses, spiculated lesions, and other abnormalities. Second, normal regions of high density pose a potential obstacle to separate from cancers. It is expected that the distributions of normal and abnormal features will not ideally behave as uni-modal distributions and some sub-patterns might be shadowed by dominating patterns. A single classifier may not be sufficient to differentiate each sub-pattern. A single classifier tends to over-fit the training data with high error when used on the complex and heterogeneous training data.

In our work, we designed a unique classifier combination method that improves the overall performance and reduces misclassification errors. The goal of our classification is to maximize the TNF, the correct classification rate of normal mammogram regions, with a very low FNF, the misclassification rate of abnormal mammogram regions as normal. Since $TNF + FPF = 1$ and $FNF + TPF = 1$ as described in Section 1.3, the above goal is the same as to minimize the FPF with a very high TPF. We follow the principle of minimizing the FPF in our classification design to achieve maximizing TNF. The reason is to make sure that we have a very high TPF, i.e. a very low FNF, since it is bad if abnormal is classified as normal while

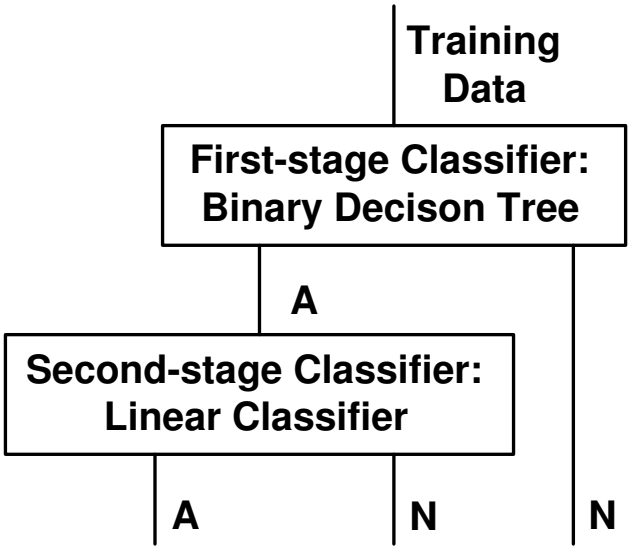


Fig. 4.1. The Structure of Our Two-stage Cascading Classifier. Class labels: A - Abnormal class and N - Normal class

performing normal detection at the screening. We use multi-stage cascading in our classification to greatly reduce the FPF, i.e. to dramatically increase the TNF. With its layered structure, it is in the family of the stacked generalization [153] or the cascade generalization [158].

4.3.1 Two-Stage Cascading Classifier

The structure of our two-stage classifier is shown in Figure 4.1, which is detailed in Figure 4.2. The emphasis of this combination scheme is that the first classifier, denoted as $\mathcal{J}_1$, should correctly classify nearly every abnormal region while separating as many normal regions as possible. $\mathcal{J}_1$ is designed to be highly cost-sensitive to make misclassification of an abnormal as "normal" very low. This is the rationale that it is unnecessary to further re-classify the data classified as "normal" by $\mathcal{J}_1$. This strategy not only reduces the complexity of the combined classification system, but also has a fundamental theoretical formulation, which most combination methods, such as the cascade generalization, lack.

Only the data classified as "abnormal" by the first classifier $\mathcal{J}_1$ are input to the second classifier, denoted as $\mathcal{J}_2$. The classifier combination scheme is simple and is easily extendable into multiple stages with cascading. Though it has a fundamentally different philosophy of combining classifiers, our two-stage cascading classification can be interpreted as a special implementation of the cascade generalization [158]. $\Im_1$ is $\mathcal{J}_1$, but $\Im_2$ is a meta-classifier as follows:

```
Define our meta-classifier ℑ2(X):  X is an input feature vector
BEGIN
    IF (X is classified by first classifier J1 as Normal )
    THEN X is Normal
    ELSE Use second classifier J2 to reclassify X
END
```

The details of our two-stage classification system are described as follows. A binary decision tree [136–138] classifier is used as the first classifier because it is one of the most powerful classification tools. Misclassification costs can be specified so that almost all of the training cancer regions are correctly classified. The decision tree algorithm [164] we used is a variation of CART [136]. This algorithm randomly divides the training data into two sets, one to build the pure leaf-node tree and the other to prune; then the roles of two sets are changed and the procedure iterates until converging to a smallest pruned subtree [164].

The decision tree classifier is based on a hierarchy of multiple features, which make it difficult to evaluate the performance using Receiver Operating Characteristic (ROC) analysis. Therefore, in addition to improving the classification performance, the second classifier can be used for ROC analysis. With the design of the first classifier, the input data to the second classifier are only the training regions classified as "abnormal" by the decision tree classifier. In our system, the second classifier is a linear discriminant classifier using Fischer's criterion [115, 116]. The linear classifier is augmented with adaptive feature selection [165, 166]. Unlike a decision tree classifier, we need to explicitly select a subset of the features. A Sequential Floating

Search [165, 166] method is one of the best feature selection methods available [167]. The principles of floating search feature selection are described as follows:

1. Let F be the selected feature subset of size k, S be the set of other features of size $N - k$, where N is the total feature number

2. Start with k=0, $F = \emptyset$, and set the stop number of $K(K <= N)$

3. Add the most significant feature f_* from S to the current subset F

4. If the added subset is the best subset of size $k + 1$, then let $k = k + 1$, $F = F + \{f_*\}$, $S = S - \{f_*\}$, and go to Step 3; otherwise, go to Step 5

5. Conditionally remove the least significant feature f_+ from the current subset F of size k

6. If removed subset is the best subset of size $k-1$ found so far, then let $k = k-1$, $F = F - \{f_+\}$, $S = S + \{f_+\}$, and go to Step 5; Else return to Step 3

7. Stop after reaching $k = K$

Implementation of the Adaptive Sequential Forward Floating Search (ASFFS) in our study is wrapper based [168], with a linear classifier as the built-in classifier. The criterion used to select the feature subset is A_z, the area under the ROC, generated by the built-in classifier. The advantage of the wrapper method is that feature selection is dependent on and consistent with the classification task.

Our two-stage cascading classifier system has the classification power of a decision tree and the simplicity of ROC analysis of a linear classifier.

4.3.2 Multi-Stage Cascading

It is easy to extend our two-stage classifier system to multi-stage. Figure 4.3 shows the extended multi-stage cascading classification system. No matter what kind of classification algorithm it uses, the first through $(n - 1)$-th stage classifier

must have the same characteristics: it should correctly classify nearly every abnormal input sample while separating as many normal input samples as possible, i.e. $TPF \simeq 1$. Each internal stage classifier must be designed to be highly cost-sensitive to make misclassification of an abnormal as "normal" very unlikely. Therefore, it is unnecessary to build another classifier at the output set of "normal." The advantage of this method is the reduction of the complexity and the reduction of the false positive rate. In each internal stage, any classification algorithm can be used, such as a decision tree, naive Bayes, or support vector machine. In order to use ROC analysis for the overall performance, the final stage classifier may be a linear or quadratic classifier for easy ROC generation. With each classifier added to the cascade, the searched feature space is reduced, which is completely opposite of the Cascade Generalization [158].

4.3.3 Normal Mammogram Region Identification

Figure 4.2 shows our two-stage classification system for identifying normal mammogram regions. A mammogram region is 512×512 pixels. The enhanced image I_E of of the region is obtained as described in Section 3.2. Four sets of features are then extracted from I_E as described in Section 3.3. From each region, a total of 86 features are obtained. These features were then used to train the two-stage cascading classifier.

The classifier was trained with 164 ground-truth cancer regions and 296 normal regions. Among the 164 ground-truth cancer regions, 53 are masses, 56 are spiculations and 55 are calcifications. Figure 4.4 shows the statistical distribution of the subtlety ratings of the 164 cancers. The subtlety rating in the DDSM [90] is used to indicate the distinguishability of a cancer by a radiologist. The lowest subtlety rating (rating 1) means the cancer is very subtle and most difficult to identify. The first-stage decision tree classifier was cost-constrained to correctly classify nearly every cancer region. This resulted in a True Positive Fraction (TPF) of 0.99 at False

Positive Fraction (FPF) of 0.29. It has a very low FNF of 0.01. Table 4.1 shows the four performance fractions of the first-decision decision tree classifier.

Table 4.1
Four Performance Fractions

TPF	FPF	TNF	FNF
0.99	0.29	0.71	0.01

The regions (including 162 true positives and 86 false positives) classified as "abnormal" were then re-classified by the second-stage linear classifier. The two-stage classifier system had the overall performance, $A_z = 0.98$, where A_z is the area under the ROC. Figure 4.5 shows the comparison with a single linear classifier, $A_z = 0.96$. Table 4.2 shows the four performance fractions of each operating point of our two-stage cascading classifier. The plot of TPF versus FPF is shown as a ROC in Figure 4.5. At the operating point shown in the fifth entry of Table 4.2, we have $TPF = 0.963$ or 96.3%, $FPF = 0.108$ or 10.8%, $TNF = 0.892$ or 89.2% and $FNF = 0.037$ or 3.7%. In terms of normal analysis, 89.2% of normal mammogram regions are correctly identified, while the misclassification rate of abnormal as normal is only 3.7%. At the operating point shown in the sixth entry of Table 4.2, we have $TPF = 0.909$ or 90.9%, $FPF = 0.064$ or 6.4%, $TNF = 0.936$ or 93.6% and $FNF = 0.091$ or 9.1%. In terms of normal analysis, 93.6% of normal mammogram regions are correctly identified, while the misclassification rate of abnormal as normal is only 9.1%. Hence, we achieved increased TNF through reducing FPF using the multi-stage classification. It clearly show that our unique multi-stage cascading classifier works.

The second-stage linear classifier uses 11 features of the 86 features selected by the adaptive floating search feature selection method described in Section 4.3.1 in comparison to the 12 features of the 86 features used by the single linear classifier. Among the selected features, at least one feature from each set: curvilinear, GLCM,

Table 4.2
Four Performance Fractions of Each Operating Point of Our Two-stage Classifier

TPF	FPF	TNF	FNF
0.988	0.291	0.709	0.012
0.982	0.267	0.733	0.018
0.982	0.216	0.784	0.018
0.976	0.172	0.828	0.024
0.963	0.108	0.892	0.037
0.909	0.064	0.936	0.091
0.860	0.020	0.980	0.140
0.756	0.010	0.990	0.244
0.530	0.003	0.997	0.470
0.299	0.003	0.997	0.701
0.104	0.000	1.00	0.896

Gabor or multiresolution feature sets were chosen, which indicates that each set is complementary to each other.

4.3.4 Performance Analysis: Why does it work?

Multistage classification provides a good solution for complex data sets. The combination method is simple and easy to understand. Receiver Operating Characteristic (ROC) curve is used for the performance analysis.

Use of Another Data Set

We have also tested our two-stage classification system on the *image* data-set from the UCI Repository of machine learning databases [169]. This image segmentation

data-set consists of 7 classes (brickface, sky, foliage, cement, window, path, grass). Each class has a total of 330 instances: 30 training and 300 test data. Each instance has 19 continuous features. In our experiment, the "cancer" class is foliage, the "normal" class consists of grass, sky, and windows to simulate the heterogenous mixture. We achieved $A_z = 0.99$ using our two-stage cascading classifier, comparing with $A_z = 0.90$ of a single linear classifier. Figure 4.6 shows ROC curve of our two-stage classifier on the *image* data-set, with comparison to a single linear classifier.

Performance Formulation

Our focus is normal mammogram analysis. We want to maximize the TNF, the rate of correctly classifying normal as normal, with very low FNF, the misclassification rate of abnormal as normal. This is the same to minimize the FPF with very high TPF since $TNF = 1 - FPF$ and $FNF = 1 - TPF$. It is clinically critical if an abnormal fails to be uncovered at the screening as described in Section 1.2.

In order to be consistent with the performance definitions of other CAD methods, the performance of our cascading classifier system is defined conventionally:

$$TPF = \frac{\#\text{ of abnormal regions correctly classified as abnormal}}{\#\text{ of total abnormal regions}} \tag{4.6}$$

$$FPF = \frac{\#\text{ of normal regions classified as abnormal}}{\#\text{ of total normal regions}} \tag{4.7}$$

$$TNF = \frac{\#\text{ of normal regions correctly classified as normal}}{\#\text{ of total normal regions}} \tag{4.8}$$

$$FNF = \frac{\#\text{ of abnormal regions classified as normal}}{\#\text{ of total abnormal regions}} \tag{4.9}$$

in the above equations denotes the number or the count. The performance gain of our two-stage classifier can be explicitly formulated, which would be very difficult in the Cascade Generalization. We will show that our two-stage cascading classifier system has a much lower FPF than that of a single classifier while having approximately the same TPF. Therefore, the classifier has a much higher TNF than that of a single classifier while keeping a very low FNF. We improve the detection rate of normal mammogram through our unique two-stage classification. Denoting the

TPF and FPF of the first classifier as TPF_1 and FPF_1, and TPF_2 and FPF_2 of the second classifier, the overall TPF and FPF of our two-stage classifier will be

$$\begin{aligned} TPF_{\nabla} &= TPF_1 \times TPF_2 \simeq TPF_2 \\ FPF_{\nabla} &= FPF_1 \times FPF_2 < FPF_2 \end{aligned}$$

since the first classifier is designed with TPF_1 nearly one. We then have

$$\begin{aligned} TNF_{\nabla} &= 1 - FPF_{\nabla} \\ FNF_{\nabla} &= 1 - TPF_{\nabla} \end{aligned}$$

This two-stage classifier is expected to reduce the overall FPF; therefore it is expected to increase the overall TNF, the correct classification rate of normal mammogram regions as normal. By cascading we maintain a very high TPF. Hence, we have a very low FNF, the misclassification rate of abnormal as normal. Therefore, our two-stage classification improves normal mammogram detection when compared to a single classifier while keeping the clinically critical misclassification of abnormal as normal very low.

For the extended multi-stage classifier system, the first through (n-1)-th stage classifier is designed with $TPF_i \simeq 1, i = 1, 2, \cdots, n-1$. Even though each classifier may use different algorithms, we can assume that FPF is roughly similar, i.e. $FPF_1 \simeq FPF_2 \simeq \cdots \simeq FPF_n = FPF_* < 1$. Hence, we can evaluate the overall performance of a multi-stage cascading classification system:

$$\begin{aligned} TPF_{\nabla} &= TPF_1 \times TPF_2 \times \cdots \times TPF_n \simeq TPF_n \\ FPF_{\nabla} &= FPF_1 \times FPF_2 \times \cdots \times FPF_n \simeq {FPF_*}^n \rightarrow 0 \end{aligned}$$

Since $1 - FPF_{\nabla} = TNF_{\nabla} \rightarrow 1$ and $1 - TPF_{\nabla} = FNF_{\nabla}$, we can have a near perfect normal mammogram classification while keeping the misclassification rate of abnormal as normal very low. For example, if the average performance of a classifier is $TPF_* = 0.99$ at $FPF_* = 0.20$, then we could greatly reduce the overall FPF after 5 stages of cascading.

$$TPF_{\nabla} = (TPF_*)^5 = 0.99^5 = 0.951$$

$$
\begin{aligned}
FPF_{\nabla} &= (FPF_{*})^{5} = 0.20^{5} = 0.00032 \\
TNF_{\nabla} &= 1 - FPF_{\nabla} = 1 - 0.00032 = 0.99968 \text{ or } 99.968\% \\
FNF_{\nabla} &= 1 - TPF_{\nabla} = 1 - 0.951 = 0.049 \text{ or } 4.9\%
\end{aligned}
$$

Theoretically, we can approach near perfect normal mammogram identification with not much increase of False Negative Fraction (FNF) if many classifiers are cascaded. However, increasing the number of stages, not only increases the complexity, the computation time and the vulnerability to noise and training data defects, but also may reduce the performance since the TPF of the internal stage classifier is not exactly 1. Is little performance gain worth the cost of computational complexity and the increasing vulnerability of noise effects? The general preference is probably best for a smaller number of stages.

Performance versus the number of stages

The optimal number of stages can be influenced by the complexity of the training data set, the type of classifier in each stage, and even the order of classifiers being cascaded.

However, the theoretical formulation of the overall performance in our cascading design can allow us to estimate the number of stages assuming the training data is perfect. The objective of multi-stage cascading is to maximize A_z. In the following, we will estimate the smallest number of stages to maximize A_z. The assumption for the estimation is that each individual classifier has a similar performance.

First, we need to estimate A_z from a typical operating point of $y = TPF_{*}$ at $x = FPF_{*}$. As Figure 4.7 shows, A_z could be estimated as:

$$A_z = \frac{1 - x + y}{2} = \frac{1 - FPF_{*} + TPF_{*}}{2} \tag{4.10}$$

The estimated A_z of n stage classifiers with an average performance of $y = TPF_{*}$ at $x = FPF_{*}$ will be:

$$A_z = \frac{1 - x^{n} + y^{n}}{2} = \frac{1 - {FPF_{*}}^{n} + {TPF_{*}}^{n}}{2} \tag{4.11}$$

To find the optimal number of stages to maximize A_z, we take the derivative of the above equation:

$$\frac{dA_z}{dn} = \frac{1}{2}(-x^n \ln x + y^n \ln y) = 0 \tag{4.12}$$

So, we have the desired number of stages by taking natural logarithms and solving for n:

$$n_* = \frac{\ln\left[\frac{\ln(x)}{\ln(y)}\right]}{\ln\left(\frac{y}{x}\right)} = \frac{\ln\left[\frac{\ln(FPF_*)}{\ln(TPF_*)}\right]}{\ln\left(\frac{TPF_*}{FPF_*}\right)},\ TPF_* \neq FPF_* \tag{4.13}$$

i.e. the number of stages depends on the average performance of each classifier. Since the performance of the classifier will be unknown before the training, the determination of the optimal number of stages would be difficult in practice. If $TPF_* = FPF_*$, i.e. random guessing, then $n_* = 1$ is the smallest number. Hence, cascading will not have any performance gain. However, even for a weak classifier with performance slightly better than random guessing ($TPF_* > FPF_*$), multi-stage cascading can improve the overall performance in theory.

Now we examine how a multi-stage classification system trained on regions can improve full-field mammogram analysis. In order to obtain a rough estimate of the number of stages there are some assumptions that need to be made. First, assume the individual classifier at each stage has similar performance of $y = TPF_*$ at $x = FPF_*$. Second, each breast area is analyzed by disjoint regions of the same size as the training region. Third, suppose cancer appears in one region in the breast area. Let n be the number of stages, and N_R be the number of normal regions in a mammogram, then the performance on full-field mammograms will be

$$\widehat{TPF} = y^n = {TPF_*}^n \tag{4.14}$$

$$\begin{aligned} \widehat{FPF} &= 1 - (\text{regional } TNF)^{N_R} = 1 - (\text{regional } FPF)^{N_R} \\ &= 1 - (1 - x^n)^{N_R} = 1 - (1 - {TPF_*}^n)^{N_R} \end{aligned} \tag{4.15}$$

Therefore, the estimated A_z for full-field analysis will be

$$\widehat{A_z} = \frac{1 - \widehat{FPF} + \widehat{TPF}}{2} = \frac{1 - (1 - (1 - x^n)^{N_R}) + y^n}{2} = \frac{(1 - x^n)^{N_R} + y^n}{2} \tag{4.16}$$

Note that if $N_R = 1$, this is the same as equation 4.11. For example, the typical square that encloses the breast area is 4000 pixels by 2000 pixels. Considering the breast area shape, the actual number of normal breast regions (512×512) is about 20, i.e. $N_R = 20$.

If the average performance of an individual classifier is $y = TPF_* = 0.99$ at $x = FPF_* = 0.2$, then one-stage full-field analysis performance is $\widehat{A_z} = 0.5$ for $N_R = 20$. For a 5-stage cascading classification system ($n = 5$), we have $\widehat{A_z} = 0.97$. Theoretically, we can dramatically improve full-field mammogram analysis with multi-stage classification.

Now, to find the optimal number of stages, n, for full-field mammogram classification, we have

$$\begin{aligned} 0 = \frac{d\widehat{A_z}}{dn} &= \frac{1}{2}\left\{N_R(1-x^n)^{N_R-1}(-x^n\ln(x)) + y^n\ln(y)\right\} \\ \Rightarrow \quad & y^n\ln(y) = N_R(1-x^n)^{N_R-1}x^n\ln(x) \end{aligned} \tag{4.17}$$

Taking natural logarithms,

$$\begin{aligned} n\ln(y) + \ln(\ln(y)) &= \ln(N_R) + (N_R-1)\ln(1-X^n) + n\ln(x) + \ln(\ln(x)) \\ \Rightarrow n\ln(\frac{y}{x}) &= (N_R-1)\ln(1-x^n) + \ln\left[N_R\frac{\ln(x)}{\ln(y)}\right] \end{aligned} \tag{4.18}$$

Since $x \ll 1$, $\ln(1-x^n) \approx 0$ for $n > 1$. Therefore, the term $(N_R-1)\ln(1-x^n)$ in the above equation can be ignored. We have

$$n_* \approx \frac{\ln\left[N_R\frac{\ln(x)}{\ln(y)}\right]}{\ln(\frac{y}{x})},\ x \neq y \tag{4.19}$$

Note that if $N_R = 1$, this is consistent with equation 4.13. Suppose the average performance of an individual classifier is $y = TPF_* = 0.99$ at $x = FPF_* = 0.2$. For the region analysis, $N_R = 1$, we have $n = 3.17$; for the full-field analysis, $N_R = 20$, we have $n = 5.05$. Thus, the number of stages in the classification system depends on the average size of the full-field mammogram, N_R, as well as the "intrinsic separability" in the feature space, characterized by the average $x = FPF_*$ and $y = TPF_*$.

Extend this to two-view (MLO and CC views) mammograms and to a case of 4 views of both left and right breasts. Let N_R be the number of normal regions in one

mammogram, $x = FPF_*$ be the average FPF of an individual classifier, $y = TPF_*$ be the average TPF of an individual classifier, $\widehat{FPF}$ be the FPF for two-view full-field mammograms, and $\widehat{TPF}$ be the TPF for two-view full-field mammograms. We have

$$\begin{aligned} \widehat{FPF} &= 1-(1-x^n)^{2N_R} && (4.20)\\ \widehat{TPF} &= 1-(1-y^n)^2 = y^n(2-y^n) && (4.21) \end{aligned}$$

Hence, for a two-view analysis, we have

$$\widehat{A_z} \approx \frac{1-\widehat{FPF}+\widehat{TPF}}{2} = \frac{(1-x^n)^{2N_R}+y^n(2-y^n)}{2} \tag{4.22}$$

Now for a case analysis of four views, there are two breasts but only one breast presented with a cancer in most cases. Therefore, TPF for a case will be the same as the two-view analysis, but FPF will be changed since there are $2 \times 2N_R$ normal regions now. So for a case of four views, we have

$$\widehat{A_z} = \frac{(1-x^n)^{4N_R}+y^n(2-y^n)}{2} \tag{4.23}$$

For $y \approx 1$, we first simplify $y^n(2-y^n)$ as $y^n(1+\delta)$, where $\delta \ll 1$ and independent of n. We then take the derivative with respect to n on both sides,

$$\begin{aligned} 0 = \frac{d\widehat{A_z}}{dn} &= \frac{1}{2}\left\{4N_R(1-x^n)^{4N_R-1}(-x^n\ln(x))+y^n\ln(y)(1+\delta)\right\} \\ \Rightarrow \quad & y^n\ln(y)(1+\delta) = 4N_R(1-x^n)^{4N_R-1}x^n\ln(x) && (4.24) \end{aligned}$$

Take natural logarithms,

$$\begin{aligned} n\ln(y)+\ln(\ln(y))+\ln(1+\delta) &= \ln(4N_R)+(4N_R-1)\ln(1-X^n)+n\ln(x)+\ln(\ln(x)) \\ \Rightarrow n\ln(\frac{y}{x}) &= \ln\left[4N_R\frac{\ln(x)}{\ln(y)}\right] && (4.25) \end{aligned}$$

Since $x \ll 1$, $\ln(1-x^n) \approx 0$ for $n > 1$, term $(4N_R-1)\ln(1-x^n)$ can be ignored. We have

$$n_* \approx \frac{\ln\left[4N_R\frac{\ln(x)}{\ln(y)}\right]}{\ln(\frac{y}{x})},\ x \neq y \tag{4.26}$$

Equation 4.13, 4.19 and 4.26 are consistent, and can be expressed as the same equation. Let N_{Total} be the total number of normal regions in a testing example to be analyzed, we can estimate the optimal number of stages in the classification system as

$$n_* \approx \frac{\ln\left[N_{Total}\frac{\ln(x)}{\ln(y)}\right]}{\ln(\frac{y}{x})} = \frac{\ln\left[N_{Total}\frac{\ln(FPF_*)}{\ln(TPF_*)}\right]}{\ln(\frac{TPF_*}{FPF_*})},\ TPF_* \neq FPF_* \tag{4.27}$$

For example, suppose an individual classifier in each stage has average performance of $y = TPF_* = 0.99$ and $x = FPF_* = 0.2$, then we have Table 4.3.

Table 4.3
The estimated optimal number of stages

	N_{Total}	n_*
Region Analysis	1	3.17
Full-field Mammogram	20	5.05
Two-view Analysis	40	5.48
Case (4-view) Analysis	80	5.91

From Table 4.3, a five-stage ideal cascading classifier may be optimal for full-field analysis. However, one of the assumptions is that the training data is infinitely dense in the feature space, and therefore an individual classifier at stage n is unbiased as the first stage classifier. However, in practice, the training data is sequentially reduced, and the bias increases dramatically in the later stages for a finite training data. For limited training data, a smaller number of stages may be preferred. In the present work, we used a two-stage classification system for our normal mammogram analysis.

Properties of Our Multi-Stage Classification System

There are some advantages of our cascading classification method compared with other classifier combination strategies.

- It is simple and easy to understand and implement. The cascading structure is very clear as shown in Figures 4.1 and 4.3.

- The later stage is simpler than the previous stage, and the later stage explores reduced search space in the feature domain. This reduces the complexity, comparing with the data extension in the cascade generalization.

- The overall performance can be theoretically formulated. We clearly showed that multistage cascading greatly increases the true negative fraction (TNF) through reducing the false positive fraction (FPF). Most of the other combination strategies lack a theoretical and formulated demonstration of performance gain.

- The overall performance can be easily obtained with ROC analysis.

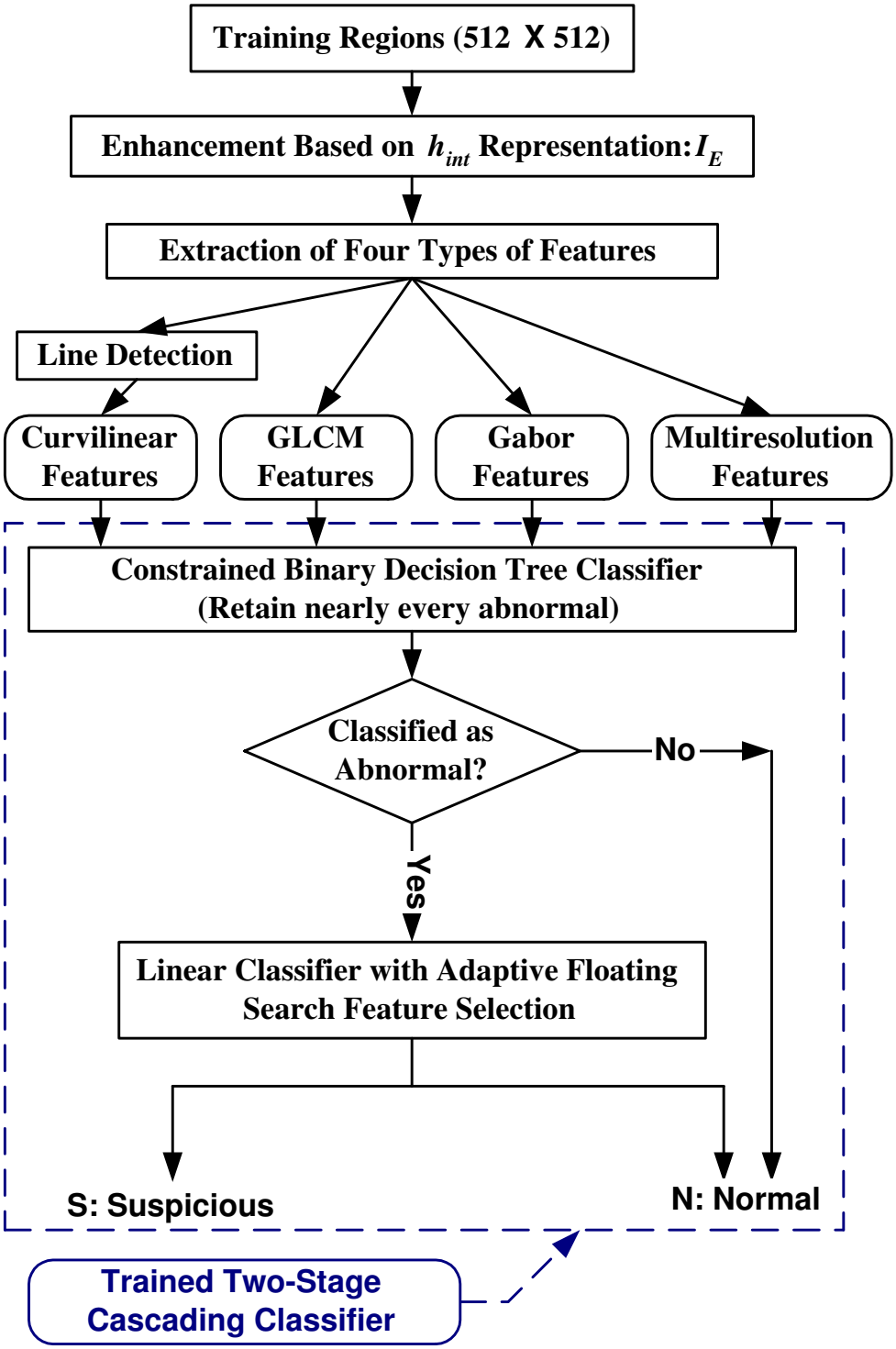


Fig. 4.2. The two-stage classification system for identifying normal regions

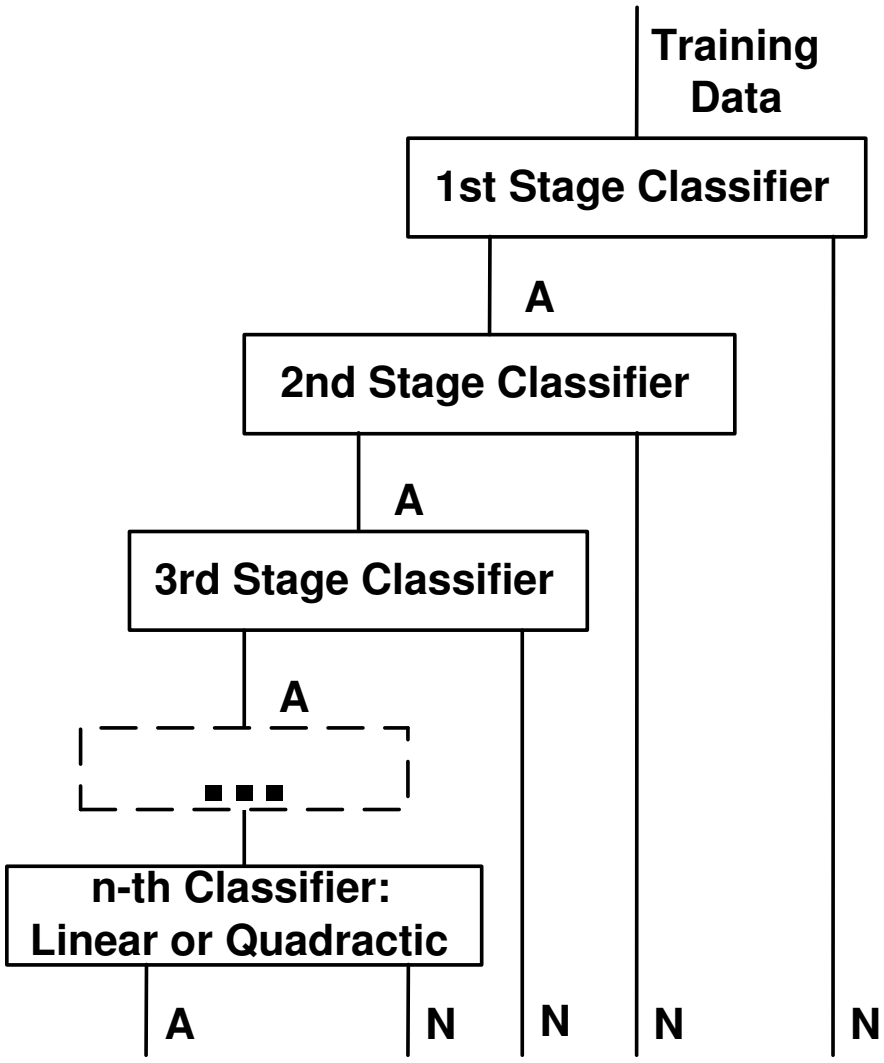


Fig. 4.3. Illustration of Extended Multi-stage Cascading Classification System. Class labels: A - Abnormal class and N - Normal class

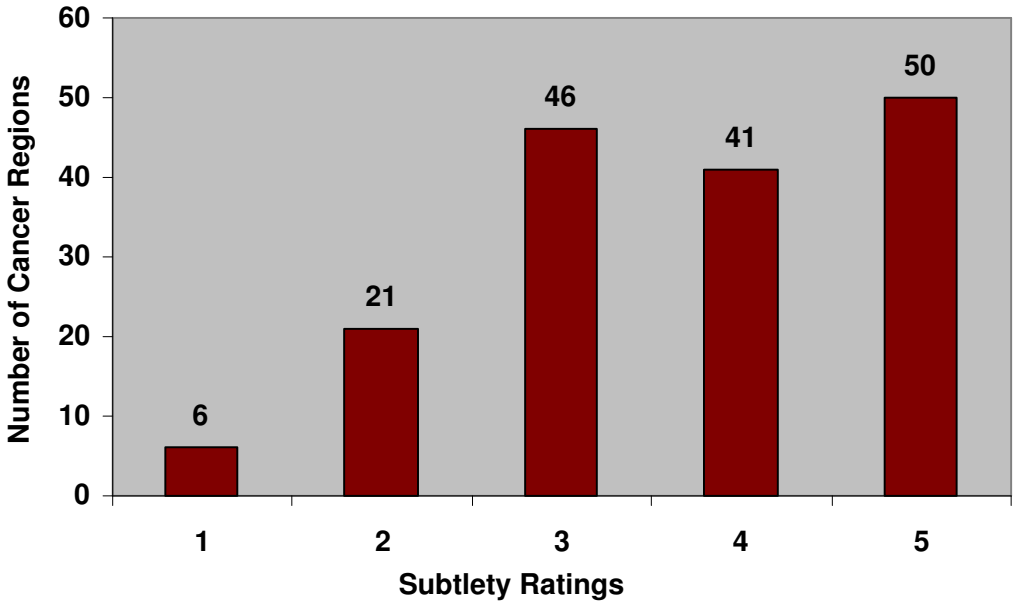


Fig. 4.4. The distribution of Subtlety Ratings of 164 Cancer Regions

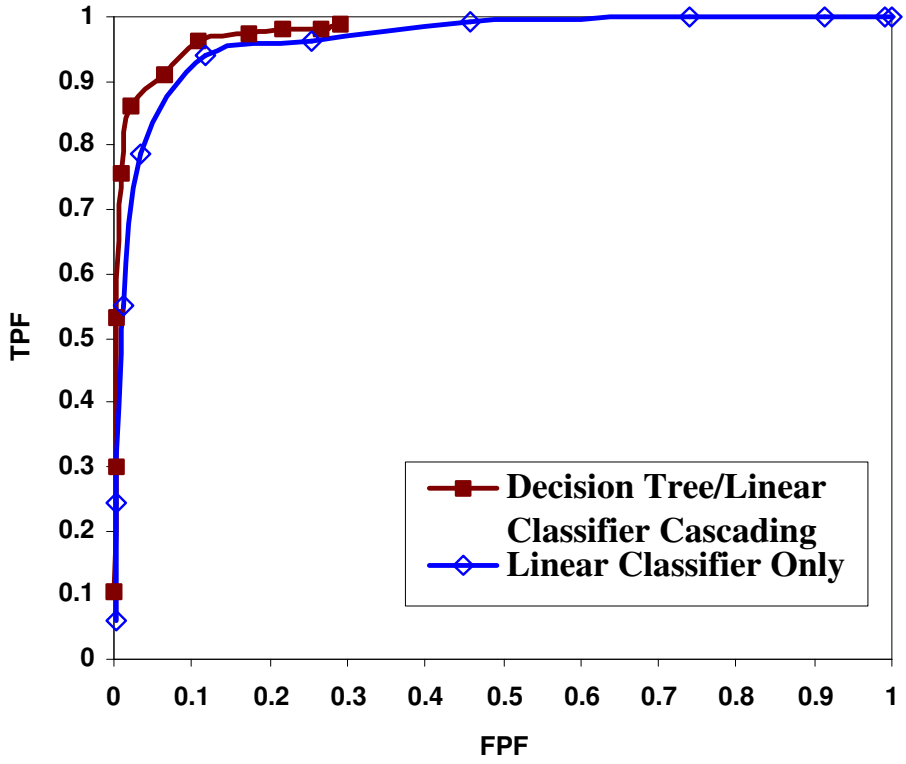


Fig. 4.5. Comparing the overall performance of our two-stage classifier with a single linear classifier on identifying normal mammogram regions

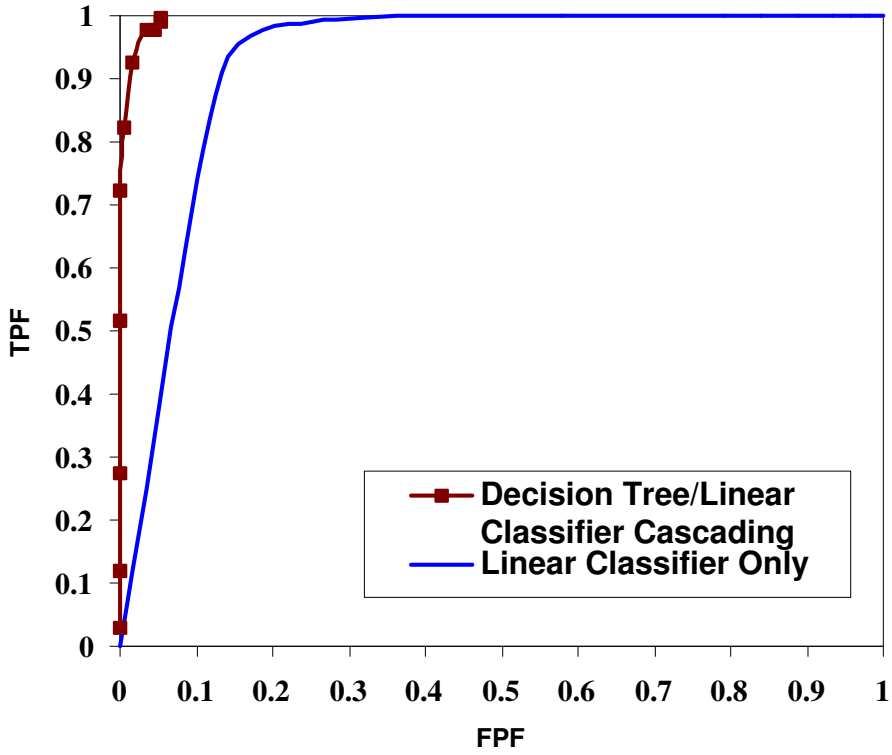


Fig. 4.6. Comparing the performance of our two-stage classifier with a single linear classifier on the *image* dataset

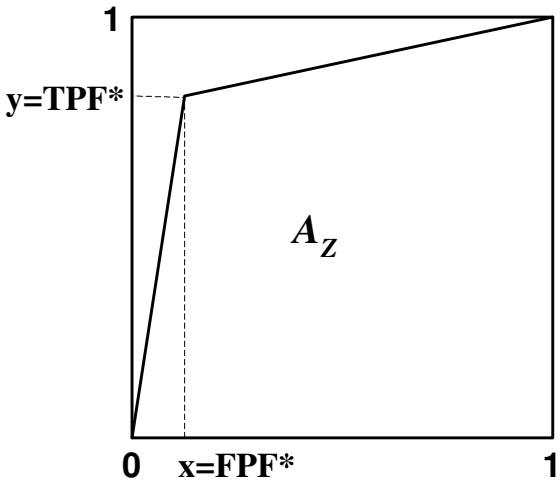


Fig. 4.7. Estimation of A_z from an average TPF^* and FPF^*

5. FULL-FIELD MAMMOGRAM ANALYSIS

In this chapter, we will present the results of full-field normal mammogram analysis using the two-stage classifier described in Section 4.3.4. The classifier will be used to analyze full-field mammograms using an overlapped moving block technique. Figure 5.1 shows the block diagram of our full-field mammogram analysis.

5.1 Breast Region and Background Segmentation

For full-field mammogram analysis, the breast region has to be segmented from the background. Otherwise, background noise and artifacts will be spuriously classified either as normal or abnormal regions. Breast-background segmentation also reduces the processing area significantly. Several segmentation approaches have been proposed [170–177]. We modified the segmentation algorithm described in [170]. The segmentation algorithm is based on histogram thresholding, morphological filtering and boundary shaping.

Figure 5.2 illustrates this method. A full-field mammogram is initially segmented by a threshold determined from its histogram. After determining the breast region peak P_{br} and background peak P_{bg} in the histogram, the threshold t_0 is selected by maximizing a local discontinuity measure. Figure 5.3 shows the automatic threshold t_0 selection, where P_{bg} indicates the mode of background area, and P_{br} indicates the breast area. The histograms shown in the figure are very typical of mammograms obtained from the DDSM [90]. The segmented image is filtered by closing and opening and then artifacts outside of the breast and the top/bottom unexposed rows are removed to obtain the final segmentation. Figure 5.4 shows the breast-background segmentation. This algorithm is relatively robust and gives satisfactory segmentation for mammograms from the DDSM.

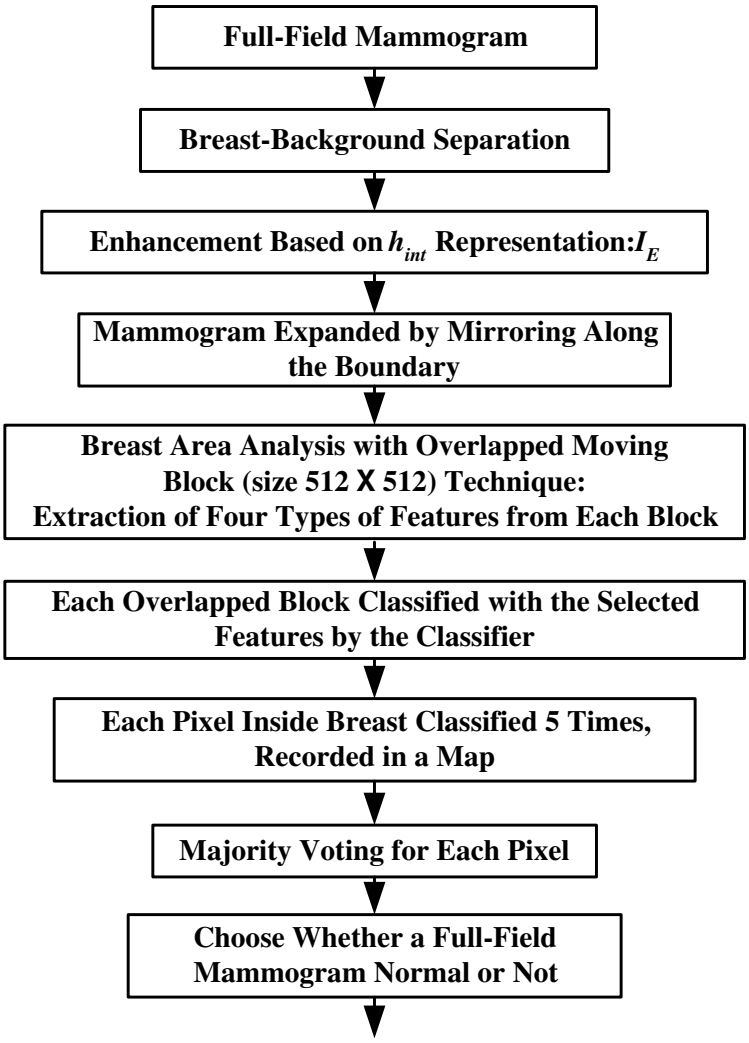


Fig. 5.1. Full-Field Mammogram Analysis

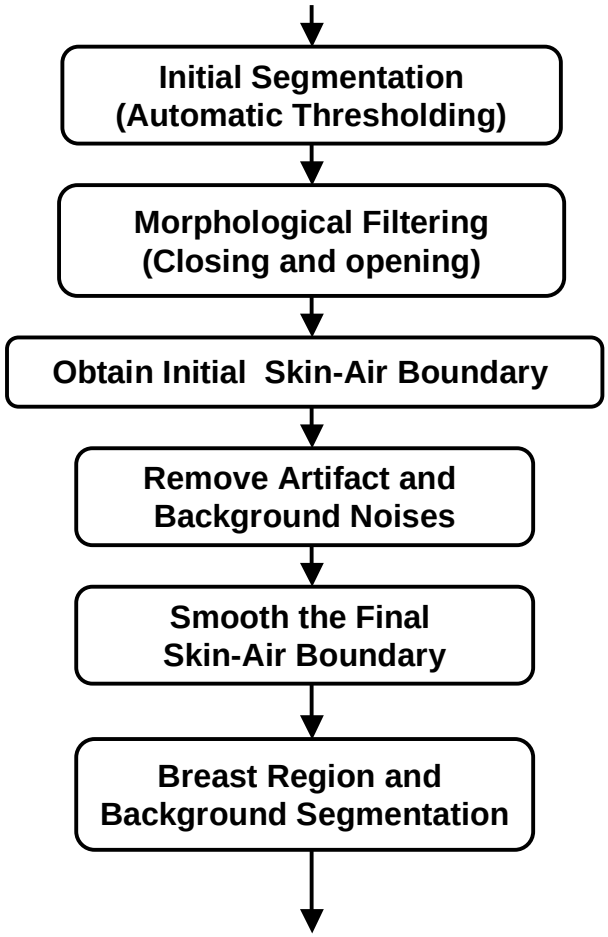


Fig. 5.2. Block diagram of breast-background segmentation

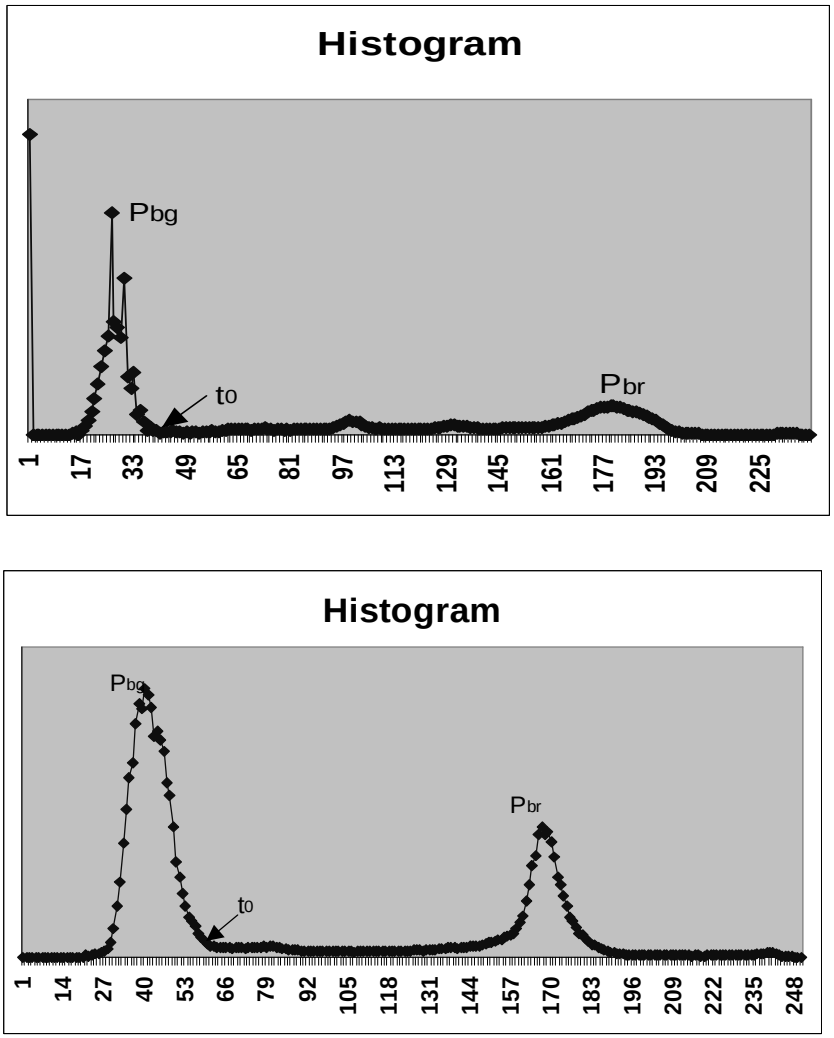


Fig. 5.3. Illustration of automatic threshold t_0 selection from the histogram of a full-field mammogram, where P_{bg} indicates the mode of the background area, and P_{br} indicates the breast area.

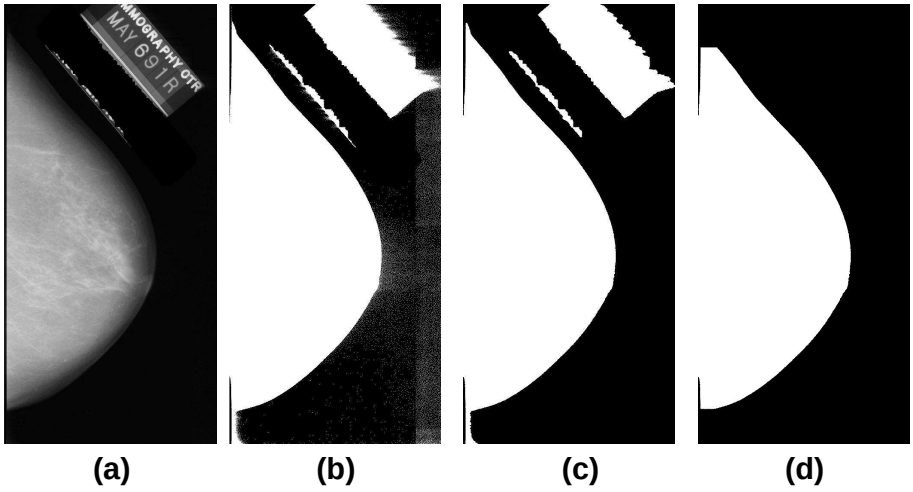


Fig. 5.4. Segmentation of the breast-background: (a) an original mammogram, (b) Initial segmentation after thresholding, (c) After morphological filtering and (d) Final segmentation after artifacts and unexposed rows are removed

5.2 Region Based Full-Field Analysis

Since the majority of screening mammograms are normal, it would be very beneficial to have a CAD system that could readily identify normal mammograms with a very high TNF, the correct classification rate of normal mammograms. The system should also have a very low FNF, the misclassification rate of abnormal mammograms as normal. Hence, a normal detection system could be used as a first reader at screening to reduce the workload of a radiologist and improve the screening performance of a radiologist. The performance is sought to be comparable to or better than the average performance of human readers. Our current goal of this work is to exceed the desired performance stated in [98]: "The desired performance is to detect roughly 40% to 50% of the normal images with a low probability of classifying abnormal images as normal."

Full-field mammogram analysis in our work is region based using the region analysis described in Section 4.3.3. A moving block scheme is implemented to analyze breast regions of mammograms using our two-stage classification system. The moving block size is 512×512 pixels, which is the same size as the training region. A breast area will be classified by 5 overlapped blocks. The first block is centered on a pixel, the block is then moved 128 pixels up, down, right, and left. Each center subregion (size 256×256 pixels) is therefore classified 5 times except for the boundary pixels. Figure 5.5 shows the five overlapped moving blocks. The overlapping block strategy reduces the risk of misclassification when an abnormality is located on the boundaries of the 256×256 center subregion. Since each pixel is classified 5 times and may have five classification labels, we need to design a scheme to assign a unique class label (either normal N or abnormal S) to each pixel in the block. We use a majority voting scheme to determine if each pixel is classified as normal or abnormal. Majority voting can be considered as an additional stage of classification. Finally, a full-field mammogram is classified as an abnormal image if one or more subregions are classified as abnormal, otherwise, the mammogram is classified as normal.

Before using this scheme, the segmented full-field mammograms are expanded 128 pixels along the boundaries in order to reduce edge effects and are mirrored from the mammogram along the edge. The following describes the procedures:

1. Segment the breast-background to obtain the breast region
2. Preprocess the mammogram using h_{int} representation to obtain I_E
3. Allocate a memory map NIM that is the same size as the mammogram
4. Analyze the breast area with the moving block (size 512×512), moving every 128 pixels in row or column directions
5. If the background in the current block is less than 35% of the block then mirror the breast pixels in the same row and fill the background within the block - otherwise ignore this position (if the entire row is in the background, fill in the row from the other end or the middle row)
6. Extract four sets of features from each block
    - 18 curvilinear features
    - 16 GLCM features
    - 32 Gabor features
    - 20 multiresolution features
7. Use the Power Function Transform to "normalize" the 86 features as described in Section 3.4
8. Classify 86 features associated with the current block using the trained two-stage classification system described in Section 4.3.4. We selected the sixth entry for region analysis at Table 4.2 as the operating point, which has $TNF = 0.936$ and $FNF = 0.091$. The corresponding threshold used in the second-stage linear classifier is 0.16667.

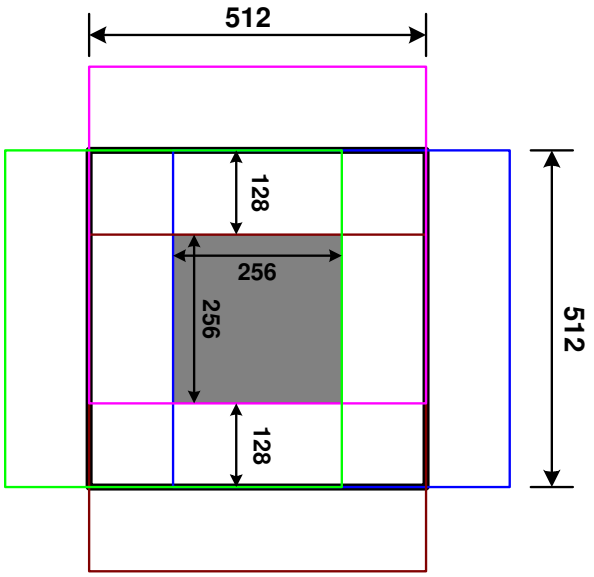


Fig. 5.5. Overlapped Block Scheme: Center subregion is classified 5 times

9. Record the class label (Normal or Abnormal) in the memory map NIM corresponding to center subregion of the current block

10. Use the majority voting on the memory map NIM to obtain the binary detection result of full-field normal analysis

11. Label a full-field mammogram as an abnormal image if one or more subregions are classified as abnormal; otherwise, as normal

5.2.1 Validation of Moving Block Indexing and Background Filling

There are some tedious implementation details associated with the overlapped moving block scheme. Moving block indexing and background filling in the boundary block are two issues. We will use phantom images to validate block indexing and boundary filling.

The purpose is to make sure that blocks are correctly moved 128 pixel in four directions and the center subregion (size 256×256) is processed 5 times. The test is carried out as follows: 1) Obtain the average pixel value I_i within the current moving block; and 2) The output for the center subregion is the sum of 5 I_i's from the 5 overlapped moving blocks instead of the majority voting used in the real-scenario.

The test image is an image with one 256×256 region (denoted as R_*) of pixel value 255 surrounded by zeros. The shape mimics the breast area. The background in the phantom image is gray. Figure 5.6 (a) shows the phantom image. Using the same moving block scheme, if the current moving block (size 512×512) is centered on R_*, R_* will be still in the block if the moving block is moved 128 pixels up, down, left and right. Hence, we will obtain the maximum output. If the center subregion of the current moving block is up, down, left or right to R_*, then only half of R_* is within the block. If the center subregion of the current moving block is at the corner of R_*, then only a quarter of R_* is within the block. Figure 5.6 (b) shows the test result, which is scaled to 0-255. We can clearly observe that the moving block indexing is correct and each subregion is classified 5 times.

The following test is carried out to verify the background filling at the boundary blocks. The test image is a phantom of periodically changing gray stripes. The output image is an image after the boundary blocks having the background filled with mirrored pixels as described in Section 5.2. Figure 5.7 (a) shows the test phantom image with Figure 5.7 (b) being the result. We can observe that the boundary filling works well. It should be emphasized that feature extraction and classification of the filled boundary blocks reduces the edge effects, however, the classification results will only be recorded for the breast area within the boundary block, not the background within the boundary block.

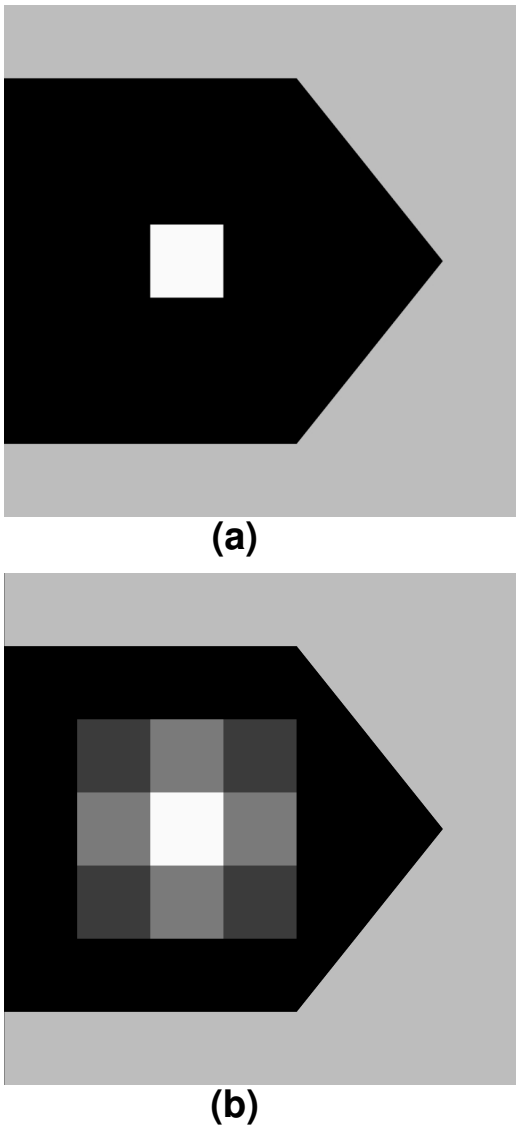


Fig. 5.6. (a) A test phantom image (b) The moving block indexing test result.

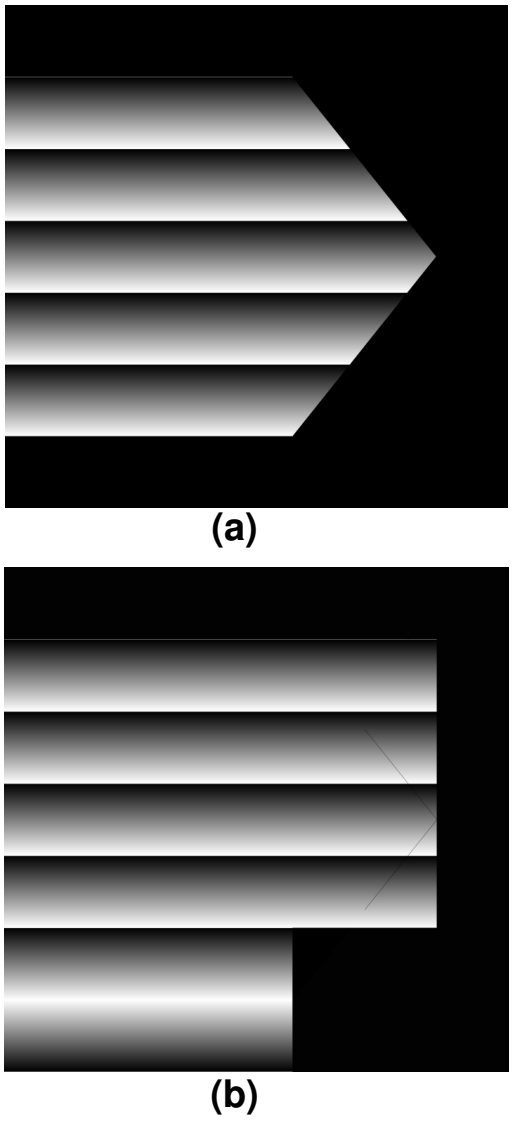


Fig. 5.7. (a) A test phantom image (b) Test result image of background filled with mirrored pixels on the boundary blocks

Table 5.1
Four Performance Fractions of Normal Analysis on Full-Field Mammograms

TPF	FPF	TNF	FNF
0.800	0.299	0.701	0.200

Table 5.2
False Negatives of Normal Analysis

Number of Misclassifications		
Calcifications	Mammograms Tested	38
	False Negatives	12
Masses	Mammograms Tested	37
	False Negatives	5
Spiculations	Mammograms Tested	35
	False Negatives	5

5.2.2 Normal Analysis on Full-Field Mammograms

The system was used to analyze full-field mammograms as described in Section 5.2. We tested 110 cancer mammograms and 144 normal mammograms. Among the 110 cancer mammograms, 38 were calcification images, 37 were mass images and 35 were spiculation images. 101 normal mammograms and 88 cancer mammograms were classified correctly. Table 5.1 shows the performance of our normal analysis. We obtain $TPF = 0.800$ or 80.0%, $FPF = 0.299$ or 29.9%, $TNF = 0.701$ or 70.1% and $FNF = 0.200$ or 20.0%. Table 5.2 shows false negatives, full-field cancer mammograms misclassified as normal.

70.1% of normal mammograms are correctly classified, and 20.0% of cancer mammogram were classified as normal. This is comparable to human readers since 20% to 40% of breast cancer fail to be detected (false negative mammogram read-

ings) at screening [11, 12]. Most of the misclassified cancer images are calcifications. The region of analysis might be too large for small clusters of calcifications. Excluding calcifications, we obtain a FNF of less than 13.9% (from Table 5.2, $\frac{5+5}{37+35} = 0.139$ or 13.9%) on mass and spiculation images. This FNF in our normal detection scheme is very comparable to the misclassification rate of abnormal as normal in most CAD mass detection systems. One of the main reasons for misclassification is due to the subtlety of the cancers. There were mainly two types of misclassification of normal mammograms: one was due to dense normal tissue; the other was due to the boundary between the pectoral muscle and the breast region, which could be removed before processing.

In order to visualize the 5 overlapped classifications of each pixel in the memory map NIM before the majority voting, the memory map NIM is initialized to zero, we add +1 to NIM if the current moving block is classified as abnormal; otherwise add -1 to NIM. This is then stretched to display as an image as shown in Figure 5.8 (d) or Figure 5.15 (c). The binary detection result is obtained after the majority voting as shown in Figure 5.8 (e) or Figure 5.15 (d). A mammogram is labelled as an abnormal image if one or more subregions are classified as abnormal; otherwise, it is labelled as normal. Hence, Figure 5.8 (a) is classified as abnormal according to Figure 5.8 (e). Figure 5.15 (a) is classified as normal according to Figure 5.15 (d). Figures 5.8, 5.9, 5.10, 5.11, 5.12, 5.13 and 5.14 show the results of abnormal mammograms which are classified correctly by our analysis. Figure 5.8 shows the correct detection of a mammogram of a circumscribed mass. Figure 5.10 shows the result of a microcalcification case. Figure 5.11 shows the detection result of a spiculated lesion.

Figures 5.15, 5.16, 5.17, 5.18, 5.19 and 5.20 show normal mammograms classified correctly. Figure 5.16 shows a mammogram of a very dense breast is correctly classified. In Figure 5.20, a screening mammogram of a fatty breast is classified correctly.

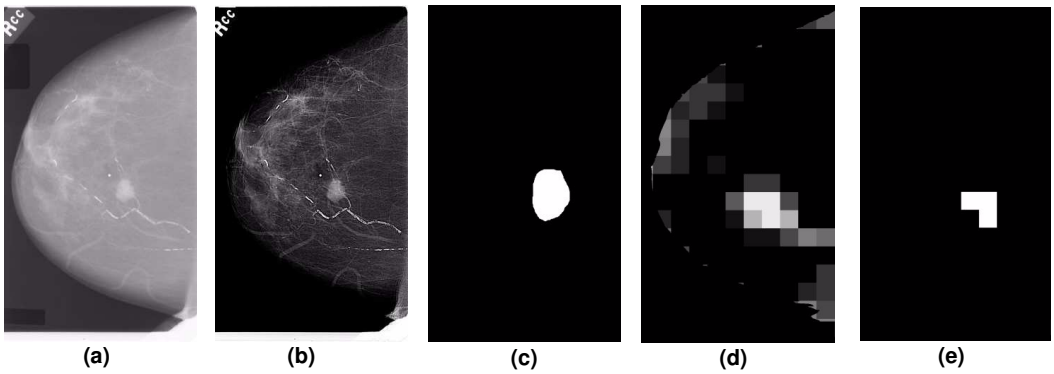


Fig. 5.8. A mass case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Ground-truth Cancer Template (d) Visualization of 5 Overlapped Classifications, and (e) Binary Detection Result

5.3 Conclusions

We presented a full-field analysis method which focuses on identifying normal mammograms instead of individual cancer detection. The challenge of this approach is that the normal class is a heterogeneous mixture of normal mammograms of different densities and the abnormal class is a heterogeneous mixture of all types of lesions with varying sizes. Our current work shows the method described above can identify normal mammograms with a fairly high TNF while having a fairly low FNF. Our FNF is below the average false negative rate (20% to 40%) of human readers [11,12]. The ideal performance of a normal detection system will be that a very high detection rate (TNF) of normal mammograms with a very low misclassification rate (FNF) of abnormal mammograms. We may have to find a trade-off operating point on the ROC curve since high sensitivity and high specificity are at the opposite ends of of the ROC curve. Though our normal analysis result is not ready for the clinical use, the TNF and FNF of our current test are very comparable to human readers and could be used as a possible "pre-screening" system.

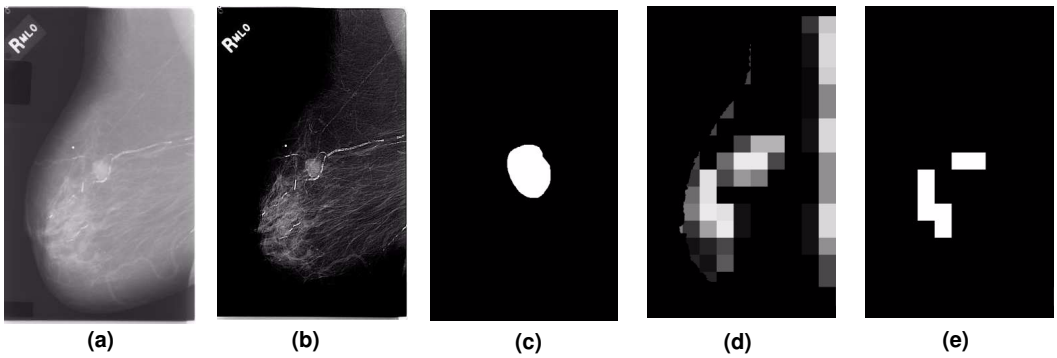


Fig. 5.9. A mass case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Ground-truth Cancer Template (d) Visualization of 5 Overlapped Classifications, and (e) Binary Detection Result

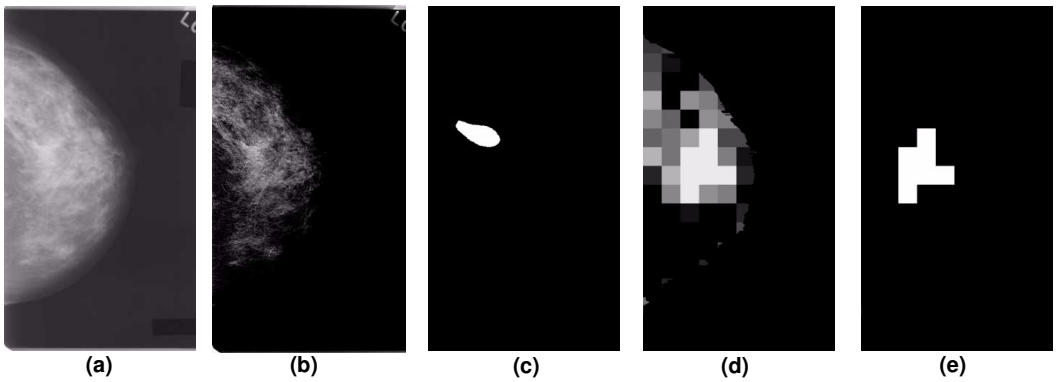


Fig. 5.10. A microcalcification case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Ground-truth Cancer Template (d) Visualization of 5 Overlapped Classifications, and (e) Binary Detection Result

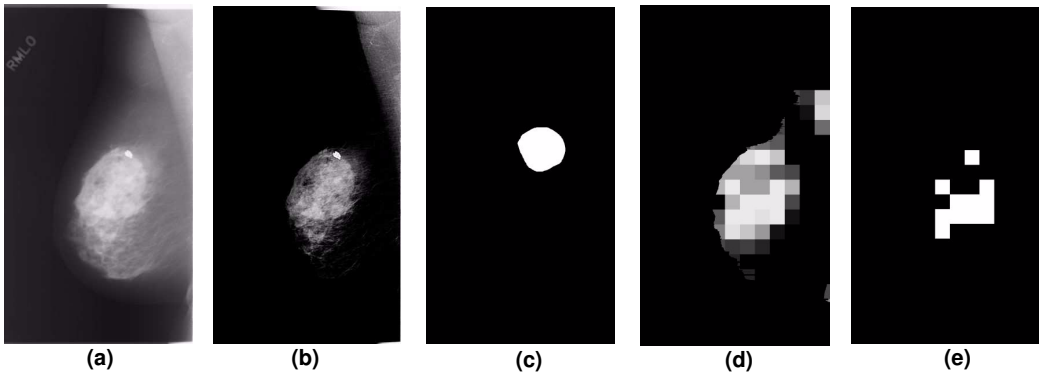


Fig. 5.11. A spiculation case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Ground-truth Cancer Template (d) Visualization of 5 Overlapped Classifications, and (e) Binary Detection Result

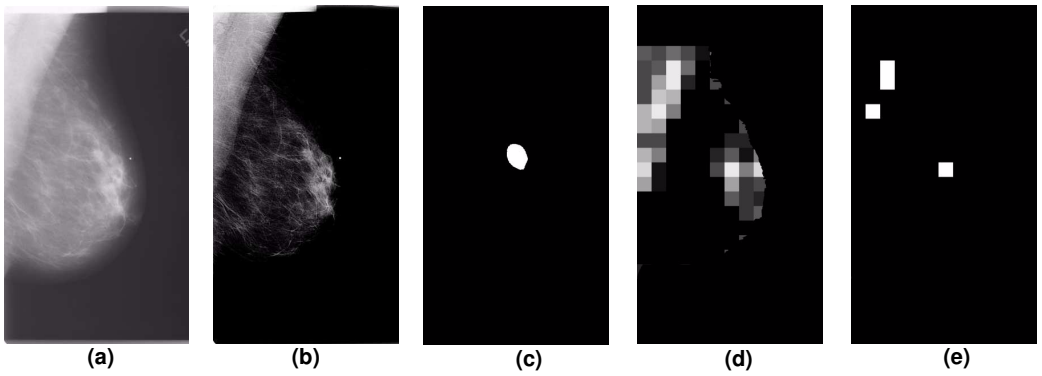


Fig. 5.12. A spiculated lesion case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Ground-truth Cancer Template (d) Visualization of 5 Overlapped Classifications, and (e) Binary Detection Result

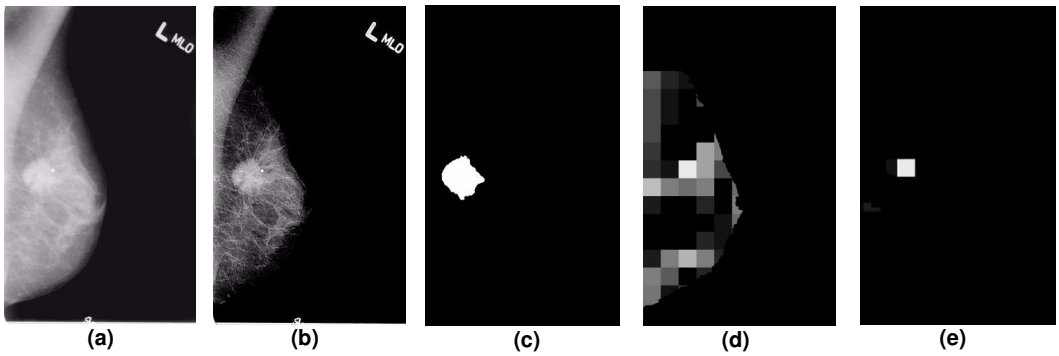


Fig. 5.13. A spiculation case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Ground-truth Cancer Template (d) Visualization of 5 Overlapped Classifications, and (e) Binary Detection Result

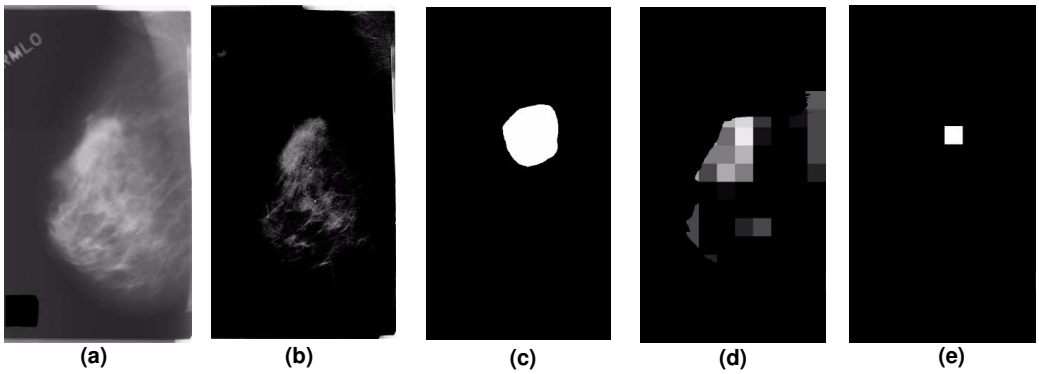


Fig. 5.14. A microcalcification/mass case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Ground-truth Cancer Template (d) Visualization of 5 Overlapped Classifications, and (e) Binary Detection Result

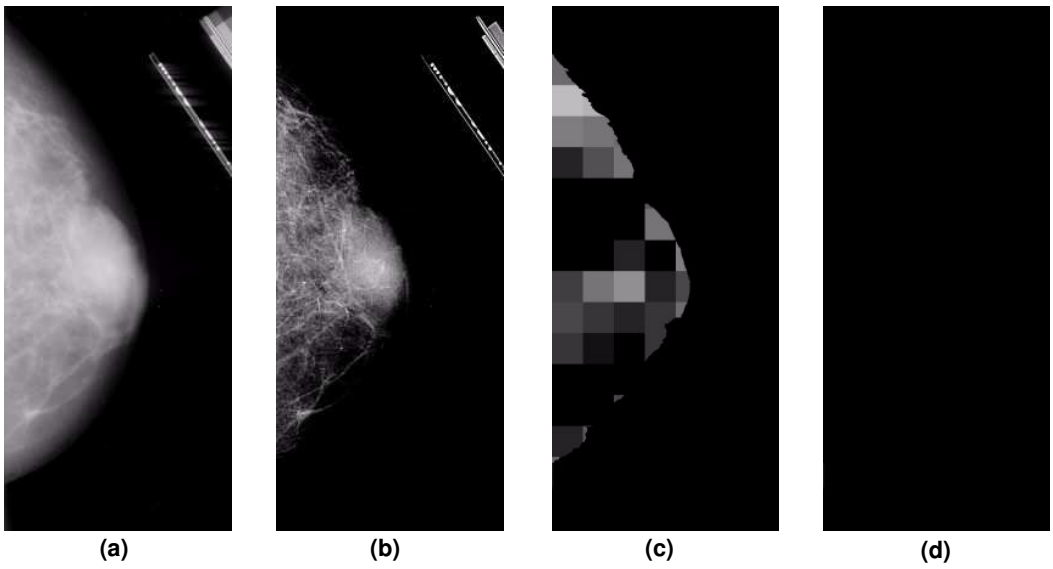


Fig. 5.15. A normal case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Visualization of 5 Overlapped Classifications, and (d) Binary Detection Result

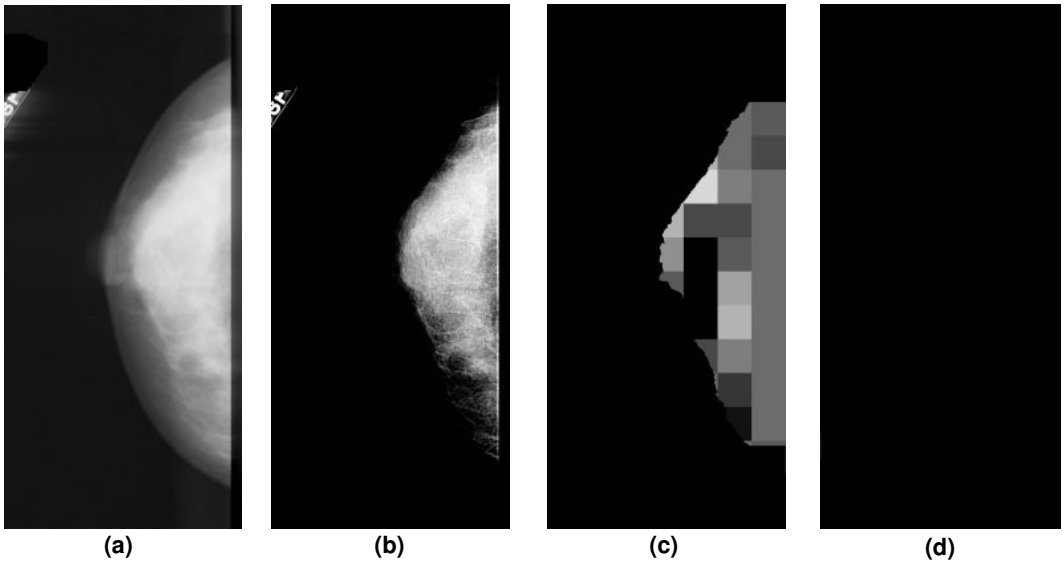


Fig. 5.16. A normal case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Visualization of 5 Overlapped Classifications, and (d) Binary Detection Result

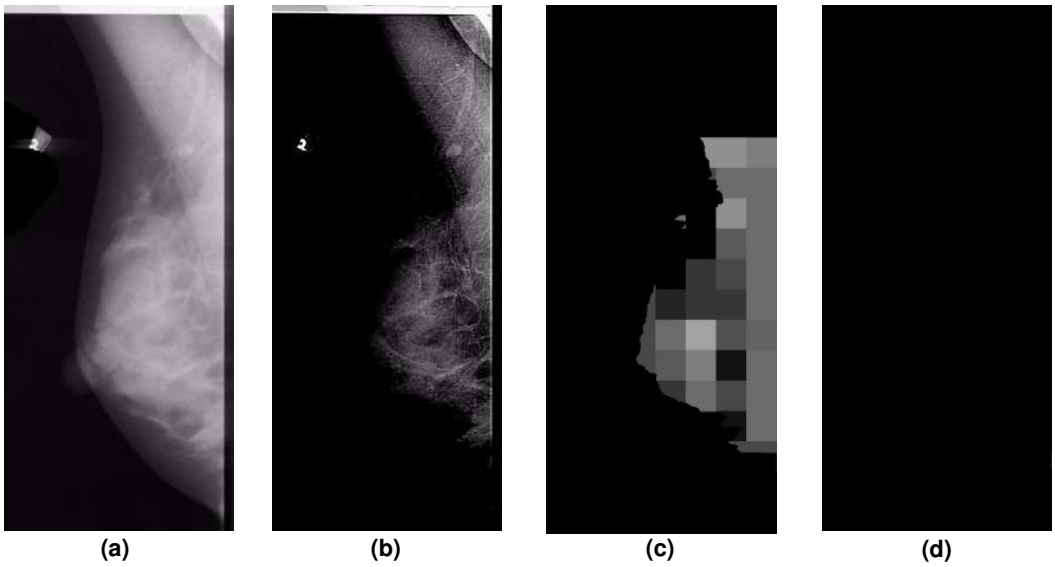


Fig. 5.17. A normal case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Visualization of 5 Overlapped Classifications, and (d) Binary Detection Result

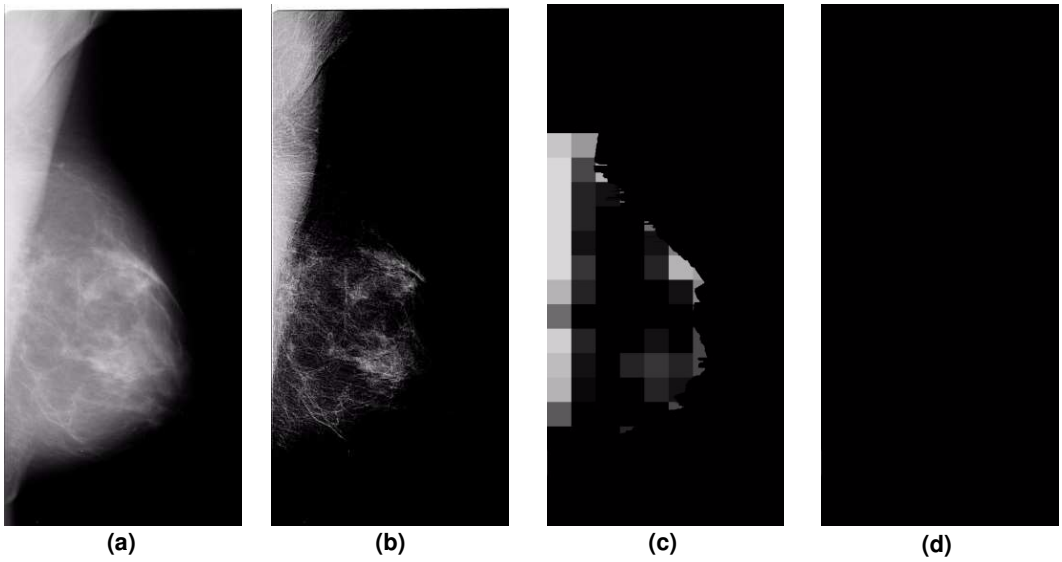


Fig. 5.18. A normal case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Visualization of 5 Overlapped Classifications, and (d) Binary Detection Result

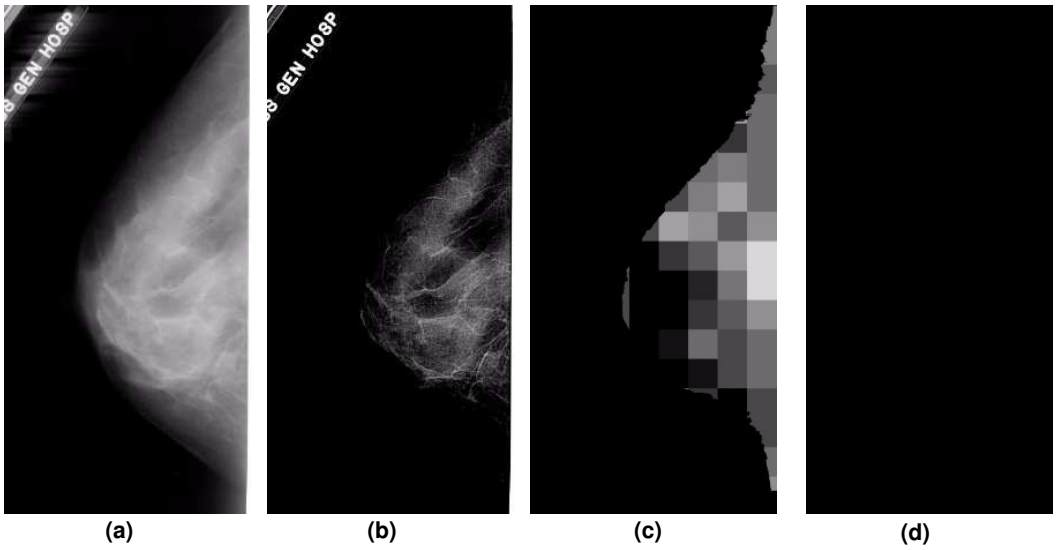


Fig. 5.19. A normal case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Visualization of 5 Overlapped Classifications, and (d) Binary Detection Result

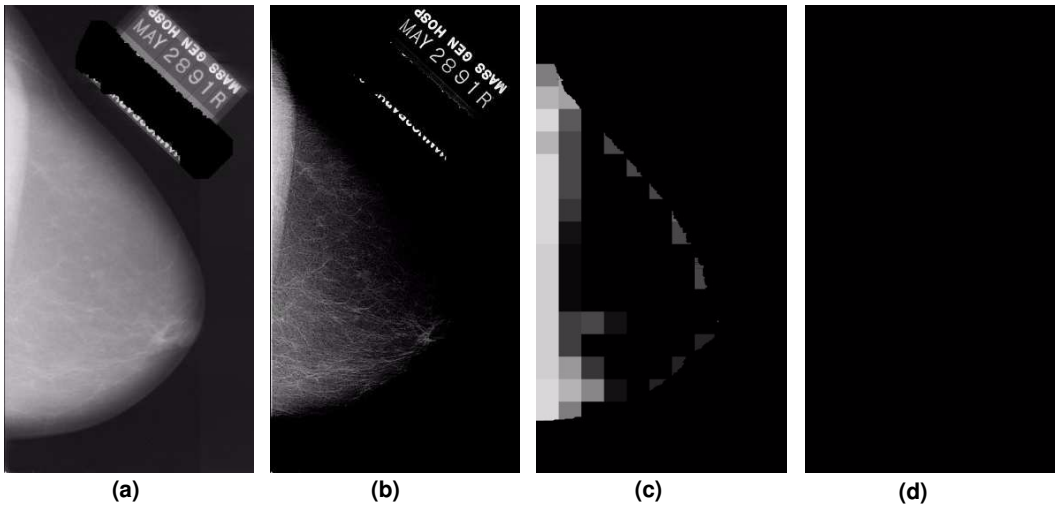


Fig. 5.20. A normal case: (a) Full-Field Mammogram, (b) Enhanced h_{int} Image: I_E, (c) Visualization of 5 Overlapped Classifications, and (d) Binary Detection Result

6. CONCLUSIONS AND FUTURE WORK

In this dissertation we proposed a computer-aided diagnosis (CAD) system which focuses on identifying normal mammograms instead of individual cancer types. Since the majority of screening mammograms are normal our strategy can help a radiologist in screening. In this chapter, we will briefly summarize the main contributions. The last section is dedicated to possible future development.

6.1 Concluding Remarks

The main contributions of this dissertation are: 1) we have defined four sets of features (Curvilinear, Gray Level Co-occurrence, Gabor and Multiresolution features) which can be used to separate normal and abnormal mammograms. Our work showed that each feature set compliments each other; 2) we have proposed a unique multi-stage cascading classification system which increases TNF, the correct classification rate of normal mammograms, while keeping a low FNF, the misclassification rate of abnormal mammograms as normal.

Our normal mammogram region analysis showed that the multi-stage classifier improves TNF through reducing FPF. The cascaded classifier has some interesting properties including the ability of obtaining ROC analysis. The classifier can classify normal mammogram regions with a high TNF, while keeping the clinically critical misclassification rate of abnormal regions as normal (FNF) low. The overall performance of the classifier can also be formulated explicitly. The classifier was then used to classify full-field mammograms using a moving block scheme. The result of our study was comparable to human readers. It showed that our system can identify normal mammogram with a high TNF, though it needs some improvement for clinical use. The performance disparity between region and full-field classification is

because the limited training regions could not represent all types of breast areas in full-field mammograms.

6.2 Future Work

We have defined a set of 86 feature from four types. We showed that each type compliments each other, and only a subset of features is used for the actual classification of normal mammograms from abnormal mammograms. One possible improvement of this system is to do a thorough study of the contributions of each feature. It could help refine a feature or a subset of features if the contribution of a feature or this subset of features to the classification is known. If some current features contribute little to the classification we could investigate new features. It could also be very helpful if we can categorize different subsets of features characterizing different mammograms or regions. It would allow channelized analysis to improve the overall classification performance.

One of the main reasons of the performance disparity between region and full-field classification is that the training region set is limited and may not represent all types of the breast areas of full-field mammograms. One possible improvement is to build a large training region set which could reduce the disparity. Another possible improvement is that another level of classification could be built on the misclassification breast areas of current full-field mammogram analysis.

Our overall system of normal mammogram analysis is computation intensive. The line detection algorithm used for extracting the curvilinear structure is slow even though it is robust and reliable. It may be desired in clinical use to increase computation performance of the overall system. It could be very helpful if a detection could be done in real-time at the screening. If a mammogram from the first view (such as CC view) could be classified as an unequivocal normal mammogram in only a few seconds, the woman would not have to endure the pain of having the second

view (such as MLO view) of the breast taken. The scenario could also reduce the exposure to X-rays.

The classification strategy is another topic for future work. Different cascading structure could be explored, such as re-classifying the training data classified as "normal" from the previous stage, i.e. changing the role of abnormal and normal classes. The study of the combinations of different classification algorithms could also help improve the overall performance. Since the data is reduced after each cascading step in our system, there may not be enough data near the last stage to make an unbiased classification. One possible improvement of the later stage classification is to combine the bootstrap [178, 179] or "boosting" [180–183] technique with the cascading classification.

LIST OF REFERENCES

LIST OF REFERENCES

[1] L. Tabar and P. Dean, *Teaching atlas of mammography.* New York: Thime, 3rd ed., 2001.

[2] L. Garfinkel, M. Catherind, C. Boring, and C. Heath, "Changing trends: An overview of breast cancer incidence and mortality," *Cancer*, vol. 74, no. 1, pp. 222–227, 1994.

[3] S. Parker, T. Tong, S. Bolden, and P. Wingo, "Cancer statistics, 1996," *Cancer J. Clin*, vol. 46, pp. 5–27, 1996.

[4] R. Doll, P. Payne, and J. Waterhause, *Cancer incidence in five continents.* Geneve, Switzerland: UICC, 1996.

[5] R. Gaughan, "New approaches to early detection of breast cancer makes small gains," *Biophotonics International*, pp. 48–53, Sept./Oct. 1998.

[6] L. Tabar, G. Fagerberg, S. Duffy, N. Day, A. Gad, and O. Grontoft, "Update of the swedish two-country program of mammographic screening for breast cancer," *The Radiologic Clinics of North America*, vol. 30, pp. 187–210, 1992.

[7] D. Kopans, *Breast Imaging.* Philadelphia: J.B. Lippincott Company, 1989.

[8] J. Howard, "Using mamography for cancer control: an unrealized potential," *CA: a Cancer Journal for Clinicians*, vol. 37, pp. 33–48, 1987.

[9] *American Cancer Society: Cancer Facts and Figures.* American Cancer Society, 2002.

[10] J. Harvey, L. Fajardo, and C. Innis, "Previous mammograms in patients with impalpable breast carcinoma: Retrospective vs blinded interpretation," *American Journal of Radiology*, vol. 161, pp. 1167–1172, 1993.

[11] C. Beam, P. Layde, and D. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists, findings from a national sample," *Archives of Internal Medicine*, vol. 156, pp. 209–213, 1996.

[12] J. Elmore, C. Wells, C. Lee, D. Howard, and A. Feinstein, "Variability in radiologists' interpretations of mammograms," *New England Journal of Medicine*, vol. 331, pp. 1493–1499, 1994.

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

[14] E. Sickles, S. Ominsky, R. Sollitto, H. Galvin, and D. Monticciolo, "Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations," *Radiology*, vol. 175, pp. 323–327, 1990.

[15] M. Linver, S. Paster, R. Rosenberg, C. Key, C. Stidley, and W. King, "Improvement in mammography interpretation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38,633 cases," *Radiology*, vol. 184, pp. 39–43, 1992.

[16] J. G. Elmore, D. L. Miglioretti, L. M. Reisch, M. B. Barton, W. Kreuter, C. L. Christiansen, and S. W. Fletcher, "Screening mammograms by community radiologists: Variability in false-positive rates," *Journal of National Cancer Institute*, vol. 94, pp. 1373–1380, 2002.

[17] E. Thurfjell, K. Lernevall, and A. Taube, "Benefit of independent double reading in a population-based mammography screening program," *Radiology*, vol. 191, pp. 241–244, 1994.

[18] R. Schmidt, D. Wolverton, and C. Vyborny, "Computer-aided diagnosis (CAD) in mammography," in *Syllabus: A Categorical Course in Breast Imaging*, pp. 199–208, 1995.

[19] M.L.Giger, "Current issues in mammography," *Proceedings of the 3rd International Workshop on Digital Mammography*, pp. 53–59, Chicago, IL, June 1996.

[20] C. Vyborny and M. Giger, "Computer vision and artificial intelligence in mammography," *American Journal of Roentgenology*, vol. 162, no. 3, pp. 699–708, 1994.

[21] R. Reid, "Professional quality assurance for mammography screening programs," *Radiology*, vol. 177, pp. 8–10, 1990.

[22] C. Metz and J. Shen, "Gains in accuracy from replicated readings of diagnostic images: Predication and assessment in terms of ROC analysis," *Medical Decision Making*, vol. 12, pp. 60–75, 1992.

[23] R. Schmidt, R. Nishikawa, and K. Schreibman, "Computer detection of lesions missed by mammography," *Proceedings of the 2nd International Workshop on Digital Mammography*, pp. 289–294, July 10-12 1994.

[24] R. Schmidt, R. Nishikawa, R. Osnis, K. Schreibman, M. Giger, and K. Doi, "Computerized detection of lesions missed by mammography," *Proceedings of the 3rd International Workshop on Digital Mammography*, pp. 105–110, June 9-12 1996.

[25] "R2 technology pre-market approval (PMA) of the M1000 image checker," *US. Food and Drug Administration (FDA) application #P970058*, approved, June 26, 1998.

[26] R. Highnam and M. Brady, *Mammographic Image Analysis*. Dordrecht: Kluwer Academic Publishers, 1999.

[27] L. Bassett, V. Jackson, R. Jahan, Y. Fu, and R. Gold, *Diagnosis of diseases of the breast*. W.B. Saunders Company, Philadelphia, PA, 1997.

[28] C. Metz, "ROC methodology in radiologic imaging," *Investigative Radiology*, vol. 21, pp. 720–733, 1986.

[29] C. Metz, "Evaluation of digital mammography by ROC analysis," *Proceedings of the 3rd International Workshop on Digital Mammography*, pp. 61–68, June 9-12 1996.

[30] C. Metz, "Receiver operating characteristic (ROC) analysis in medical imaging," *ICRU News*, pp. 7–16, 1997.

[31] D. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Medical Physics*, vol. 16, p. 561, 1989.

[32] D. Chakraborty and L. Winter, "Free-response methodology: alternate analysis and a new observer-performance experiment," *Radiology*, vol. 174, p. 873, 1990.

[33] R. Swensson, "Unified measurement of observer performance in detecting and localizing target objects on images," *Medical Physics*, vol. 23, p. 1709, 1996.

[34] R. Wagner, S. Beiden, and C. Metz, "Continuous vs. categorical data for ROC analysis: Some quantitative considerations," *Academic Radiology*, vol. 8, pp. 328–334, 2001.

[35] N. Karssemeijer, "Stochastic model for automated detection of calcifications in digital mammograms," *Image and Vision Computing*, vol. 10, no. 6, pp. 369–375, 1992.

[36] N. Karssemeijer and W. Veldkamp, "Normalisation of local contrast in mammograms," *IEEE Transactions on Medical Imaging*, vol. 19, no. 7, pp. 731–738, 2000.

[37] K. McLoughlin, P. Bones, and P. Dachman, "Connective tissue representation for detection of microcalcification in digital mammograms," *Proceedings of SPIE Medical Image 2002: Image Proceesing*, vol. 4684, pp. 1246–1256, 2002.

[38] M. Gurcan, Y. Yardimci, A. Cetin, and R. Ansari, "Detection of microcalcifications in mammograms using higher order statistics," *IEEE Signal Processing Letters*, vol. 4, no. 8, pp. 213–216, Aug. 1997.

[39] N. Karssemeijer, "Adaptive noise qualization and image analysis in mammography," *13th International Conference on Information Processing in Medical Imaging*, pp. 472–486, June 14-18 1992.

[40] A. Bagci, T. Yardimci, and A. Cetin, "Detection of microcalcification clusters in mammogram images using local maxima and adaptive wavelet transform analysis," *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3856 –3859, May 13 - 17 2002.

[41] M. Gurcan, H. Chan, B. Sahiner, L. Hadjiiski, N. Petrick, and M. Helvie, "Optimal neural network architecture selection: effects on computer-aided detection of mammographic microcalcifications," *Proceedings of SPIE Medical Image 2002: Image Proceesing*, vol. 4684, pp. 1325–1330, 2002.

[42] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, and R. Nishikawa, "Support vector machine learning for detection of microcalcifications in mammograms," *Proceedings of the IEEE 2002 International Symposium on Biomedical Imaging*, pp. 201–204, July 7-10 2002.

[43] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, and R. Nishikawa, "A support vector machine approach for detection of microcalcifications in mammograms," *Proceedings of the IEEE 2002 International Conference on Image Processing*, pp. 953–956, Sept. 22-25 2002.

[44] W. Zhang, H. Yoshida, R. Nishikawa, and K. Doi, "Optimally weighted wavelet transform based on supervised training for detection of microcalcification in digital mammograms," *Medical Physics*, vol. 25, pp. 949–956, June 1998.

[45] B. Sahiner, M. Gurcan, H. Chan, L. Hadjiiski, N. Petrick, and M. Helvie, "Use of joint two-view information for computerized lesion detection on mammograms: improvement of microcalcification detection accuracy," *Proceedings of SPIE Medical Image 2002: Image Proceesing*, vol. 4684, pp. 754–761, 2002.

[46] W. Veldkamp, N. Karssemeijer, J. Otten, and J. Hendriks, "Automated classification of clustered microcalcification into malignant and benign types," *Medical Physics*, vol. 27, pp. 2600–2608, November 2000.

[47] H. Soltanian-Zadeh, S. Pourabdollah-Nezhad, and F. Rafiee-Rad, "Shape-based and texture-based feature extraction for classification of microcalcifications in mammograms," *Proceedings of SPIE Medical Image 2001: Image Proceesing*, vol. 4322, pp. 301–310, 2001.

[48] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions On Systems, Man, and Cybernetics*, vol. SMC-3, pp. 610–621, November 1973.

[49] C. Ping, B. Sahiner, N. Petrick, M. Helvie, L. K. Leung, D. Adler, and M. Goodsitt, "Computerized classificaiton of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network," *Physics in Medicine and Biology*, vol. 42, no. 3, pp. 549–567, 1997.

[50] H.-P. Chan, B. Sahiner, K. Lam, N. Petrick, M. Helvie, M. Goodsitt, and D. Adler, "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," *Medical Physics*, vol. 25, no. 10, pp. 2007–2019, 1998.

[51] J. Lo, M. Gavrielides, M. Markey, and J. Jesneck, "Computer-aided classification of breast microcalcification clusters: merging of features from image processing and radiologists," *Proceedings of SPIE Medical Imaging 2003: Image Processing*, vol. 5032, pp. 882–889, 17-20 February 2003.

[52] *Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas).* American College of Radiology, 5th ed., 2003.

[53] L. Hadjiiski, H. Chan, B.Sahiner, N. Petrick, M. Helvie, M. Roubidoux, and M. Gurcan, "Computer-aided characterization of malignant and benign microcalcification clusters based on the analysis of temporal change of mammographic features," *Proceedings of SPIE Medical Image 2002: Image Proceesing*, vol. 4684, pp. 749–753, 2002.

[54] S. Pohlman, K. Powell, N. Obuchowski, W. Chilcote, and S. Grundfest-Broniatowski, "Quantative classification of breast tumor in digitized mammograms," *Medical Physics*, vol. 23, no. 8, pp. 1337–1345, 1996.

[55] D. Guliato, R. Rangayyan, W. Carnelli, J. Zuffo, and J. Desautels, "Segmentation of breast tumors in mammograms by fuzzy region growing," *Proceedings of 20th Annual International Conference IEEE Engineering in Medicine and Biology*, pp. II: 1002–1004, Oct.29-Nov.1, 1998.

[56] D. Guliato, R. Rangayyan, J. Zuffo, and J. Desautels, "Detection of breast tumor boundaries using iso-intensity contours and dynamic thresholding," *Proceedings of the 4th International Workshop on Digital Mammography*, pp. 253–260, Nijmegen, Netherlands, June 1996.

[57] N. Petrick, H. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Transactions on Medical Imaging*, vol. 15, pp. 59–67, Jan. 1996.

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

[59] N. Mudigonda, R. Rangayyan, and J. Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1215–1227, Dec. 2001.

[60] J. Suckling, J. Parker, D. Dance, S. Astley, I. H. anc C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, "The mammographic image analysis society digital mammogram database," *Proceedings of the 2nd International Workshop on Digital Mammography*, pp. 375–378, July 10-12 1994.

[61] D. Catarious, A. Baydush, C. Abbey, and C. Floyd, "A mammographic mass cad system incorporating features from shape, fractal, and channelized hotelling observer measurements: preliminary results," *Proceedings of SPIE Medical Imaging 2003: Image Processing*, vol. 5032, pp. 111–119, 17-20 February 2003.

[62] M. L. Comer, S. Liu, and E. J. Delp, "Statistical segmentation of mammograms," *Proceedings of the 3nd International Workshop on Digital Mammography*, pp. 475–478, June 9-12 1996.

[63] H. Li, M. Karllergi, L. Clarke, V. Jain, and R.A.Clark, "Markov random field for tumor detection in digital mammography," *IEEE Transactions on Medical Imaging*, vol. 14, pp. 565–576, Sept. 1995.

[64] C. Chen and G. Lee, "Image segmentation using multiresolution wavelet analysis and expectation-maximization (EM) algorithm for digital mammography," *International Journal of Imaging System and Technology*, vol. 8, no. 5, pp. 491–504, 1997.

[65] Y. Chang, X. Wang, L. Hardesty, C. Hakim, B.Zheng, W. Good, and D. Gur, "Incorporation of negative regions in a knowledge-based computer-aided detection scheme," *Proceedings of SPIE Medical Image 2002: Image Proceesing*, vol. 4684, pp. 726–732, 2002.

[66] I. Christoyianni, E. Dermatas, and G. Kokkinakis, "Automatic detection of abnormal tissue in mammography," *Proceedings of the IEEE 2001 International Conference on Image Processing*, pp. 877 –880, October 7-10 2001.

[67] S.-C. B. Lo, H. Li, Y. Wang, L. Kinnard, and M. Freedman, "A multiple circular path convolution neural network system for detection of mammographic masses," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 150–158, Feb. 2002.

[68] B. Sahiner, H.-P. Chan, D. Wei, N. Petrick, M. Helvie, D. Adler, and M. Goodsitt, "Image feature selection by a genetic algorithm: application to classification of mass and normal breast tissue," *Medical Physics*, vol. 23, no. 10, pp. 1671–1684, 1996.

[69] Y. Xu, S. Neu, C. Ornes, J. Owens, J. Sklansky, and D. Valentino, "Optimization of active contour model parameters using genetic algorithms: segmentation of breast lesions in mammograms," *Proceedings of SPIE Medical Image 2002: Image Proceesing*, vol. 4684, pp. 1406–1414, 2002.

[70] Y. Chu, L. Li, D. Goldgof, and R. Clark, "Classification of masses on mammograms using support vector machine," *Proceedings of SPIE Medical Imaging 2003: Image Processing*, vol. 5032, pp. 940–948, 17-20 February 2003.

[71] X.-P. Zhang, "Multiscale tumor detection and segmentation in mammograms," *Proceedings of the IEEE 2002 International Symposium on Biomedical Imaging*, pp. 213–216, July 7-10 2002.

[72] S. Kwok, R. Chandrasekhar, and Y. Attikiouzel, "Adaptation of the daugman-downing texture demodulation to highlight circumscribed mass lesions on mammograms," *Proceedings of the 14th International Conference on Digital Signal Processing*, vol. 1, pp. 449 –452, 2002.

[73] Y. Chu, L. Li, and R. Clark, "Graph-based region growing for mass segmentation in digital mammography," *Proceedings of SPIE Medical Image 2002: Image Proceesing*, vol. 4684, pp. 1690–1697, 2002.

[74] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 369–378, May 1999.

[75] G.Ertas, H. Gulcur, E. Aribal, and A. Semiz, "Feature extraction from mammographic mass shapes and development of a mammogram database," *Proceedings of the 23rd Annual EMBS International Conference*, pp. 2752–2755, Oct. 25-28 2001.

[76] L. Kinnard, S.-C. Lo, P. Wang, M. Freeman, and M. Chouikha, "Separation of malignant and benign masses using maximum-likelihood modeling and neural networks," *Proceedings of SPIE Medical Image 2002: Image Proceesing*, vol. 4684, pp. 733–741, 2002.

[77] L. Hadjiiski, H. Chan, B. Sahiner, M. Helvie, M. Roubidoux, C. Blane, C. Paramagul, N. Petrick, J. Bailey, K. Klein, M. Foster, S. Patterson, D. Adler, A. Nees, and J. Shen, "Roc study: effects of computer-aided diagnosis on radiologists' characterization of malignant and benign breast masses in temporal pairs of mammograms," *Proceedings of SPIE Medical Imaging 2003: Image Processing*, vol. 5032, pp. 94–101, 17-20 February 2003.

[78] C. Varela, J. Muller, and N. Karssemeijer, "Mammographic mass characterization using sharpness and lobulation measures," *Proceedings of SPIE Medical Imaging 2003: Image Processing*, vol. 5032, pp. 120–129, 17-20 February 2003.

[79] L. Arbach, D. Bennett, J. Reinhardt, and G. Fallouh, "Classification of mammographic masses: comparison between backpropagation neural network (BNN) and human readers," *Proceedings of SPIE Medical Imaging 2003: Image Processing*, vol. 5032, pp. 810–818, 17-20 February 2003.

[80] S. Liu, C. Babbs, and E. Delp, "Multiresolution detection of spiculated lesions in digital mammograms," *IEEE Transactions on Image Processing*, vol. 10, pp. 874–884, June 2001.

[81] S. Liu, *The Analysis of Digital Mammograms: Spiculated Tumor Detection and Normal Mammogram Characterization*. Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, May 1999.

[82] H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Transactions on Medical Imaging*, vol. 15, pp. 235–245, June 1996.

[83] N. Karssemeijer and G. te Brake, "Detection of stellate distortion in mammograms," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 611–619, Oct. 1996.

[84] T. Parr, R. Zwiggelaar, C. Taylor, S. Astley, and C. Boggis, "The detection of stellate lesions in digital mammography," *Proceedings of Medical Image Understanding and Analysis*, pp. 69–72, 1997.

[85] T. Parr, R. Zwiggelaar, C. Taylor, S. Astley, and C. Boggis, "Detecting stellate lesions in mammograms via statistical models," *Proceedings of the 8th British Machine Vision Conference*, pp. 182–192, 1997.

[86] R. Zwiggelaar, T. Parr, J. Schumm, I. Hutt, C. Taylor, S. Astley, and C. Boggis, "Model-based detection of spiculated lesions in mammograms," *Medical Image Analysis*, vol. 3, no. 1, pp. 39–63, 1997.

[87] J. W.P. Kegelmeyer, "Evaluation of stellate lesion in a standard mammogram data set," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 6, pp. 1477–1492, 1993.

[88] W. Kegelmeyer, "Evaluation of stellate lesion detection in a standard mammogram data set," *Proceedings of SPIE Biomedical Image Processing and Biomedical Visualization*, vol. 1905,, pp. 787–798, 1993.

[89] J. W.P. Kegelmeyer, J. Pruneda, P. Bourland, A. Hillis, M. Riggs, and M. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology*, vol. 191, no. 2, pp. 331–337, May 1994.

[90] M. Heath, K. Bowyer, D. Kopans, R. Moore, and J. P. Kegelmeyer, "The digital database for screening mammography," *Proceedings of the 5th International Workshop on Digital Mammography*, pp. 212–218, June 11-14 2000.

[91] L. Christopher, E. Delp, and C. M. P. Carson, "3-d bayesian ultrasound breast image segmentation using the em/mpm algorithm," *Proceedings of 2002 IEEE International Symposium on Biomedical Imaging*, pp. 86 – 89, 7-10 July 2002.

[92] L. Christopher, *Bayesian Segmentation Of Three Dimensional Images Using The EM/MPM Algorithm.* Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, May 2003.

[93] S.-F. H. R.-F. C. D.-R. C. W. Moon, "Characterization of spiculation on ultrasound lesions," *IEEE Transactions on Medical Imaging*, vol. 23, no. 1, pp. 111 – 121, 2004.

[94] E. Fear, J. Sill, and M. Stuchly, "Experimental feasibility study of confocal microwave imaging for breast tumor detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 887 – 892, March 2003.

[95] L. Dun, P. Meaney, and K. Paulsen, "Conformal microwave imaging for breast cancer detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 4, pp. 1179 – 1186, 2003.

[96] P. Gamagami, M. Silverstein, and J. Waisman, "Infra-red imaging in breast cancer," *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology society, 1997*, vol. 2, pp. 677 – 680, 30 October - 2 November 1997.

[97] J. Lo, J. Baker, P. Kornguth, J. Iglehart, and C. Floyd, "Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features," *Radiology*, vol. 203, pp. 159–163, 1997.

[98] J. Heine, S. R. Deans, D. Cullers, R. Stauduhar, and L. Clarke, "Multiresolution statistical analysis of high-resolution digital mammograms," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 503–515, 1997.

[99] J. Heine, S. R. Deans, and L. Clarke, "Multiresolution probability analysis of random fields," *Journal of the Optical Society of America. Part A*, vol. 16, no. 1, pp. 6–16, 1999.

[100] J. Heine, S. R. Deans, R. Velthuizen, and L. Clarke, "On the statistical nature of mammograms," *Medical Physics*, vol. 26, no. 11, pp. 2254–2265, 1999.

[101] L. Blot and R. Zwiggelaar, "Extracting background texture in mammographic images: a co-occurrence matrices based approach," *Proceedings of the 5th International Workshop on Digital Mammography*, pp. 142–148, June 11-14 2000.

[102] L. Blot and R. Zwiggelaar, "Background texture extraction for the classification of mammographic parenchymal patterns," *Proceedings of Medical Image Understanding and Analysis*, pp. 145–148, July 16-17 2001.

[103] R. Zwiggelaar and C. Rubin, "Separating anatomical structures and background texture in mammographic images," *Proceedings of the 5th International Workshop on Digital Mammography*, pp. 427–435, June 11-14 2000.

[104] S. Liu, C. Babbs, and E. Delp, "Normal mammogram analysis and recognition," *Proceedings of the IEEE International Conference on Image Processing*, pp. 727–731, October 4-7 1998.

[105] T. Parr, C. Taylor, S. Astley, and C. Boggis, "Statistical modelling of oriented line patterns in mammograms," *Proceedings of SPIE Medical Image 1997: Image Proceesing, Newport Beach, CA*, vol. 3034, pp. 44–55, 1997.

[106] R. Zwiggelaar, T. Parr, and C. Taylor, "Finding oriented line pattern in digital mammographic images," *Proceedings of the 7th British Machine Vision Conference*, pp. 715–724, 1996.

[107] R. M. R. Zwiggelaar and C. Boggis, "Detection of linear structures in mammographic images," *Proceedings of the 4th Conference on Medical Image Understanding and Analysis '00*, pp. 97–100, 2000.

[108] R. Zwiggelaar and R. Marti, "Detecting linear structures in mammographic images," *Proceedings of the 5th International Workshop on Digital Mammography*, pp. 436–442, June 11-14 2000.

[109] R. Zwiggelaar and C. Boggis, "Classification of linear structures in mammographic images," *Proceedings of Medical Image Understanding and Analysis*, pp. 1–4, July 16-17 2001.

[110] N. Cerneaz and M. Brady, "Finding curvilinear structures in mammograms," *Lecture Notes in Computer Science*, vol. 905, pp. 372–382, 1995.

[111] V. Schenk and M. Brady, "Finding CLS using multiresolution oriented local energy feature detection," *Digital Mammography IWDM 2002: 6th International Workshop on Digital Mammography*, pp. 64–68, June 22-25 2002.

[112] C. Evans, K. Yates, and M. Brady, "Statistical characterization of normal curvilinear structures in mammograms," *Digtial Mammography IWDM 2002: 6th International Workshop on Digital Mammography*, pp. 285–291, June 22-25 2002.

[113] L. Shen, R. Rangayyan, and J. Desautels, "Application of shape analysis to mammographic calcifications," *IEEE Transactions on on Medical Imaging*, vol. 12, pp. 263–274, June 1994.

[114] J. Marti, X. Cufi, J. Regincos, J. Espanol, J. Pont, and C. Barcelo, "Shape-based features selection for microcalcification evaluation," *Proceedings of SPIE Medical Image 1998: Image Proceesing*, vol. 3338, pp. 1215–1224, 1998.

[115] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic Press, 2nd ed., 1990.

[116] R.O.Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons, Inc., 1973.

[117] R. Bellman, *Adaptive Control Process*. Princeton: Princeton University Press, 1961.

[118] R. Highnam, M. Brady, and B. Shepstone, "Computing the scatter component of mammographic images," *IEEE Transactions on Medical Imaging*, vol. 13, no. 2, pp. 301 –313, June 1994.

[119] R. Highnam, M. Brady, and B. Shepstone, "A representation for mammographic image processing," *IEE Colloquium on Digital Mammography*, pp. 4/1 –4/6, 1996.

[120] R. Highnam, M. Brady, and R. English, "Detecting film-screen artifacts in mammography using a model-based approach," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 1016 –1024, Oct. 1999.

[121] B. Jahne, *Digital Image Processing: concepts, algorithms, and scientific applications*. New York : Springer, 1997.

[122] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, pp. 51–59, 1996.

[123] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Mutliresolution gray scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[124] R. Conners, M. Trivedi, and C. Harlow, "Segmentation of a high-resolution urban scene using texture operators," *Computer Vision, Graphics and Image Processing*, vol. 25, no. 3, pp. 273–310, 1984.

[125] T. Reed and H. Wechsler, "Segmentation of textured images and gestalt organization using spatial/spatial-frequency representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 1–12, 1990.

[126] B. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, vol. 18, pp. 837–842, August 1996.

[127] S. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on gabor filters," *IEEE Transactions on Image Processing*, vol. 11, pp. 1160–1167, October 2002.

[128] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, pp. 467–476, April 2002.

[129] J. Kovačević and M. Vetterli, "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for Rn," *IEEE Transactions on Information Theory*, vol. 38, pp. 535–555, March 1992.

[130] J. Kovačević and M. Vetterli, "Nonseparable two- and three-dimensional wavelets," *IEEE Transactions on Signal Processing*, vol. 43, pp. 1269–1273, May 1995.

[131] A. Mojsilovic, S. Markovic, and M. Popovic, "Texture analysis and classification with the nonseparable wavelet transform," *Proceedings of the IEEE 1997 International Conference on Image Processing*, pp. 182–185, Oct.26-29 1997.

[132] R. Andrews and D. Nguyen, "Separable and quincunx image coding," *Proceedings of the 6th IEEE Workshop on ISPACS*, Nov. 1998.

[133] E. Adelson and E. Simonceli, "Orthogonal pyramid transform for image coding," *Proceedings of SPIE Proceedings of Visual Communications and Image Processing II*, pp. 50–58, 1987.

[134] D. Wei and S. Guo, "A new approach to the design of multidimensional nonseparable two-channel orthonormal filterbanks and wavelets," *IEEE Signal Processing Letters*, vol. 7, pp. 327–330, November 2000.

[135] S. Theodoridis and K. Koutroumbas, *Pattern Classification*. San Diego: Academic Press, 3rd ed., 1999.

[136] L.Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.

[137] J. Quinlan, "Introduction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986.

[138] J. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[139] J. Pearl, "Fusion, propagation, and structuring in belief networks," *Artificial Intelligence*, vol. 29, pp. 241–288, 1986.

[140] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[141] J. Pearl, *Causality*. Cambridge, 2000.

[142] F. Jensen, *Bayesian Networks and Decision Graphs*. Springer, 2001.

[143] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*. Springer-Verlag, 1999.

[144] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford: University Press, 1995.

[145] S. Haykin, *Neural Networks*. Prentice Hall, 1999.

[146] B. Ripley, *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996.

[147] V. Vapnik, *Statistical Learning Theory*. Wiley-Interscience, 1998.

[148] C. Burgess, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[149] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[150] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

[151] L. Davis, *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold, 1991.

[152] Z. Michalewicz, *Genetic algorithms + Data Structures = Evolution Programs*. New York: Springer-Verlag, 1992.

[153] D. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[154] P. Chan and S. Stolfo, "Learning arbiter and combiner trees from partitioned data for scaling machine learning," *Proceedings of International Conference on Knowledge Discovery and Data Mining*, pp. 39–44, 1995.

[155] D. Lee, *A theory of classifier combination: the neural network approach*. Ph.D. Thesis, State University of New York at Buffalo, 1995.

[156] K. Ali and M. Pazzani, "Error reduction through learning multiple descriptions," *Machine Learning*, vol. 24, no. 3, pp. 173–202, September 1996.

[157] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226 –239, March 1998.

[158] J. Gama and P. Brazdil, "Cascade generalization," *Machine Learning*, vol. 41, no. 3, pp. 315–343, December 2000.

[159] L. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 281 –286, February 2002.

[160] T. Ho, J. Hull, and S. Srihari, "Decision combination in multiple classifier systems," *IEEE Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66–75, 1994.

[161] B. Zhang and S. Srihari, "Class-wise multi-classifier combination based on dempster-shafer theory," *Proceedings of the Seventh International Conference on Control, Automation, Robotics and Vision (ICARV 2002)*, pp. 698–703, December 2-5 2002.

[162] L. Xu, A. Krzyzak, and C.Suen, "Methods of combining multiple classifiers and their applications to handwritten recognition," *IEEE Transaction on Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.

[163] K. Ting and I. H. Witten, "Issues in stacked generalization," *Journal of Artificial Intelligence Research*, vol. 10, pp. 271–289, 1999.

[164] S. Gelfand, C. Ravishankar, and E. Delp, "An iterative growing and pruning algorithm for classification tree design," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 13, pp. 163–174, 1991.

[165] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.

[166] P. Somol, P. Pudil, J. Novovicova, and P. Paclik, "Adaptive floating search methods in feature selection," *Pattern Recognition Letters*, vol. 20, pp. 1157–1163, 1999.

[167] A. Jain and D. Zongker, "Feature-selection: Evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153–158, 1997.

[168] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *Machine Learning: Proceedings of the Eleventh International Conference*, pp. 121–129, Morgan Kaufmann, 1994.

[169] C. Blake and C. Merz, "UCI repository of machine learning databases," [http://www.ics.uci.edu/∼mlearn/MLRepository.html], University of California, Irvine, Dept. of Information and Computer Sciences, 1998.

[170] T. Ojala, J. Näppi, and O. Nevalainen, "Accurate segmentation of the breast region from digitized mammograms," *Computerized Medical Imaging and Graphics*, vol. 25, pp. 47–59, 2001.

[171] M. Abdel-Mottaleb, C. Carman, C. Hill, and S. Vafai, "Locating the boundary between the breast skin edge and the background in digitized mammograms," *Proceedings of the 3rd International Workshop on Digital Mammography*, pp. 467–470, June 9-12 1996.

[172] A. Mendez, P. Tahoces, M. Lado, M. Souto, J. Correa, and J. Vidal, "Automatic detection of breast border and nipple in digital mammograms," *Computer Methods and Programs in Biomedicine*, vol. 49, pp. 253–262, 1996.

[173] J. Heine, M. Kallergi, S. Chetelat, and L. Clarke, "Multiresolution wavelet approach for separating the breast region from the background in high resolution digital mammography," *Proceedings of the 4th International Workshop on Digital Mammography*, pp. 295–298, 1998.

[174] R. Chandrasekhar and Y. Attikiouzel, "Automatic breast border segmentation by background modelling and subtraction," *Proceedings of the 5th International Workshop on Digital Mammography*, pp. 560–565, June 11-14 2000.

[175] R. Ferrari, R. Rangayyan, J. Desautels, and A. Frere, "Segmentation of mammograms: Identification of the skin-air boundary, pectoral muscle, and fibroglandular disc," *Proceedings of the 5th International Workshop on Digital Mammography*, pp. 573–579, June 11-14 2000.

[176] M. Masek, Y. Attikiouzel, and C. DeSilva, "Automatic removal of high-intensity labels and noise from mammograms," *Proceedings of the 5th International Workshop on Digital Mammography*, pp. 587–592, June 11-14 2000.

[177] S. Kwok, R. Chandrasekhar, and Y. Attikiouzel, "Automatic pectoral muscle segmentation on mammograms by straight line estimation and cliff detection," *The Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001*, pp. 67 –72, Nov. 18-21 2001.

[178] B. Efron, "Bootstrap methods: Another look at the jackknife," *Annual Statistics*, vol. 7, pp. 1–26, 1979.

[179] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. New York: CRC Press, 1994.

[180] R. Shapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990.

[181] Y. Freund, "Boosting a weak learning algorithm by majority," *Information and Computation*, vol. 121, no. 2, pp. 256–285, 1995.

[182] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug., 1997.

[183] Y. Freund and R. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, no. 5, pp. 771–780, 1999.

VITA

VITA

Yajie Sun was born in Ningbo, P.R. China. In 1991, he was admitted into Shanghai Jiao Tong University exempted from the National Entrance Exams. He received his double Bachelor of Engineering (highest honors) degrees in July 1995, majoring in Biomedical Engineering and Instrumentation, and minoring in Industrial Management Engineering. He was honored twice as "Excellent Student in Shanghai Jiao Tong Univerity."

From 1995 to 1998, he was a research assistant for the late Professor Weixue Lu at Zhejiang University, where his research focused on heart model and simulation. He received his Master of Engineering degree in July 1998.

Since the Fall of 1998, he has been pursuing his doctoral degree in the School of Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana. He is a research assistant in the Video and Image Processing Laboratory (VIPER) under the supervision of Professor Edward J. Delp. His research interests include signal and image processing, medical imaging, pattern classification, data mining, and algorithms and modelling development.

During the summer of 2002, he worked as an intern at CADx Systems Inc, Beavercreek, Ohio. He developed a decision tree classifier which is being currently used in their commercial CAD products.