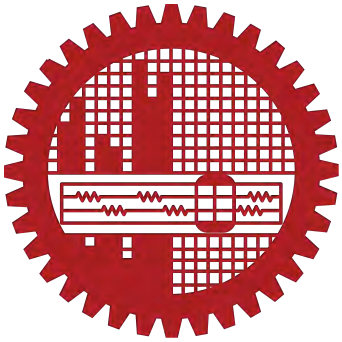


Breast Cancer Classification from Ultrasonic Images Based on Sparse Representation

by

Abdullah Al Helal

MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

2013

The thesis titled "**Breast Cancer Classification from Ultrasonic Images Based on Sparse Representation**", submitted by Abdullah Al Helal, Roll no: 0409062253, session April 2009, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Master of Science in Electrical and Electronic Engineering on July 24, 2013.

BOARD OF EXAMINERS

1. _____ Chairman
   Dr. Md. Saifur Rahman (Supervisor)
   Professor, Dept. of EEE
   Bangladesh University of Engineering and Technology, Dhaka-1000.

2. _____ Member
   Dr. Shaikh Anowarul Fattah
   Associate Professor, Dept. of EEE
   Bangladesh University of Engineering and Technology, Dhaka-1000.

3. _____ Member
   Dr. Pran Kanai Saha (Ex-Officio)
   Professor and Head, Dept. of EEE
   Bangladesh University of Engineering and Technology, Dhaka-1000.

4. _____ Member
   Dr. Mohammad Rakibul Islam (External)
   Professor, Dept. of EEE
   Islamic University of Technology, Gazipur-1704.

CANDIDATE'S DECLARATION

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

_____

Abdullah Al Helal

To my parents

Acknowledgments

First of all, I would like to express my earnest gratitude to my supervisor Professor Dr. Saifur Rahman for his tremendous support and guidance. He helped me a lot in every aspect of my work and guided me with proper directions whenever I sought one. I was exposed to the real wonderful and fascinating field of Biomedical Image Processing by his careful supervision.

I would also like to express my acknowledgment and heartfelt gratitude to Dr. S. Kaisar Alam from Riverside Research, New York, for his constant guidance, invaluable suggestions and perpetual support throughout the progress of my work. It would not be possible to get hold of highly qualified research resources without his kind help. His words of encouragement led me to the completion of this thesis. I also acknowledge his generosity in sharing a database of ultrasound images.

I would like to thank Dr. Khawza I. Ahmed from United International University, Dhaka, for his helpful discussions and sharing of ideas regarding this thesis. I would also like to thank him for his thoughts and inputs in improving the write-up of this thesis.

I appreciate the thoughtful comments from Dr. Shaikh Anowarul Fattah, and also thank him for giving significant amount of time commenting on my work. Some of the major revisions in this work have been possible due to his valuable suggestions.

In this regard, I remain ever grateful to my beloved parents who always exist as a source of inspiration behind every success I have ever made. I also remember my sister for her consistent support in my everyday living.

Abstract

This thesis presents a novel Sparse Representation-based Classifier (SRC) that provides superior performance in terms of high Area Under the ROC Curve (AUC) in classifying benign and malignant lesions of breasts captured in ultrasound images. Although such a classifier was initially proposed for face recognition, the use of this has been proposed in medical diagnosis from ultrasonic images in this dissertation for the first time. The classifier is based on ℓ_1-norm based sparse representation of a patient's test data in terms of linear combination of the features of the benign and malignant test lesions available in the training set. The proposed classifier uses an index called Sparsity Rank (SR) for the classification obtained from the normalized energy of the weights as a linear combination of the global sparse representation of the ultrasound images of the training set. The performance of the classifier is further enhanced to a great extent by two ways; first, by the use of a method that intelligently combines the features extracted from the multiple ultrasound scan of the same patient, and the second, by using the reduced feature set. The combining principle of the multiple data scans is based on averaging of the SRs of all the scans. The near-to-optimal feature set is obtained by a suboptimal strategy to evade the time exhaustive brute force approach that has a combinatorial search space.

With all the enhancements an AUC of 0.9754 has been achieved, when training and testing sets are chosen by leave-one-out approach from the data set.

Contents

List of Tables

List of Figures

List of Acronyms

SRC	Sparse Representation-based Classifier
SR	Sparsity Rank
CAD	Computer-Aided Diagnosis
BUS	Breast ultrasound
FDA	Food and Drug Administration
PPV	Positive Predictive Value
NPV	Negative Predictive Value
SVM	Support Vector Machine
RF	Radio Frequency
ACR	The American College of Radiology

BI-RADS	Breast Imaging Reporting and Data System
FNPA	4-neighborhood pixels algorithm
PCA	Principal Component Analysis
TPR	True Positive Rate
FPR	False Positive Rate
ROC	Receiver Operating Characteristic
AUC	Area Under the ROC Curve
ATL	Advanced Technology Laboratories
PMA	pre-market approval
SNR	Signal to Noise Ratio
ROI	Region of Interest
LDA	Linear Discriminate Analysis
LOGREG	Logistic Regression

Chapter 1

Introduction

1.1 Brief Description of the Problem

The average age of the human life has been extended to 80 years in the developed countries and around 60 years in the developing countries [1]. This is mainly due to the advancement of medical science and engineering. The major reasons for death have been shifted toward the cardiovascular diseases and cancers from the earlier reasons of airborne and direct-contact infectious diseases [1], [2]. Among various cancers, breast cancer is the second most lethal one among women after lung cancer [2]. However, early detection of it reduces the risk of death to a great extent.

Ultrasound image or mammogram of the breast can effectively be used for early detection of cancer by CAD. This thesis attempts to identify a breast tumor to have benign or malignant masses using a labeled database of ultrasound images.

1.2 The Global Burden of Breast Cancer

The global burden of cancer continues to increase both in developed and developing countries [3], [4]. The present world faces a death from cancer out of every eight deaths [2]. Breast cancer is the most common cancer among females [4], and the most lethal cancer only second to lung cancer [2]. Nevertheless, death rates for breast cancer have decreased steadily for the last 25 years.

The decrease in death from breast cancer has been possible due to, among various factors, early detection of the malignant neoplasms responsible for the cancer. Studies have shown that early detection through mammography can be attributed to saving lives by increasing options. However, being cost prohibitive, this approach is not feasible in most economically developing countries [5]. A recent survey of cancer survival rates indicates substantially lower survival rates in Africa, India, and the Philippines than in China, South Korea, and Singapore [4].

The traditional approach of diagnosing a tumor involves performing biopsies of a sample of the identified tumor. However, 65% to 85% biopsy operations turn out to be unnecessary due to the low specificity of mammography. It is even less sensitive for young women due to their radiologically dense breasts, though tumors tend to grow faster in younger estrogen-rich women, and statistics show that these women constitute 40% of all patients undergoing screening mammography. These unnecessary biopsies not only increase the cost, but also make the patients suffer from anxiety and put both the patients and the radiologists under risk of the radiation of mammography [6]. High health-care cost, minor risk and patient anxiety arisen from these unneeded biopsies warrant the development of alternative

technology and quantitative, reproducible and expertise-independent classification algorithms to supplement the biopsy results.

1.3 The Role of Breast Ultrasound

Breast ultrasound (BUS) has been proven superior to the mammography in the following respects:

1. Being not radioactive, ultrasound makes it safer and more convenient for patients for frequent screening than mammography.

2. Being cheaper, ultrasound is especially fit for the developing countries, allowing feasible treatment to a large number of patients.

3. BUS exhibits higher sensitivity for detecting tumors in dense breasts, making it more valuable than mammography for younger women.

4. BUS exhibits higher specificity implying lower False Positive Rate (FPR), consequently avoiding a large number of biopsies. The accuracy of BUS imaging in diagnosing a simple cyst can reach 90% to 100%.

All these facts have made Food and Drug Administration (FDA) recently approve the use of BUS imaging techniques in addition to mammography for women with dense breast tissue.

Conventional ultrasound images can display most breast cancers, including more than 90% of invasive ductal carcinomas. To improve the accuracy of diagnosis from BUS images, The American College of Radiology (ACR) developed a lexicon

of features named BI-RADS describing how the breast lesions appear in an ultrasonogram or a mammogram. The lexicon lists six possible outcomes, ranging from Category 1 to Category 6.

This thesis attempts to identify each breast lesion into either benign or malignant.

1.4 Ultrasound Imaging

Ultrasound imaging is a widely used modality for diagnostic applications. Currently, approximately 20 percent of all medical images are made with ultrasound, and the proportion is reported to be increasing [7]. Ultrasound is a type of mechanical wave named acoustic oscillation. The name arises from the operating frequency of ultrasound being beyond the upper limit of human hearing. The propagation of ultrasound through a medium is attributed to the oscillation of particles around their mean position in an elastic medium due to intra-particle forces, which can be described by the wave equation. As the ultrasound travels through human tissues, reflection occurs at the discontinuity of the mechanical characteristics. It is the variation among tissue types in terms of their ability to reflect ultrasound that makes it possible to interpret their presence and difference in the reflected signal. In other words, the variation among tissue types becomes visible in the received ultrasound image.

1.5 Use of Computer-Aided Diagnosis

Ultrasonography is much more operator-dependent than mammography, and requires interpretation of an expert radiologist, yet suffers from a high inter-observer variation rate. Computer-Aided Diagnosis (CAD) can play a supplementary role in this issue by helping radiologists in detecting and diagnosing breast cancer. The main motivation behind the use of CAD is to reduce or completely eliminate the dependency on the subjective judgement of the radiologist on the presence and the diagnosis of the breast lesion for its state of being benign or malignant. This can only happen when CAD becomes consistent in giving the correct diagnosis decision always. Technically, this implies that a classifier should have the zero false negative fraction, i.e., 100% True Positive Rate (TPR) or *sensitivity* in order to avoid any case that would mistakenly diagnose a cancerous lesion as benign to the effect that the patient eventually risks his life for such wrong diagnosis. Table 1.1 presents different parameters used in quantifying the performance of a classifier used for medical diagnosis.

1.6 Performance Criterion of a Classifier

Improving TPR is as important as improving FPR from a scientific point of view. Usually attempting to increase TPR results in an increase of FPR, implying that TPR can be traded off for FPR. A Receiver Operating Characteristic (ROC) curve succinctly represents the simultaneous variation of TPR and FPR in the same plot. Figure 1.1 shows several ROC curves for different hypothetical classifiers. For

Table 1.1: Highlights of the various parameters associated with the performance of a CAD for classification of lesions into benign and malignant masses.

		True Condition	
		Malignant (Pos)	Benign (Neg)
Test Outcome	Malignant (Y)	TP	FP
	Benign (N)	FN	TN

where

Condition Positive	$= Pos$
Condition Negative	$= Neg$
True Positive	$= TP$
False Positive	$= FP$
True Positive Rate (TPR) or Sensitivity	$= \frac{TP}{Pos}$
False Positive Rate (FPR) or 1− Specificity	$= \frac{FP}{Neg}$
Accuracy	$= \frac{TP + TN}{Pos + Neg}$
Positive Predictive Value (PPV)	$= \frac{TP}{Y}$
Negative Predictive Value (NPV)	$= \frac{TN}{N}$
Precision	$= \frac{TP}{Y}$
Recall	$= \frac{TP}{Pos}$
F-measure	$= \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}$

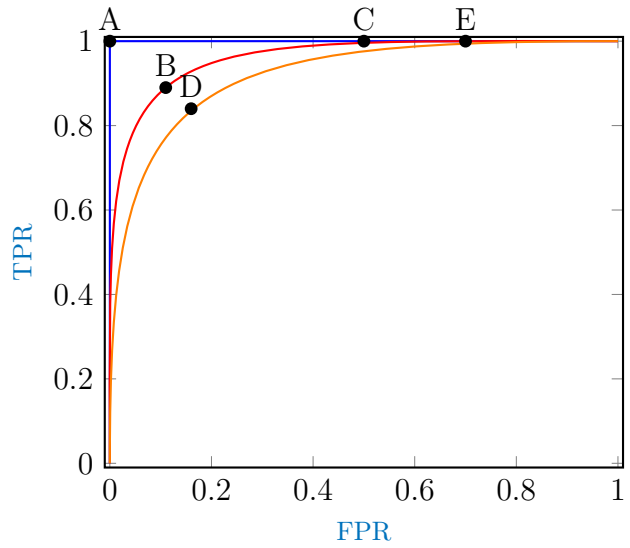


Figure 1.1: Selection of operating points in ROC of a realistic classifier used in medical diagnosis: 'A' is a point in an ideal classifier, whereas 'B', 'C', 'D' and 'E' are the points in realistic classifiers.

an ideal classifier the ROC curve has 100% Area Under the ROC Curve (AUC). However, a real classifier's AUC falls below that. One logical question arises where the operating point on ROC should be chosen. This depends on the application, e.g., medical diagnosis or radar detection, related with the ROC. In cancer diagnosis in the breast, the operating point needs to be chosen so that TPR or *sensitivity* is 100% even at the cost of high FPR or $1 -$ specificity and thus the overall accuracy. However, such high value of FPR implies false cases diagnosed as cancerous and this ultimately would require the biopsy procedure to rule out if the case is false. Therefore, to reduce the cost of biopsy a good classifier needs to have an operating point that would rule out any false negative case as well as false positive case. In other words, the *accuracy* of the classifier needs to be close to 100%. In the presence of such classifier only the cases diagnosed as malignant would be forwarded for the

biopsy. Thus, such reduction of the number of biopsy will reduce the monitory involvement and the use of the other resources of the patients and the relevant authorities. For a realistic classifier employed for cancer diagnostic, it needs to operate at the point of ROC curve where sensitivity is 100% and $1 -$ specificity is minimum.

The point 'A' in Figure 1.1 denotes the operating point of an ideal classifier that represents 100% TPR with 0% FPR, i.e., 100% accuracy, thus lies on an ideal ROC curve with a perfect AUC of 1. A random classifier, i.e., one that classifies randomly, would form a diagonal line as a ROC curve with an AUC of 0.5. Points 'B', 'C', 'D' and 'E' shows the operating points for more realistic classifiers. While points 'B' and 'D' are the operating points that may correspond to the highest accuracy rate, points 'C' and 'E' need to be chosen as operating points for the realistic classifiers that correspond to 100% TPR with the minimum FPR. The closer the point 'C' or 'E' towards 'A', the better the classifier. For example, point 'C' lies on a better classifier compared to point 'E'. In this work, an enhanced classifier based on Sparse Representation-based Classifier (SRC) is presented that provides us near-to-ideal operating points in the ROC curve with an AUC of 0.9754.

1.7 Challenges

A basic difference between the problem of classifying a biometric class, e.g., face recognition, and that of classifying a disease group, e.g., tumor classification, is the nature of the classes themselves. For example, each class in face recognition refers to a particular person, whose data samples, i.e., images inherently contain a good

deal of overlapping information resulting in redundancy. On the contrary, each class in tumor classification refers to a particular tumor type, but not a particular patient or even a particular tumor, thus de-emphasizing the effect of inclusion of redundancy. Specifically, a test sample in face recognition is ideally to be matched against the training samples of the same person, whilst a test sample in tumor classification is ideally to be matched against the training samples of the same type rather than the same tumor. Repeated scanning of the same lesion may enable us to exploit the inherent redundancy of the data by providing a complementary view.

Furthermore, ultrasonogram itself is a noisy process in that it not only measures tissue variations but also includes a portion of noisy signals due to the fact that, unlike other popular medical imaging modalities, ultrasonic images are degraded by speckle noise [7]. For example, for fully developed speckle, scatter density is greater than 5 per wavelength, resulting in a very low value of Signal to Noise Ratio (SNR) of 1.91.

1.8 State of the Art

Since CAD tools allow for investigating abnormal features that may otherwise be invisible on the screening image, several groups have worked on developing CAD tools to aid in automating the diagnosis of tumor lesions.

Tan et. al. worked on classifying breast tumors from Automated 3-D Breast Ultrasound Imaging instead of the traditional 2-D ultrasound, since spiculation, an

important characteristic of malignant tumors, is either absent or not clearly visible in 2-D ultrasound, but captured in coronal planes parallel to the transducer. They applied Support Vector Machine (SVM) by selecting a radial basis function as the kernel to classify a labeled database of 88 patients. With all features included, the resulting ROC gives an AUC of 0.93 [8].

Shankar et. al. utilized statistical parameters to diagnose breast tumors from ultrasound images. Non-Rayleigh distribution of the envelope of the backscattered echo from ultrasonogram had been shown to be useful in characterizing breast tissue, and Nakagami distribution and its statistics had been successfully applied to classify breast tumors. Shankar et. al. implemented a multi-parameter approach to involve these statistical features, and developed a discriminant to combine them linearly. Classification using this discriminant results over a database of 99 patients in AUC of 0.94 [9].

Alam et. al. employed a multi-parameter approach by developing quantitative descriptors of the suggested BI-RADS features which had been reported to improve the accuracy of diagnosing breast tumors. Their work used a database of 502 tumor scans, among which many of the scans correspond to the same tumor. They suggested a reduced feature set, and applied Logistic Regression (LOGREG) as a classifier to come up with an AUC of 0.95 [10].

Joo et. al. utilized morphological features of the ultrasonograms of the breast nodules, and implemented an artificial neural network technique in the form of a general multilayer perceptron with the back propagation learning rule for training the classifier over a database of 584 patients. The performance of the classifier was

evaluated by testing over a database of 266 patients, and the AUC was found to be 0.95 [11].

Huang and Chen utilized interpixel texture features and implemented SVM with the radial kernels to classify breast tumors from ultrasound images. They have used two databases: one with 250 patients, resulting in an AUC of 0.96, and the other one with 140 patients, resulting in a somewhat better AUC of 0.97 [12].

Many of the work did not involve any ROC analysis, and reported the performance of classification using some other performance criteria.

For example, Cheng et. al. used fractal dimension of the breast tumors as a measure of texture by using the fractal Brownian motion, followed by an application of morphological operations. They implemented the k-means classification method over a database of 250 patients, reporting a PPV of 89.74 and a NPV of 96.24 [13].

Yanjiao et. al. used the lobulated contour feature obtained by implementing a boundary-based corner count to classify breast tumors. They applied SVM with radial based function as the kernel over a database of 240 patients, obtaining an accuracy of 0.95 [14].

Bocchi et. al. worked on classifying breast tumors from ultrasound video, so that each frame of the video describes the underlying tumor from a different point of view. They used shape features of the tumor to classify tumors. Each frame was processed independently for classification, and the obtained results were integrated to result in a more reliable classification with an accuracy up to 0.97 [15].

This list is not exhaustive. A good number of work did not consider the effect of ultrasound scanning system, hence did not compensate for the contribution of the system.

1.9 Motivation behind the Proposed Technique

Sparse representation is a recently developed classification technique based on the non-conventional compressive sensing principle. This representation falls into the category of reconstructive methods. Different reconstructive and discriminative methods have been implemented in the literature for breast cancer classification. While discriminative methods aim at maximizing the separation of signal classes, and thus are, in general, capable of classifying ideal signals more efficiently than reconstructive methods [16, 17], their performance degrades in presence of noise and outliers due to their nonrobust nature [18]. Sparse representation is shown to achieve state-of-the-art performance in image denoising and face recognition [19]. The fact that ultrasound is inherently a noisy process motivates us to investigate sparse representation in classifying cancer images.

1.10 Objectives and Scope of the Thesis

The objectives of the research with specific aims are as follows:

1. To study ultrasonic imaging technique and medical features recommended in BI-RADS for classification and search for an optimal set of features to aid in classification.

2. To study the recent sparse representation technique in order to judge its merit in classifying ultrasonic images of cancer lesions.

3. To study the information theoretic view of redundancy in data so as to exploit the presence of redundancy in repeated ultrasonic scans of the same tumor by exploiting an appropriate setup of scans.

The outcomes are listed as follows:

1. A subset of standard features is recommended which is capable of giving better result in classification than the available ones.

2. The sparse representation is employed to classify benign and malignant tumors from ultrasonic images more accurately.

3. A setup is proposed to include redundancy by taking multiple scans of the same lesion, and a modified classifier is chosen which will harness the redundancy for better classification.

1.11 Outline of the Thesis

A classifier based on sparse representation has been successfully used in face recognition [20]. The framework of the methodology includes implementation of such sparse representation in classifying cancer lesions from labeled ultrasonic

images. Raw labeled data in `.eye` format are collected from Riverside Research, New York. The raw images are converted to matrix form using a `.eye` reader and MATLAB routines. Selected features are considered for classification and an appropriate weight is assigned to each feature before classification. Redundancy is also incorporated and the improvement in classification results is observed. Finally, the findings are matched against the labeled database. A comparative study of the performance of the present research with that of the other work is also demonstrated.

1.12 Contribution of the Thesis

The prime contributions of the current research are as follows. Sparse Representation-based Classifier (SRC) as a classifier is introduced for the first time in the problem of medical diagnosis from ultrasonograms. The performance of Sparse Representation-based Classifier (SRC) in face recognition has proven to yield high accuracy [20], where the test image can usually be expressed as the linear combination of one of the classes, i.e., persons, in the database and thus sparsity lies in all the other classes with respect to the matched class. However, in classification of tumor, as explained earlier, it may not always be possible to express the test tumor of unknown class in terms of the linear combination of the training tumors of one of the two classes. This may result in non-sparsity and eventual failure of the correct detection. Also, the degree of sparsity may become poor as only two classes, i.e., benign and malignant are considered. This limitation of SRC is turned into blessings when the information of multiple scans of the same

tumors is combined by the averaging of the Sparsity Rank (SR) during testing and afterwards the decision is made comparing the average SR with the set threshold in the classifier.

To the best of the author's knowledge, none has so far reported a combining techniques of multiple scans that results in superior performance, i.e., AUC as high as 0.9754, although there are reports where the features available from multiple scans of the same tumor are averaged before being used during training and testing [10]. To improve the performance of SRC further, an optimization strategy for the feature set to be used in the classifier is proposed that searches in a reduced space compared to the computationally prohibitive combinatorial search space.

1.13 Organization of the Thesis

The rest of the thesis is organized as follows. This chapter describes the premise of diagnosing breast cancer and available tools for its early detection. Chapter 2 develops the proposed technique in classifying breast tumors. Chapter 3 deals with the implementation steps of the experiments. The results are explained and discussed in Chapter 4. Finally, Chapter 5 summarizes the outcome of the current research and concludes with suggestions for future work.

In this chapter, we have discussed the burden of breast cancer, and how ultrasound imaging can play a role in its early detection as a better alternative to mammography by obviating unnecessary biopsies and thus potentially provide a

feasible treatment to the economically developing countries. The significance of CAD and available CAD tools in classifying breast tumors have also been discussed.

The next chapter describes the proposed technique of this thesis, and the logical walk-through to develop the same.

Chapter 2

The Proposed Sparse Representation Technique for Breast Cancer Classification

Having established the significance of developing a CAD tool to aid to diagnose breast cancer from ultrasound images, we will now discuss how the sparse representation can be employed to adjust to the challenges of classifying tumors from ultrasound images.

2.1 Problem Statement

The task of tumor classification translates to the problem of correctly assigning every test tumor sample to its benign or malignant status using labeled training

tumor samples. In other words, given a large enough ultrasound database of breast tumors labeled as benign or malignant, we want to label any new sample into one of the given classes. The working database includes Radio Frequency (RF) ultrasonogram scans of patients. Each image in the database is denoted by $\mathcal{I} = (\mathbf{P}, R)$ where $\mathbf{P}$ is a two-dimensional (2D) grid of pixels (x, y) and $R(x, y) : (x, y) \in \mathbf{P}$ is the image intensity function that corresponds to the RF signal received through ultrasonogram scans. Two typical ultrasound images $\mathcal{I}$ are shown in Figure 3.1. The lesion $\mathbf{L} \subset \mathbf{P}$, visible in the RF image, is traced and separated using a trained human guided boundary tracing program, so that $\mathbf{L}$ defines the set of all pixels within the boundary of the lesion.

Thus, we will identify an ultrasonogram tumor scan $\mathcal{L} = (\mathbf{L}, R)$ into one of the N_c classes based on the available training samples. Two typical breast tumors $\mathcal{L}$ are shown in Figure 3.2. In this work, the tumor has been identified to have either benign or malignant masses. This implies that $N_c = 2$. We could say, for example, benign tumors form class 1 and malignant tumors, class 2.

2.2 Outline of Methodology

The salient steps of the proposed method of classification are outlined in Figure 2.1. The steps are explained in Sections 2.3 to 2.7.

2.3 Feature Extraction

ACR has developed a host of features listed in BI-RADS [21], [22]. In this work, quantitative descriptors have been used to simulate the BI-RADS features so as to make our evaluation reproducible and to reduce the inter-observer discrepancy. A complete list of the descriptors used in this work can be found in Section 3.4.

Apart from using the BI-RADS features, other feature extraction techniques such as Principal Component Analysis (PCA) and Linear Discriminate Analysis (LDA), and downsampled pixel values could also be used.

2.4 Test Samples as Sparse Linear Combinations of the Training Samples

Having generated the feature set, we can now interpret any test tumor sample $(\mathbf{L}, R)$ as projected on a *feature space* represented by the *feature vector*

$$\mathbf{b} = [b^{(1)}, b^{(2)}, \ldots, b^{(m)}]^\intercal \in \mathbb{R}^m \tag{2.1}$$

in an m-dimensional space, where m is the feature dimension, and each element $b^{(i)}$ of the vector $\mathbf{b}$ represents a feature. Thus

$$\mathbf{b} = \begin{bmatrix} b^{(1)} \\ b^{(2)} \\ \vdots \\ b^{(m)} \end{bmatrix} \tag{2.2}$$

is an m-directional vector corresponding to a tumor, and each $b^{(i)}$ is a scalar corresponding to a feature of the tumor.[1]

A simple but efficient model is incorporated in the current research which assumes that the samples from a single class lie on a linear subspace. Subspace models contain enough flexibility to represent much of the variation of a real data set. For example, images of a particular face, projected in a linear subspace, have been shown to achieve lower recognition error rates [16].

If we have n_i number of training samples from class i, the total matrix for the training samples of class i can arranged as follows:

$$\mathbf{A}_i = [\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, \ldots, \mathbf{b}_{i,n_i}] \tag{2.3}$$

where $\mathbf{b}_{i,j}$ is the m-dimensional feature vector of the j-th training sample from class i.

Assuming any test sample with feature vector $\mathbf{b}$ from this class[2] as the linear combination, i.e., weighted sum, of the feature vectors of the training samples $\mathbf{b}_{i,j}$, $\mathbf{b}$ can be written as

$$\mathbf{b} = a_{i,1}\mathbf{b}_{i,1} + a_{i,2}\mathbf{b}_{i,2} + \cdots + a_{i,n_i}\mathbf{b}_{i,n_i} \tag{2.4}$$

where $a_{i,j}$, $j = 1, 2, \ldots, n_i$, are real coefficients.

[1]For example, if we go for BI-RADS features according to Section 3.4, $b^{(1)}$ corresponds to echogenecity, is equal to -28.4 for the 1st tumor sample, and is found to vary from -54.1 to -12.6 for all tumor samples.

[2]Throughout this thesis, $\mathbf{b}$ without subscript represents sample from the testing set, and $\mathbf{b_i}$ or $\mathbf{b_{i,j}}$ with subscript, from the training set.

As the membership i of the test sample is not known during training, we could use a big fat matrix consisting of feature vectors of all n training samples from all classes, as follows:

$$\mathbf{A} = [\mathbf{A}_1|\mathbf{A}_2|\dots|\mathbf{A}_{N_c}] \tag{2.5}$$

with columns $\mathbf{b}_\lambda$, where $\{\, \lambda \in \Lambda_i \,\}$ is a set of indices that corresponds to training samples $\mathbf{A}_i$ from class i, enabling us to write the equation set as follows:

$$\mathbf{b} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n \tag{2.6}$$

with only n_i non-zero entries a_λ, $\lambda \in \Lambda_i$ associated with class i.

To explain Equation (2.6) further, let us take an example of a test tumor. Let us say that we have a test tumor, whose m features have to be related to the weighted sum of all the features of the training tumor samples:

$$\begin{bmatrix} b^{(1)} \\ b^{(2)} \\ \vdots \\ b^{(m)} \end{bmatrix} = a_1 \begin{bmatrix} b_1^{(1)} \\ b_1^{(2)} \\ \vdots \\ b_1^{(m)} \end{bmatrix} + a_2 \begin{bmatrix} b_2^{(1)} \\ b_2^{(2)} \\ \vdots \\ b_2^{(m)} \end{bmatrix} + \cdots + a_n \begin{bmatrix} b_n^{(1)} \\ b_n^{(2)} \\ \vdots \\ b_n^{(m)} \end{bmatrix} \tag{2.7}$$

so that Equation (2.6) follows.

If the right side of Equation (2.7) is rearranged, $\mathbf{b}$ may be considered as the product of a matrix and a vector as follows:

$$\begin{bmatrix} b^{(1)} \\ b^{(2)} \\ \vdots \\ b^{(m)} \end{bmatrix} = \begin{bmatrix} b_1^{(1)} & b_2^{(1)} & \dots & b_n^{(1)} \\ b_1^{(2)} & b_2^{(2)} & \dots & b_n^{(2)} \\ \vdots & \vdots & \dots & \vdots \\ b_1^{(m)} & b_2^{(m)} & \dots & b_n^{(m)} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{2.8}$$

or,

$$\mathbf{b} = \mathbf{A}\mathbf{x}_r \tag{2.9}$$

where

$$\mathbf{x}_r = [a_1, a_2, \ldots, a_n]^\intercal \tag{2.10}$$

with only n_i non-zero entries a_λ, $\lambda \in \Lambda_i$ associated with class i. Equation (2.6) represents any test sample as a linear combination of the training samples. The task of classification dictates that we need to find the coefficients a_λ (which, in turns, demands to find $\mathbf{x}_r$), since identifying the location of non-zero entries translates to detecting the native class of the tumor.

What makes the apparently obvious and standard approach of solving the linear system in Equation (2.9) difficult is that the solution is not unique dictated by the fact that m and n need not be equal due to the independence between the training sample size and the number of features extracted from the samples.

If $m > n$, i.e., the number of features surpasses the number of training samples, the system is overdetermined, there may or may not be a unique solution. However, considering the usual simplified modeling and measurement error, methods exist to find an 'approximate' solution to the system, and these methods usually find the correct solution $\mathbf{x}_r$.

On the other hand, if $m < n$, the system is underdetermined and there is no unique solution. To make things worse, test samples, when corrupted with noise, may not satisfy Equation (2.9) even in case of an overdetermined system. The problem of non-uniqueness of the solution can be dealt with looking for the minimum ℓ_2-norm

solution:

$$\mathbf{x}_2 = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_2 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{2.11}$$

where $\|\cdot\|_2$ is the ℓ_2-norm, which can also be viewed as the square root of the energy of the vector:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_i a_i^2} \tag{2.12}$$

Not only that this solution does not equal the correct solution $\mathbf{x}_r$, it generally is dense in that it contains non-zero entries in almost all coordinates, making it ineffective to provide information about the sample's identity.

The representation $\mathbf{x}_r$, as represented in Equation (2.10), is *sparse* in the sense that (at most) n_i entries are non-zero, when the test sample is from class i. Thus the solution vector $\mathbf{x}_r$ is a *sparse representation* of a test sample in terms of the training samples. Intuitively, the more terms will be zero, the more accurately a test sample can be associated with its native class, which inspires us to seek for the 'sparsest' solution, i.e., a solution with the maximum number of zero coefficients, or alternatively, the minimum number of non-zero coefficients:

$$\mathbf{x}_0 = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.13}$$

where $\|\cdot\|_0$ is the ℓ_0-norm that counts the non-zero elements of a vector:

$$\|\mathbf{x}\|_0 = \#\{\, i | a_i \neq 0 \,\} \tag{2.14}$$

It has been shown that if $\mathbf{x_r}$ is sparse enough, it is the unique spartest solution of Equation (2.13): $\mathbf{x_0} = \mathbf{x_r}$ [23].

Though theoretically this sounds good, the fact that solving Equation (2.13) requires combinatorial search [24] warrants a practically feasible algorithm or method to solve the same. It has been shown that if $\mathbf{x}_r$ is sparse enough, it can be solved for by implementing a totally different program called ℓ_1-minimization

$$\mathbf{x}_1 = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.15}$$

where $\|\cdot\|_1$ is the ℓ_1-norm that sums the elements of a vector,

$$\|\mathbf{x}\|_1 = \sum_i |a_i| \tag{2.16}$$

making the ℓ_1-norm solution equal to the ℓ_0-norm solution [25]. The program ℓ_1-minimization is found to recover $\mathbf{x_r}$, thus choosing $\mathbf{x}_1$ equal to $\mathbf{x_r}$, with overwhelming probability.

If $\mathbf{x}_r$ is not sparse enough, it has been shown that the recovery error $\mathbf{x_r} - \mathbf{x}$ for $\mathbf{x}_r$ is given by

$$\|\mathbf{x_r} - \mathbf{x}\|_2 \leq C_1 \frac{\|\mathbf{x_r} - \mathbf{x}_K\|_1}{\sqrt{K}} \tag{2.17}$$

for $K \ll m$, where $\mathbf{x}_K$ is the vector $\mathbf{x_r}$ with all but the largest K coefficients set to zero [26].

2.5 Ranking and Classification of the Test Samples

Ideally, the sparse representation of a test sample will constitute of non-zero values only at the indices of a single class. For example, Figure 2.2 displays an ideal sparse representation of a benign test sample found through ℓ_1-minimization where

samples from index 1 to index 402 correspond to the actual class and other indices correspond to the wrong class. It is noteworthy that samples from the wrong class possess zero coefficients.

The single class with non-zero coefficients can then be identified as the true class of the test sample as follows:

$$\text{detection}(\mathbf{b}) = \arg\,\text{any}\{\,\lambda \in \Lambda_i \colon a_\lambda \neq 0\,\}. \tag{2.18}$$

In practical cases, however, noise and measurement errors may cause small but non-zero coefficients associated with the wrong class. Figure 2.3 displays a typical sparse representation of a benign test sample where samples from index 1 to index 402 correspond to the actual class and other indices correspond to the wrong class. It is obvious that samples from the wrong class possess small but non-zero coefficients.

We could design a classification scheme engaging the global sparse representation that can take advantage of the subspace structure of each class. For example, ℓ_1-norm of the values of a single class can provide information about how the whole class contributes to the solution. The problem with this choice is that ℓ_1-norm does not discriminate between small (hence not significant) values over large (hence significant) values much as norms of higher order (or any of their strictly monotonic functions), e.g., ℓ_2-norm, do.

In a multiple-class scenario, to get rid of this problem, we may introduce an index which assigns a score or rank to each solution measuring how close it is to be a 'perfect' member of class i as $\text{index}_i(\mathbf{b}) = \frac{\|\{\,a_\lambda \colon \lambda \in \Lambda_i\,\}\|_2}{\|\mathbf{x}\|_2}$, where the normalization is required when the classes are of different sizes. Having done this, we can detect

the class with the highest index. However, when applied in a binary-class scenario, this does not have any option to adjust TPR and FPR as per need, thus cannot allow for ROC analysis.

Thus, to deal with binary classification, as the present research is concerned with, we introduce a ranking function called Sparsity Rank (SR) (or simply, rank), which assigns a score or rank to each solution measuring how close it is to be a 'perfect' member of *class* 2 as follows:[3]

$$\text{rank}(\mathbf{b}) = \frac{\|\{\, a_\lambda \colon \lambda \in \Lambda_2 \,\}\|_2}{\|\mathbf{x}\|_2}. \tag{2.19}$$

The normalization introduced in Equation (2.19) enables us to cancel out the effect of non-uniform class sizes, i.e., when class 1 and class 2 do not have equal number of samples:

$$\#\{\, \mathbf{b}_\lambda \colon \lambda \in \Lambda_1 \,\} \neq \#\{\, \mathbf{b}_\lambda \colon \lambda \in \Lambda_2 \,\} \tag{2.20}$$

Since the formulation of rank requires having the knowledge of Λ_i, thus the labels of the training samples, the proposed technique can be considered as a supervised classification.

Ideally all class-1 samples will have a rank of 0, and all class-2 samples, 1. In practical cases, recovery error will occur, which can be related to rank using

$$\mathbf{x_r} - \mathbf{x} = \begin{cases} \{\, a_\lambda \colon \lambda \in \Lambda_2 \,\} & \text{for class 1} \\ \{\, a_\lambda \colon \lambda \in \Lambda_1 \,\} & \text{for class 2} \end{cases} \tag{2.21}$$

[3]We could have alternatively defined rank with class 1 involved without loss of generality; this would require altering the decision sign from $>$ to $\leq$ in Equation (2.22).

so that a rank will identify a sample through its proximity to one of the two extremities, and it would be enough to compare the rank of a sample with a threshold rank_{th} to determine its class:

$$\begin{aligned} &\text{if } \text{rank}(\mathbf{b}) > \text{rank}_{\text{th}} \\ &\quad \text{detection}(\mathbf{b}) = 2, \\ &\text{else } \text{detection}(\mathbf{b}) = 1. \end{aligned} \tag{2.22}$$

The performance of a classifier for a particular threshold can be measured by a paired value of TPR and FPR.

If we agree to label class-2 samples as *positives*, so that class-1 samples are *negatives*, we may define the relevant terms as follows:

$$\text{positive}, Pos = \#\{\, \mathbf{b}_\lambda \colon \lambda \in \Lambda_2 \,\} \tag{2.23a}$$

$$\text{negative}, Neg = \#\{\, \mathbf{b}_\lambda \colon \lambda \in \Lambda_1 \,\} \tag{2.23b}$$

$$\text{true positive}, TP = \#\{\, \text{detection}(\mathbf{b}_\lambda) = 2 \colon \lambda \in \Lambda_2 \,\} \tag{2.23c}$$

$$\text{false positive}, FP = \#\{\, \text{detection}(\mathbf{b}_\lambda) = 2 \colon \lambda \in \Lambda_1 \,\} \tag{2.23d}$$

$$\text{true negative}, TN = Neg - FP \tag{2.23e}$$

$$\text{false negative}, FN = Pos - TP \tag{2.23f}$$

$$TPR = \frac{TP}{Pos} \tag{2.23g}$$

$$FPR = \frac{FP}{Neg} \tag{2.23h}$$

$$\text{accuracy} = \frac{TP + TN}{Pos + Neg} \tag{2.23i}$$

Figure 2.4 shows the spread of the rank of each class for a typical run of the classification method. For better visualization, the distribution of each class is also shown in the form of histograms.

One interesting thing about ranking the samples (as opposed to classifying outright) is that not all class-2 samples have to precede a predetermined threshold value rank_{th}, rather it is only necessary that all class-2 samples precede all class-1 samples, so that a suitable threshold can be chosen, which separates the two classes.

A straight-forward implication of using ranking is that we can vary the threshold which essentially selects the operating point of classification. Flexibility of choosing the operating point, e.g., for a particular TPR is desired in some biomedical applications. For instance, legal issues about diagnosing a patient dictates that a radiologist/oncologist/physician would prefer a classification scheme with a high TPR over the one with a low TPR, even at the the price of a high FPR.

Varying the threshold over all possible values of rank gives us $(FPR_\lambda, TPR_\lambda)$ for each b_λ, allowing us for cross validation. The set

$$\{\,(FPR_\lambda, TPR_\lambda)\,\}, \tag{2.24}$$

when plotted on an xy plane, is called the ROC curve. While each point on the curve corresponds to an operating point accuracy (or other measures), the overall performance of the classifier can thus be measured by the AUC.

For an ideal classifier, all class-2 samples precede all class-1 samples, so that $TPR_\lambda = 1$ or $FPR_\lambda = 0$ for each λ, giving an AUC of 1. Thus a classifier is said to be better than a second one if its AUC is closer to 1, hence greater than that of the second one.

2.6 Robustness of the Classifier and Exploitation of Redundancy

From the point of view of information theory, redundancy in a piece of information can be used to recover it more efficiently or detection any corruption, and even correct any potential corruption. This motivates us to build a database by including redundant labeled data, which can be exploited for better classification. In the context of ultrasonogram, we can scan a tumor several times, e.g., from different angles. Multiple images taken by varying the beam angle are used in speckle reduction, since real data are unaffected by changes in beam angle, while speckle patterns are affected, and frame-averaging results in reduced speckles [7]. In a face recognition context, this can refer to taking photos of the same person with different facial expression and lighting condition.

The key point of including redundancy is that the internal structure of data be the same. Of course, it would carry no further information if the exact data is obtained several times. If we consider measurement errors, measurement of the same data will give different measures, but essentially the internal structure will be the same, up to a noise margin. This may be represented as follows:

$$\mathbf{b} = \mathbf{A}\mathbf{x}_r + \mathbf{e}. \tag{2.25}$$

Here, $\mathbf{e}$ represents the noise contributing to a measurement error. Furthermore, ultrasonogram itself being a noisy process in that it captures speckle noise, even with no measurement errors, ultrasound scans will result in corruption with noise.

It has been shown for noisy cases that a measurement error is attributed to the expression of recovery error [26] as follows:

$$\|\mathbf{x_r} - \mathbf{x}\|_2 \leq C_1 \frac{\|\mathbf{x_r} - \mathbf{x}_K\|_1}{\sqrt{K}} + C_2\mathbf{e} \tag{2.26}$$

Thus the measurement error $C_2\mathbf{e}$ is proportional to the noise. This makes sparse representation a robust method in that a small noise would cause a small perturbation in the representation. This also proves SRC to be one of the best choices to classify the noisy ultrasonograms.

Moreover, the linearity of the measurement error with the noise motivates us to device a mechanism to act on the measurement error, which will be propagated to the source of the noise to reduce it as well.

Assuming that the noise involved in Equation (2.25) is zero-mean, it follows from Equation (2.26) that

$$\begin{aligned} E[\|\mathbf{x_r} - \mathbf{x}\|_2] &\leq E\left[C_1 \frac{\|\mathbf{x_r} - \mathbf{x}_K\|_1}{\sqrt{K}}\right] + E[C_2\mathbf{e}] \\ &= C_1 \frac{\|\mathbf{x_r} - \mathbf{x}_K\|_1}{\sqrt{K}} \end{aligned} \tag{2.27}$$

so that the expectation of the recovery error is as good as the recovery error of a noiseless case as in Equation (2.17).

If we define the scans of the same tumor as a *group*, as opposed to each scan which is defined as a *sample*, the direct relation between the rank and the recovery error by means of Equation (2.21) enables us to estimate the left side of Equation (2.27) through *Group Sparsity Rank*, or simply group rank as follows:

$$\text{rank}_{\text{group}}(\mathbf{b}^\theta) = \mathcal{A}[\text{rank}(\mathbf{b}^{\theta_j})] \tag{2.28}$$

where $\{ \mathbf{b}^{\theta_j} \}$, $j = 1, 2, \ldots, N^\theta$ belong to the samples or 'realizations' of θ-th group out of n^θ groups, and $\mathcal{A}[\cdot]$ indicates averaging (over j).

The group sparsity rank is optimum in the sense that it suppresses the measurement error completely, hereby minimizing the recovery error.

2.7 Reduction of Feature Set

It should be noted here that the projection of data in lower subspaces can lead to improved classification, which warrants the search for appropriate feature extraction. This is related to the principle of shortest data description in model selection which specifies the most compact representation of a model to be preferred for classification [27]. In the ultrasonic imaging context, a good portion of an ultrasonic scan contains low details, thus making it hard for the classifier to extract discriminatory information from the samples of different classes to separate them. One additional benefit of extracting feature from the samples is the reduced computational cost associated with data of low-dimension.

The BI-RADS dictionary suggested by ACR describes a set of features of ultrasound images of breast tumors, which can potentially improve the accuracy of classification, hence diagnosis. This work implements quantitative descriptors of the mentioned features, which is traditionally assessed subjectively by physicians, to remove dependence on the operator and the expertise of the physicians.

To contrast the relative significance of the BI-RADS features against one another, an exhaustive search can be employed, which attempts to find the most significant

features contributing to the correct separation of classes. In particular, for $\mathbf{b} = \{ f_k \}$, $k = 1, 2, \ldots, m$, every possible subspaces of $\mathbf{b}$, i.e., every possible subsets of $\{ f_k \}$ can be substituted for $\mathbf{b}$ to search for the subspace that is responsible for the maximum AUC. This would be the optimum result considering the inclusion or exclusion of each feature in the feature space. If the computational cost associated with the exhaustive search is impractical, as found to be the case with the current work, a suboptimal search can be implemented which seeks recursively for a reduced feature set by initializing from $\{ f_k \}$ and deciding upon exclusion of each exclusion of each feature if the exclusion improves the resulting AUC and then going for next feature until the result reaches a saturation. The outcome will be suboptimal since choice of the sequence of features will affect the result. Consequently, the optimum solution may not be reached.

2.8 Classification of a Sample with Unknown Diagnosis

Once the classifier is designed, given a new tumor sample with unknown diagnosis, classifying this sample involves extracting the reduced feature set determined earlier, and then generating the Group Sparsity Rank of this sample. Comparing it with a pre-determined threshold will result in diagnosis of the sample identifying it as benign or malignant. Choosing the threshold depends on the operating point in the ROC curve, and thus, e.g., may be chosen by the radiologist.

Figure 2.5 depicts the procedure of classifying a tumor sample whose diagnosis is yet to be determined.

We have used the inherent sparse representation model of the test samples to classify tumors, and introduced a rank to facilitate ROC analysis. Information has been extracted from duplicate scans, and selection of a reduced feature set has been discussed.

The following chapter will describe how to implement the proposed technique to run an experiment over the available data set.

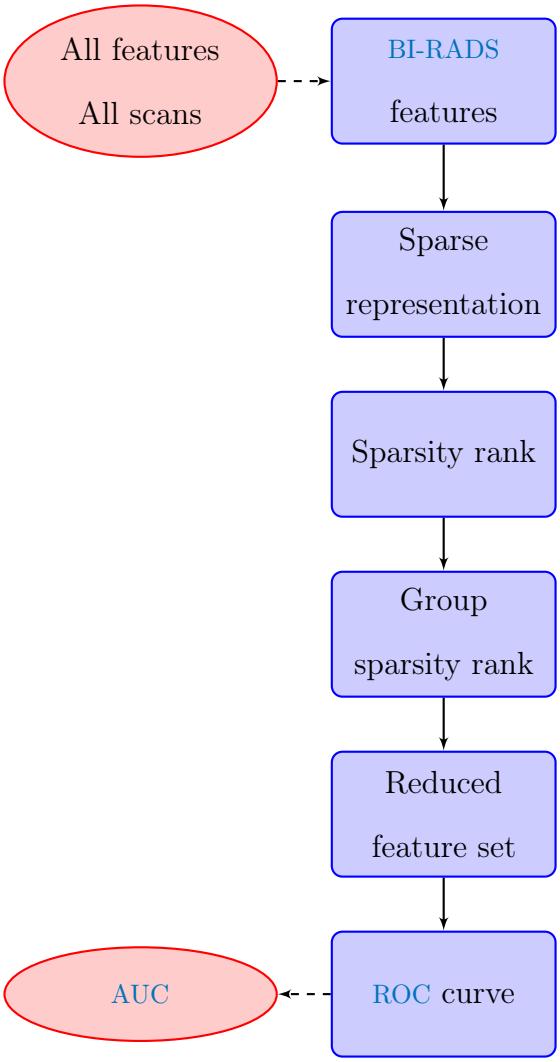


Figure 2.1: Flowchart depicting the classification procedure.

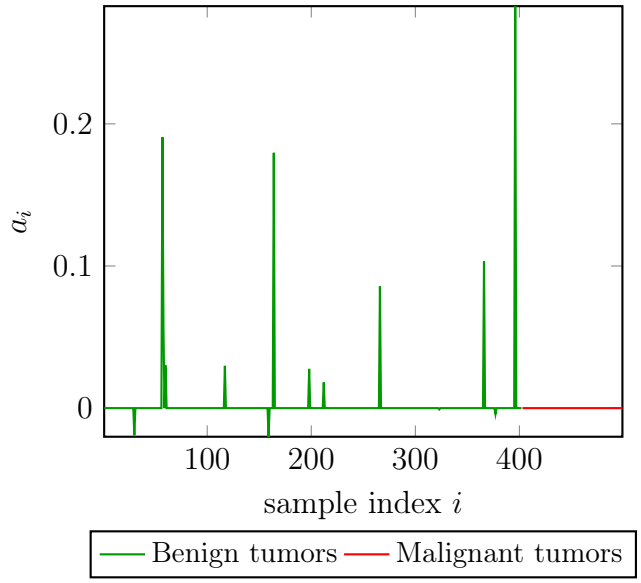


Figure 2.2: An ideal simulation result showing the contribution of the training samples to build up the test sample. Samples from index 1 to index 402 correspond to the actual class. The coefficients for the other samples are zero.

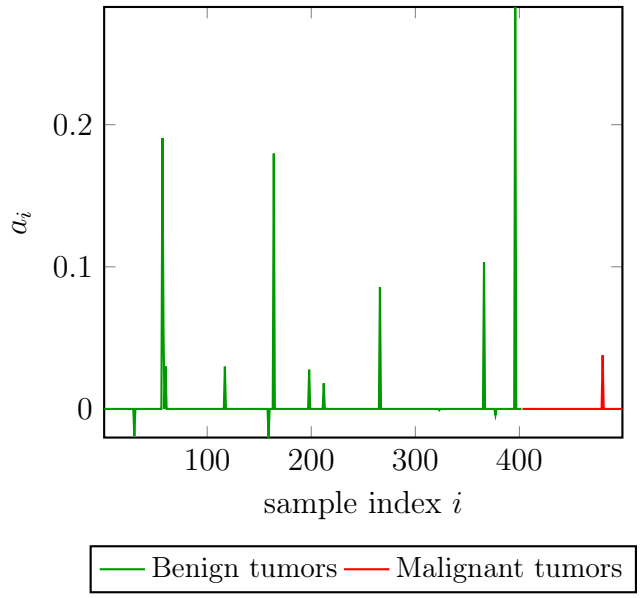
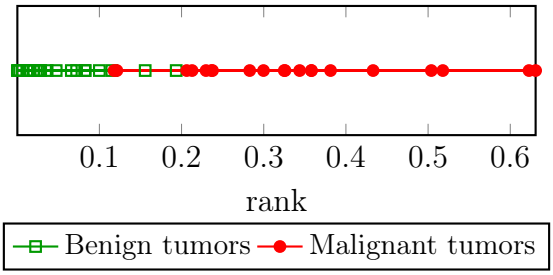
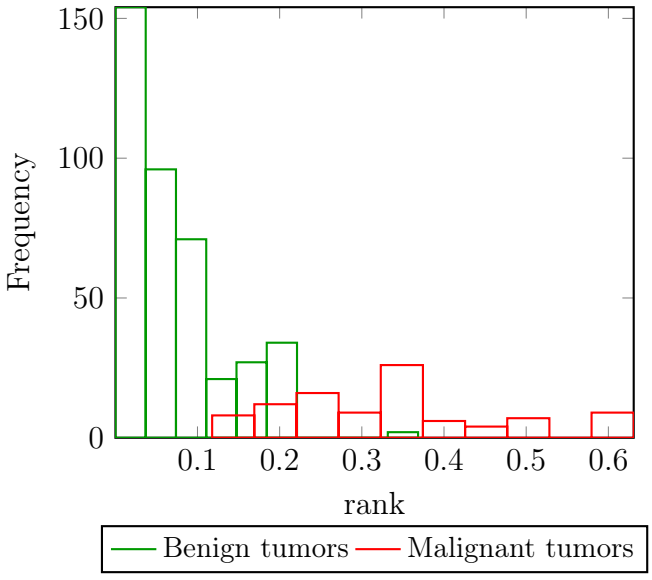


Figure 2.3: A typical simulation result showing the contribution of the training samples to build up the test sample. Samples from index 1 to index 402 correspond to the actual class. Some of the coefficients of the other samples are small but non-zero.



(a) Spread of rank



(b) Histogram of rank

Figure 2.4: A typical simulation result showing the distribution of the rank of each class.

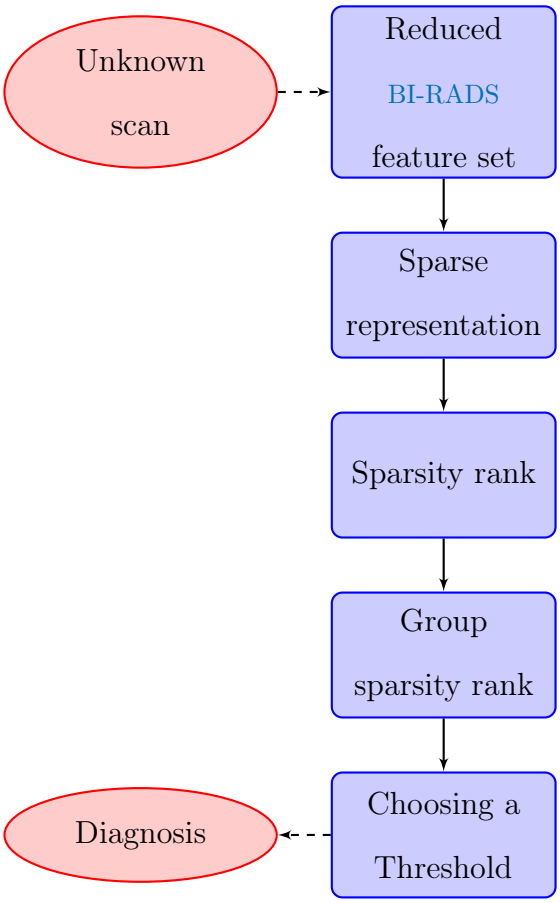


Figure 2.5: Flowchart depicting the classification procedure of a sample with unknown diagnosis.

Chapter 3

Implementation of the Proposed Technique

The last chapter has developed a novel technique to classify tumors. The current chapter will discuss the technical details to go through an experiment to classify breast tumors from a database of BUS images.

3.1 ATL Database

We have performed our experiments on a database which has been generated by Advanced Technology Laboratories (ATL) during its pre-market approval (PMA) studies in 1994, referred to as ATL database. Data were acquired at the Thomas Jefferson University, the University of Cincinnati, and Yale University. The data are contributions from Riverside Research. The data set contains the RF data and

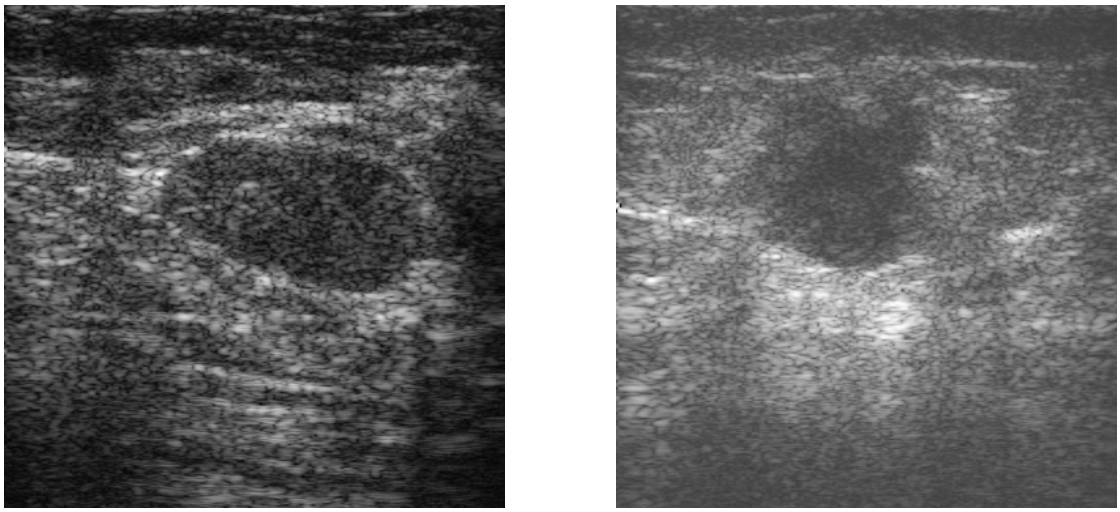
diagnosis results of $n = 502$ breast tumor scans out of which $n_1 = 405$ were found to have benign masses, and $n_2 = 97$ to have malignant masses through biopsy. Many of the tumor lesions were scanned at multiple intersecting scanning planes, N^θ ranging from 0 to 25, providing complementary data about the tumors. A total of $n^\theta = 130$ unique tumors, 104 benign and 26 malignant, were scanned, and lesions showed up in these patients' mammograms.

An expert sonographer or radiologist examined the masses using a Philips Ultrasound UM-9 HDI scanner with a 7.5 MHz L10-5 linear-array transducer. Standard breast examination procedures were followed. The operator set the transducer at a constant power level and a constant transmit focal length. Interfacing the scanner with a Spectrasonics Inc. data acquisition module allowed for digital acquisition of RF echo-signal data, taken at a sampling rate of 20 MHz with a resolution of 14 bits.

Envelope detection was performed on the RF data to generate a B-mode image (or envelope image). Figure 3.1 shows the B-mode images from the scan of two typical breast masses, one benign and one malignant.

3.2 Preprocessing

The boundary corresponding to the tumors was traced using custom software in Riverside Research, and the extracted boundary was set as the Region of Interest (ROI). Figure 3.2 shows the traced tumor from the scan of two typical breast masses, one benign and one malignant.



(a) benign tumor (b) malignant tumor

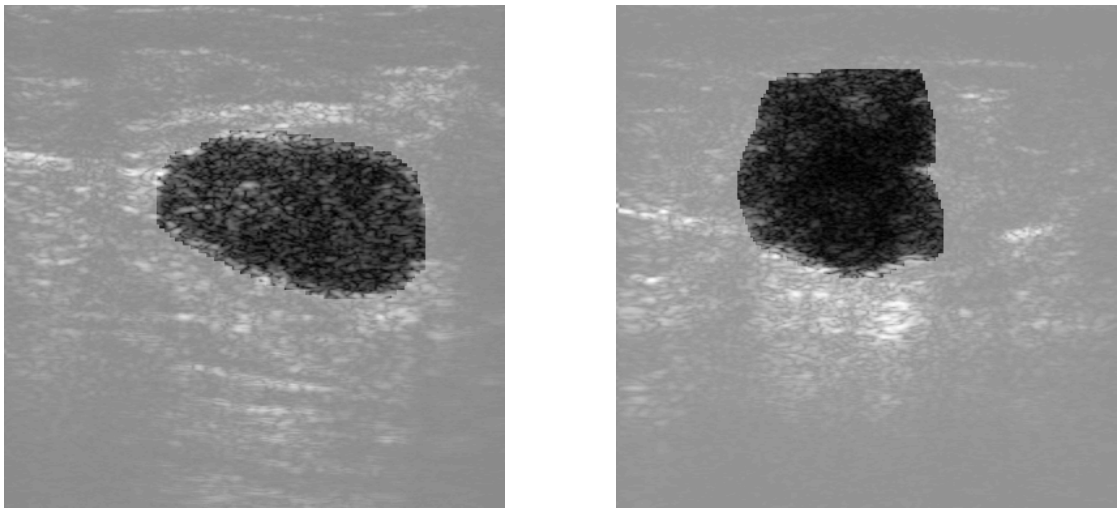
Figure 3.1: B-mode images $\mathcal{P}$ of breast masses.

A set of analysis regions along with the tumor boundary was demarcated to facilitate feature extraction explained later in Section 3.4. Figure 3.3 shows the traces of the nine analysis regions superimposed on the ultrasound images. With respect to the tumor, these analysis regions are: left anterior (la), tumor anterior (ta), right anterior (ra), left lateral (ll), tumor (L), right lateral (rl), left posterior (lp), tumor posterior (tp), and right posterior (rp).

Calibration was needed to account for the three contributions captured in the RF images other than the tissue information, which are:

1. the combined two-way transfer function of the transducer and the data acquisition module,

2. the two-way range-dependent diffraction function raised from beam properties, and



(a) benign tumor (b) malignant tumor

Figure 3.2: Traced tumors $\mathcal{L}$ in breast masses.

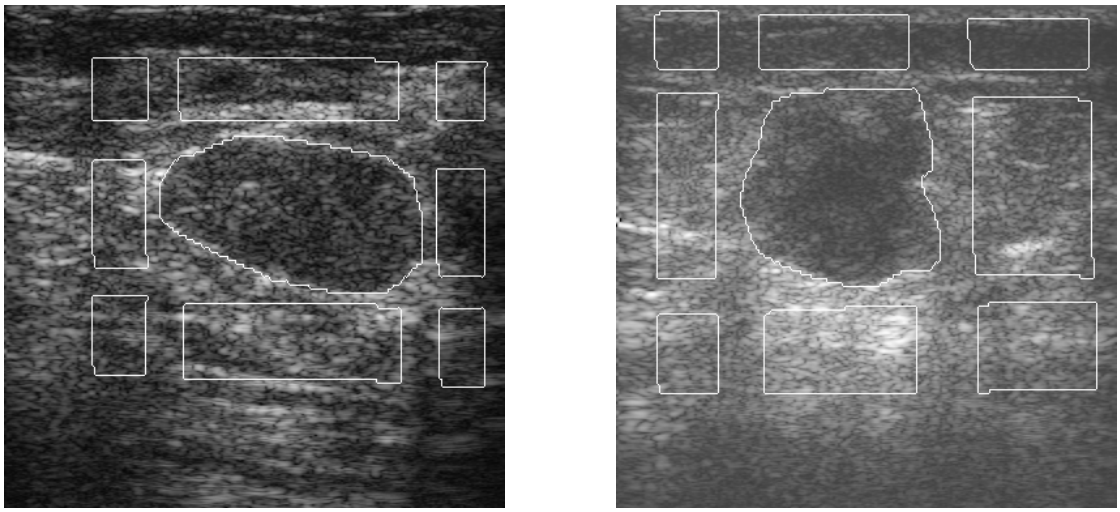
3. the acoustic attenuation.

The estimation of the three functions required setting up of experiments and using empirical results as follows:

1. The transfer function was estimated using a planar reflection method from RF data acquired from the planar surface of a Room Temperature Vulcanizing silicon block put in a water bath.

2. The diffraction function was estimated using a reference-phantom methods from RF data acquired from a rubber block containing a diffuse suspension of 10 µm-diameter glass spheres.

3. The acoustic attenuation was estimated using an empirical attenuation coefficient of $1\,\mathrm{dB\,MHz^{-1}\,cm^{-1}}$.

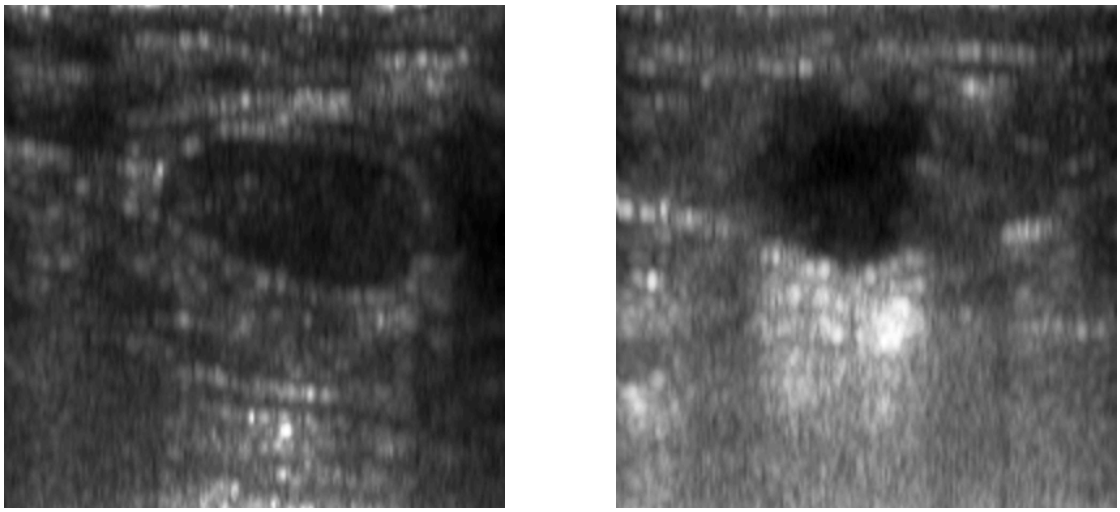


(a) benign tumor (b) malignant tumor

Figure 3.3: B-mode images of breast masses with traces of analysis regions superimposed.

Spectral analysis was performed by computing a Fourier transform using FFT over the RF data windowed by a 2.4 mm Hamming window, calculating and expressing the power spectrum in 1 dB. The resulting power spectrum was calibrated by subtracting the transfer function, diffraction function and the acoustic attenuation. The signal over a bandwidth of 6 dB was considered for the linear regression during the analysis of the spectrum: the parameters of interest were the midpoint of the regression line, or midband (M), its intercept at zero frequency (I), and its slope (s). Progressively sliding the windows over all RF data in an overlapping fashion and repeating the above sequence generates images corresponding to these parameters: the spectral midband image or M image, the spectral intercept image or I image, and the spectral slope image or s image.

Figure 3.4 shows the M image from the scan of two typical breast masses, one benign and one malignant.



(a) benign tumor (b) malignant tumor

Figure 3.4: M images of breast masses.

3.3 Classification

We deploy the classifier with feature space dimension of 25. The traced tumors seem to vary from pixel size 111×192 to 265×192, which have been converted to a vector[1] of a general length of 25 by means of interpolation for each of the image types, which are: RF data, envelope data, M image, I image, and s image. The length has been chosen so as to allow performing the simulation in a reasonable time. We compute the sparsity rank from Equation (2.19), and generate the ROC curve using the rank from Equation (2.24). We use the leave-one-out approach for the cross-validation of the classification process. Finally we compute the AUC

[1]Since the pixels are considered as independent elements of a vector, conversion of the tumors from a matrix form to a vector form does not affect the representation of a test sample in terms of the training samples.

using the trapezoidal rule:

$$AUC = \sum_{\lambda=1}^{n-1} \frac{1}{2}[TPR_{\lambda+1} + TPR_{\lambda}][FPR_{\lambda+1} - FPR_{\lambda}] \tag{3.1}$$

The steps are shown in Figure 3.5.

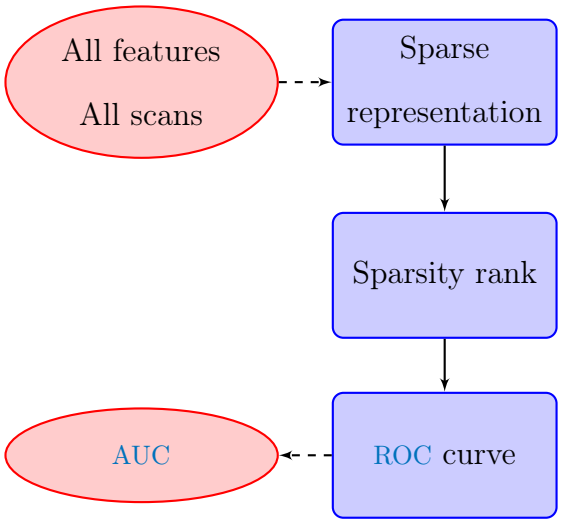


Figure 3.5: Flowchart depicting the classification procedure using the images.

3.4 The Role of Feature Extraction

BI-RADS features are reported to allow for better classification results. Twenty five different, but not necessarily independent, features from the BI-RADS criteria and their variants have been implemented in the current research. These are: echogenicity, heterogeneity, co-occurrence contrast, co-occurrence contrast with image in dB, 4-neighborhood pixels algorithm (FNPA) to measure texture, 4-neighborhood pixels algorithm (FNPA) with image in dB, Hurst coefficient to measure heterogeneity, Hurst coefficient with image in dB, two different

implementations of shadowing, two different implementations of relative absorption, margin definition (gradient-based), margin definition (edge-based), lesion aspect ratio, lesion area, lesion aspect ratio divided by area, lesion compactness, lesion roundness, lesion convexity, lesion solidity, lesion form factor, lesion Kolmogorov dimension to measure border irregularity, lesion Minkowski dimension to measure border irregularity, lesion Hausdorf dimension to measure border irregularity.

Table 3.1 shows the within-class compactness and the between-class separability of each feature by listing the standard deviations of the two classes and the difference of the means of the classes.

Figure 3.6 shows the steps to classify using these extracted features.

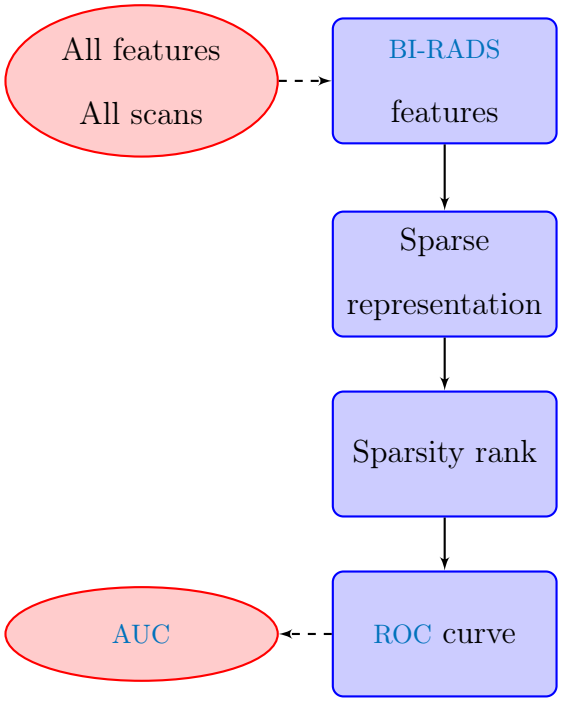


Figure 3.6: Flowchart depicting the classification procedure using the BI-RADS features.

Following are the descriptions of a few selected feature:

Echogenicity Echogenicity can be defined as the mean of the I image within the lesion:

$$\text{echogenicity} = \mu_{I_L} \tag{3.2}$$

The benign mass is found to be more echogenic compared with the malignant mass.

Heterogeneity Heterogeneity can be defined as the standard deviation of the M image within lesion:

$$\text{heterogeneity} = \sigma_{M_L} \tag{3.3}$$

Since M images are less noisy than I and s images, they are more suitable to measure heterogeneity of the lesion itself. The benign mass is found to be less heterogeneous, i.e., more homogeneous, compared with the malignant mass.

FNPA 4-neighborhood pixels algorithm (FNPA) can be defined as the normalized gradient of the M image within lesion:

$$\begin{aligned}\text{FNPA} = \mathcal{A}\Big[\frac{1}{4}\Big\{&|M(k,l) - M(k-1,l)| + |M(k,l) - M(k,l-1)| \\ &+|M(k,l) - M(k+1,l)| + |M(k,l) - M(k,l+1)|\Big\}\Big]\end{aligned} \tag{3.4}$$

FNPA is a descriptor of texture, and the benign mass is found to be have lower FNPA compared with the malignant mass.

Co-occurrence Contrast Co-occurrence Contrast is a measure calculated from the co-occurrence matrix, defined as the 2D normalized autocorrelation

coefficient of the M image within lesion:

$$\gamma(\Delta k, \Delta l) = \frac{A(\Delta k, \Delta l)}{A(0,0)} \tag{3.5}$$

where

$$A(\Delta k, \Delta l) = \mathcal{A}[|\{M(k,l) - \mu_M\}\{M(k+\Delta k, l+\Delta k) - \mu_M\}|] \tag{3.6}$$

Co-occurrence contrast is a descriptor of texture, and the benign mass is found to be have lower co-occurrence contrast compared with the malignant mass.

Shadowing Shadowing can be defined as the normalized difference between mean M values in shadowed and unshadowed posterior region:

$$\text{shadowing} = \frac{\frac{1}{2}\left[\mu_{M_{\text{tp}}} - \mu_{M_{\text{lp}}} + \mu_{M_{\text{tp}}} - \mu_{M_{\text{rp}}}\right]}{d_L} \tag{3.7}$$

where d_L is the lesion thickness. The benign mass is found to be have anti-shadowing or enhancement compared with the shadowing of the malignant mass.

Relative Absorption Relative Absorption can be defined as:

$$\begin{aligned}\text{relative absorption} = &\frac{1}{2}\left[\frac{\mu_{M_{\text{lp}}} - \mu_{M_{\text{ll}}}}{d_{\text{lp,ll}}} + \frac{\mu_{M_{\text{rp}}} - \mu_{M_{\text{rl}}}}{d_{\text{rp,rl}}}\right] \\ &- \frac{\mu_{M_{\text{tp}}} - \mu_{M_{\text{L}}}}{d_{\text{tp,L}}}\end{aligned} \tag{3.8}$$

where $d_{x,y}$ is the distance between the centroids of the analysis regions x and y. The benign mass is found to be have lower relative absorption compared with the malignant mass.

Aspect Ratio Aspect Ratio can be defined as the maximum vertical lesion dimension divided by maximum horizontal lesion dimension. The benign

mass is found to be have lower aspect ratio compared with the malignant mass.

Gradient-based Margin Definition Gradient-based Margin Definition can be defined as the normalized absolute gradient sum of the M image evaluated over the lesion contour:

$$\text{gradient-based margin definition} = \frac{\sum|\nabla M_{L,\text{contour}}|}{\sum|M_{L,\text{contour}}|} \tag{3.9}$$

The benign mass is found to be have higher gradient-based margin definition compared with the malignant mass.

All features mentioned in the current research are described in a greater detail in a work by Alam et. al. [10].

3.5 Exploiting Redundancy

Many of the tumors of the ATL database have several scans at different scanning planes. For example, 25 different scans correspond to tumor JR00371. We employ $\text{rank}_{\text{group}}$ on the obtained rank vector using Equation (2.28), and generate the ROC curve. The steps are shown in Figure 3.7.

For comparison, we have removed the outliers from the rank of the duplicate scans, where an outlier is defined as any rank value at least twice the standard deviation away from the mean.

Also, we have selected the class that is detected most in the duplicate scans, which is essentially an implementation of the majority rule.

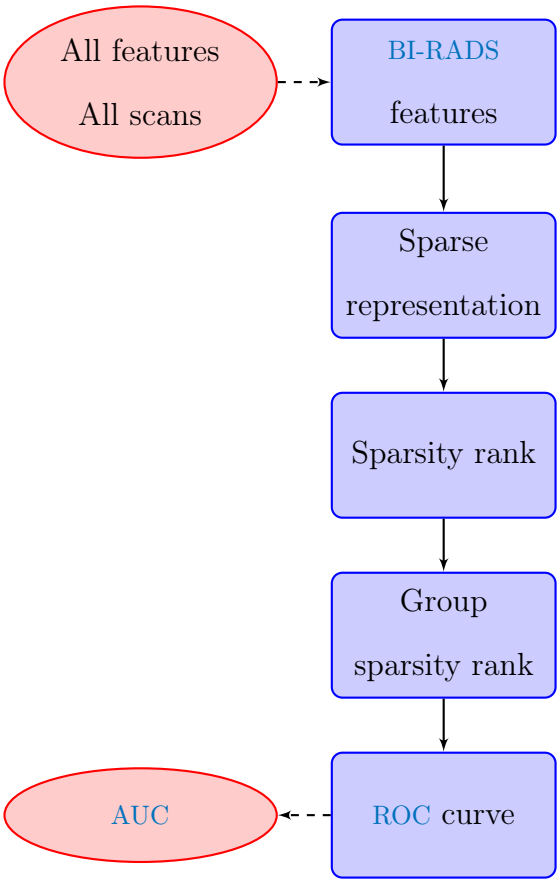


Figure 3.7: Flowchart depicting the classification procedure employing the group rank.

3.6 Optimizing Feature Set

To come up with a reduced feature set, we have implemented a recursive search among the 25 features used in classification. To avoid the risk of over-training, the training set and the cross-validation set have been chosen to be mutually exclusive by randomly splitting the data set in two halves.

The whole procedure has been summarized in Figure 2.1, and is reproduced in Figure 3.8 for convenience.

For comparison, we went through a p-value analysis of the feature list. We sorted the list in an ascending order of the p-values of the features. We kept on removing the last feature from the current list until the resulting AUC stopped improving.

We have also run PCA over the data set and opted for keeping only the first 15 and 10 significant principal components respectively.

This chapter acts as a reference to an experimental setup to classify breast tumors from ultrasound images. The process of data acquisition, the necessary operating points, preprocessing and segmentation, feature extraction, exploitation of duplication, and optimization of feature set have been discussed.

The obtained results and their interpretations will be described in the following chapter.

Table 3.1: Statistics of BI-RADS features.

Feature	Difference of means	Standard deviation of benign samples	Standard deviation of malignant samples
Echogenicity	3.48	7.18	6.47
Heterogeneity	0.48	1.53	1.11
Co-occurrence Contrast	1.88	4.59	1.75
Co-occurrence Contrast with Image in dB	0.09	0.64	0.05
FNPA	-0.01	0.02	0.01
FNPA with Image in dB	0.02	0.04	0.03
Hurst Coefficient	0.07	0.11	0.08
Hurst Coefficient with Image in dB	0.30	0.43	0.20
Shadowing	1.88	15.84	9.72
Shadowing-2	4.56	14.75	10.19
Relative Absorption	-0.10	0.38	0.28
Relative Absorption-2	-0.14	0.36	0.26
Gradient-based Margin Definition	0.03	0.03	0.02
Edge-based Margin Definition	0.04	0.05	0.04
Aspect Ratio	-0.27	0.22	0.30
Area	-0.44	0.81	1.04
Aspect Ratio Divided by Area	0.10	2.54	2.13
Compactness	-0.06	0.12	0.08
Roundness	-0.08	0.16	0.12
Convexity	0.03	0.02	0.04
Solidity	0.03	0.04	0.06
Form Factor	-0.01	0.13	0.12
Kolmogorov Dimension	-0.07	0.11	0.08
Minkowski Dimension	0.04	0.07	0.04
Hausdorf Dimension	0.00	0.01	0.01

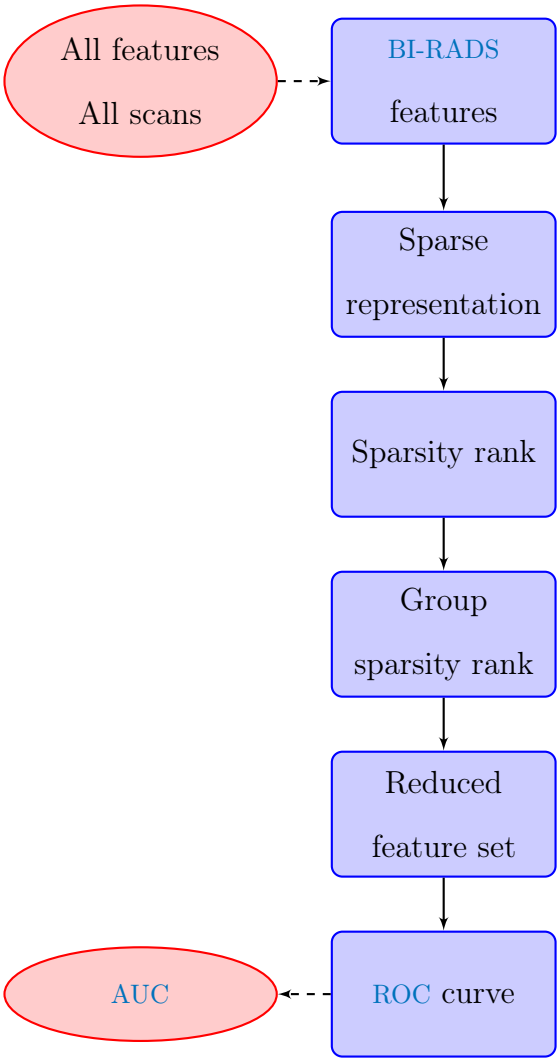


Figure 3.8: Flowchart depicting the classification procedure.

Chapter 4

Results and Discussion

The last chapter has discussed designing and conducting an experiment to verify the proposed technique on breast tumor classification. The current chapter lists the obtained results, and explains and interprets them to have a better understanding of the technique.

4.1 ℓ_1-Minimization and Size of Training Set

It is well established that a larger training database helps training a classifier more accurately. This work uses YALL1 package[1] to implement ℓ_1-minimization. We have plotted the variation of the ℓ_2-norm of the recovery error $\|\mathbf{x}_r - \mathbf{x}_1\|_2$, normalized by the length of $\mathbf{x}$ in Figure 4.1. The trend of the curve demonstrates that a larger training set implies a reduced error in sparse representation. Thus it

[1] yall1.blogs.rice.edu

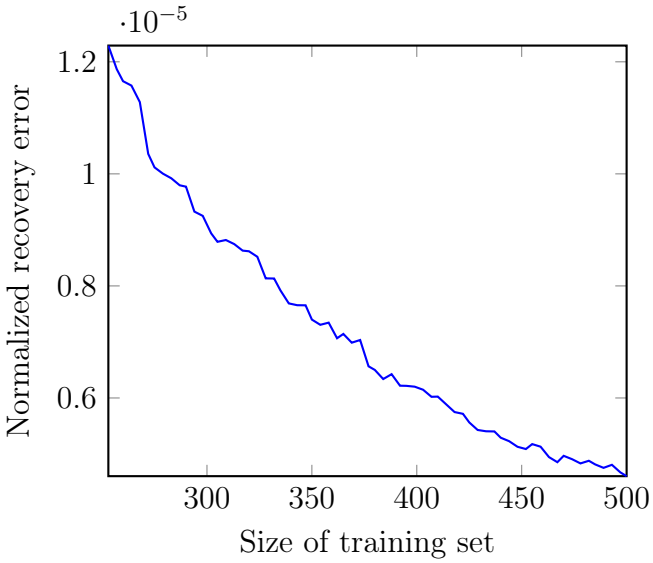


Figure 4.1: Effect of increasing size of training set on the recovery error of the sparse representation of the test sample vector.

is best to choose the largest possible training set, which motivates us to follow a leave-one-out approach.

Moreover, the amplitude of the error, combined with the effect of increasing data size, as can be seen from the graph, can be attributed to the convergence of the aforementioned ℓ_1-minimization algorithm.

4.2 Classification

Implementing the proposed classifier, as mentioned in Section 3.3, on each of the image types, RF data, envelope data, M image, I image, and s image, give AUCs between 0.48 and 0.57, as shown in Table 4.1. The corresponding ROC curves are shown in Figure 4.2. The ROC curves show a tendency towards the diagonal

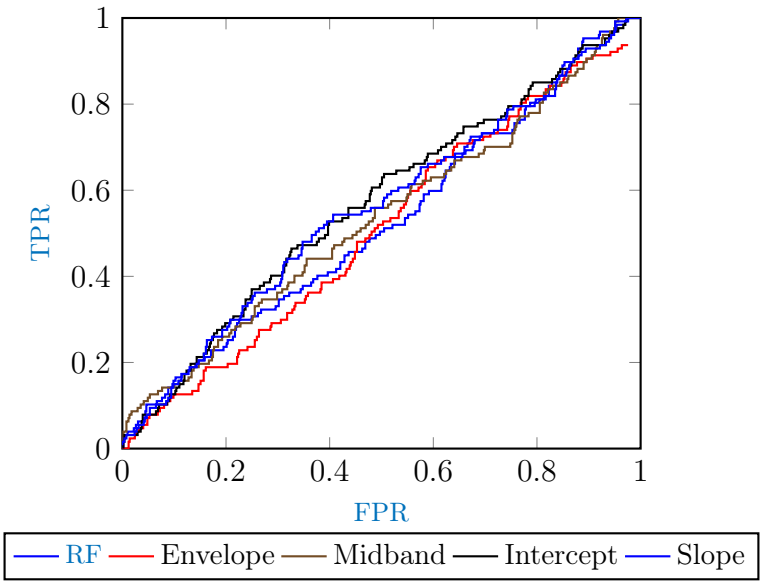


Figure 4.2: Classification with images.

line $TPR = FPR$, explaining the AUC values around 0.5. Thus the five images themselves contain little more information than a random image, detecting the right class in half of the cases and detecting the wrong one in the other half, since only a small portion of all the pixels carry useful and discriminating information about their class, and get suppressed by the rest of the non-discriminating pixels. Choosing in a downsampling fashion, rather than going through some intelligent transformation of the pixels, maintains the same ratio of useful and non-useful pixels.

4.3 The Role of Feature Extraction

Implementation of BI-RADS features and using the rank to classify the tumors gives the ROC curve as shown in Figure 4.3. It is found that an AUC of 0.8278 could be

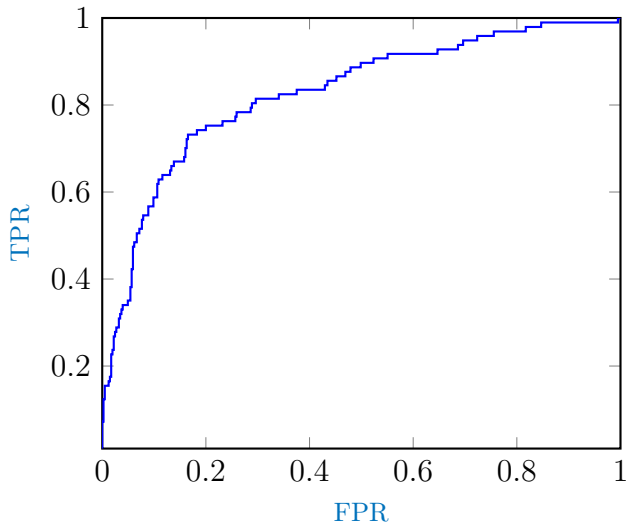


Figure 4.3: Classification with BI-RADS feature set using the rank.

achieved, which is a vast improvement over the performance on the classification using images alone without using BI-RADS features. In particular, the features give highly discriminatory information necessary to separate the two classes, as expected.

4.4 Exploiting Redundancy

Employing $\text{rank}_{\text{group}}$ on the database gives the ROC curve as shown in Figure 4.4. Employing this technique improves the performance further achieving an AUC of 0.9459. Thus the group rank has proven to successfully harness the inherent redundancy included in the data set.

In comparison, instead of averaging over the rank, removing the outliers results in an AUC of 0.8674.

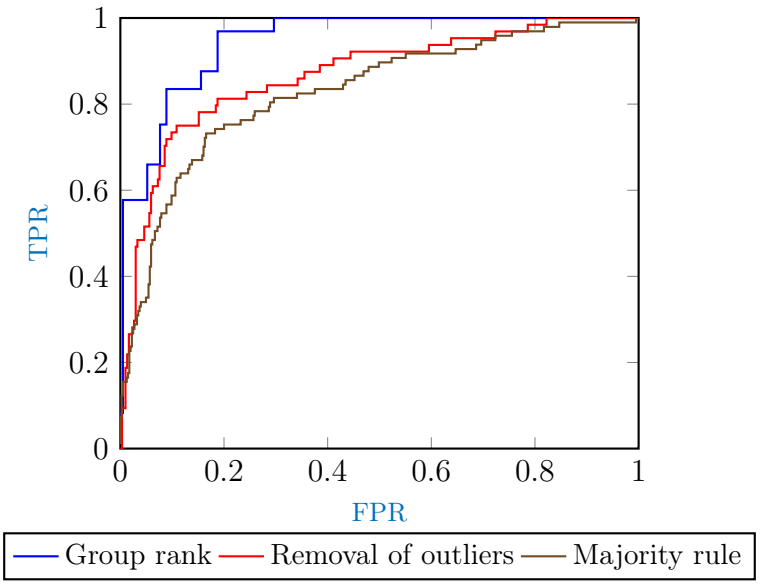


Figure 4.4: Classification with BI-RADS feature set (a) using group rank, (b) removing outliers, and (c) using the majority rule.

Also, implementing a majority rule gives the ROC curve with an AUC of 0.8278, no visible improvement.

4.5 Optimizing Feature Set

Among the 25 features used in classification, choosing the reduced set of features by recursive search improves the classification performance, as shown in Figure 4.5, giving the ROC curve with an AUC of 0.9754.

In contrast, choosing the best AUC by taking the set of features with lowest p-values gives the ROC curve that has only a marginal improvement with the improved AUC of 0.9495.

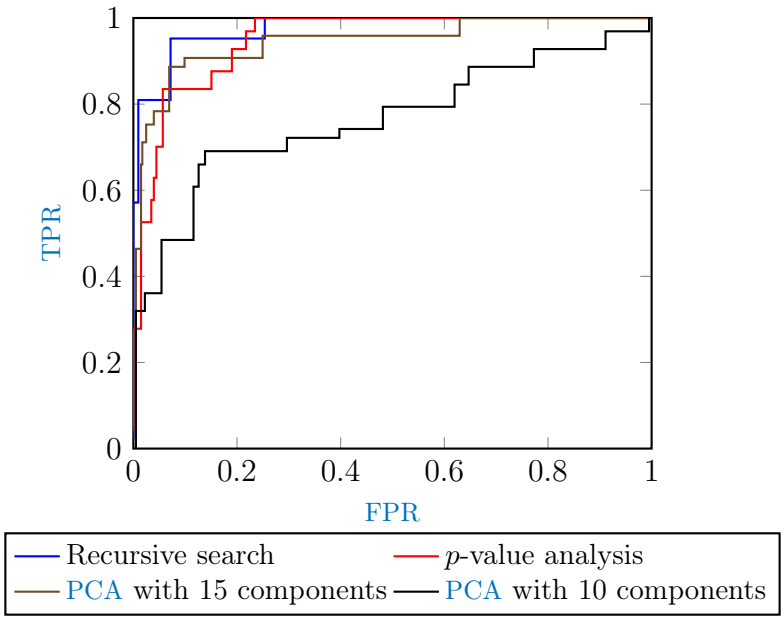


Figure 4.5: Classification with (a) reduced BI-RADS feature set, (b) BI-RADS feature set from p-value analysis, (c) first 15 principal components, and (d) first 10 principal components.

Using the first 15 significant principal components gives the ROC curve with an AUC of 0.9450, and the first 10 of them correspond to an AUC of 0.7646. Thus choosing principal components as features makes a zero or negative improvement.

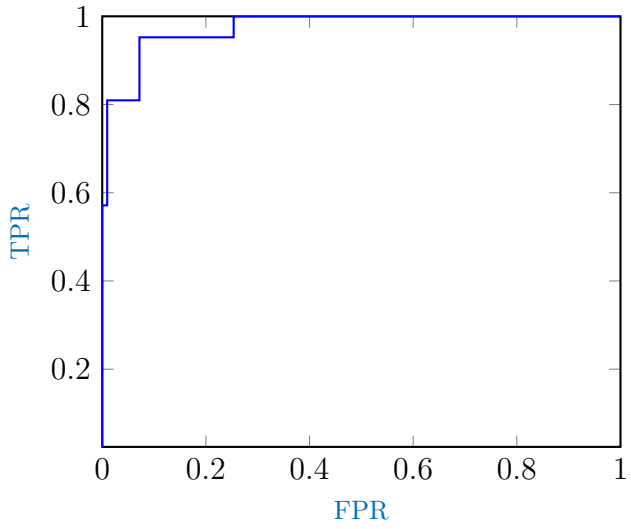
The reduced set includes 15 features, namely, echogenicity, heterogeneity, co-occurrence contrast, 4-neighborhood pixels algorithm (FNPA) to measure texture, 4-neighborhood pixels algorithm (FNPA) with image in dB, Hurst coefficient to measure heterogeneity, Hurst coefficient with image in dB, shadowing, relative absorption, lesion aspect ratio, lesion area, lesion compactness, lesion roundness, lesion form factor, and lesion Hausdorf dimension to measure border irregularity.

4.6 Comparative Study

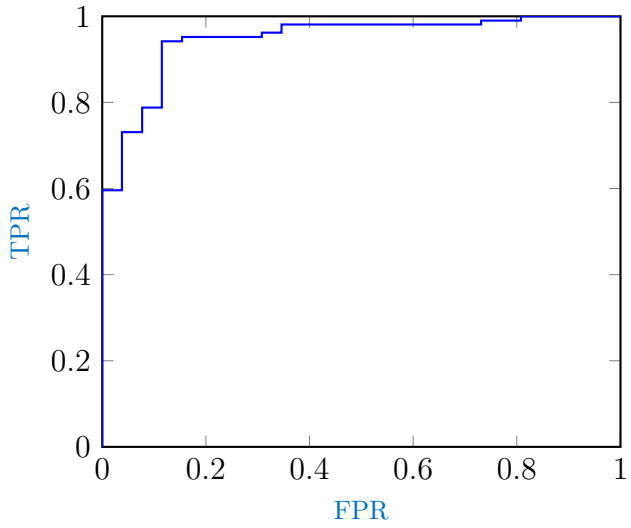
This section accumulates the performance result from the proposed technique as well as those reported by some of the work available in the literature as presented in Section 1.8.

With the ROC curve obtained from the proposed classifier shown in Figure 4.6a, Figure 4.6b shows the ROC curve obtained by reproducing the work of Alam et. al. [10], which uses the same data set as used in the current research. Though not a comparison study per se due to the difference in the data set involved, ROC curves and AUCs from the work of Tan et. al. [8], Shankar et. al. [9], Joo et. al. [11], and Huang and Chen [12] are shown in Figures 4.6c to 4.6f and Table 4.2.

These results insinuate that the proposed method surpasses or is comparable to the other ones reported in the literature. It can be noted that the work of Shankar et. al., Huang and Chen, and Tan et. al. employed a Gaussian-based ROC curve, while our work as well as those of Alam et. al. and Joo et. al. implemented a trapezoidal rule which gives a slightly lower estimate of AUC than the smooth Gaussian estimate [28]. Methods listed in Section 1.8 that do not perform any ROC analysis are not considered for the comparative study.

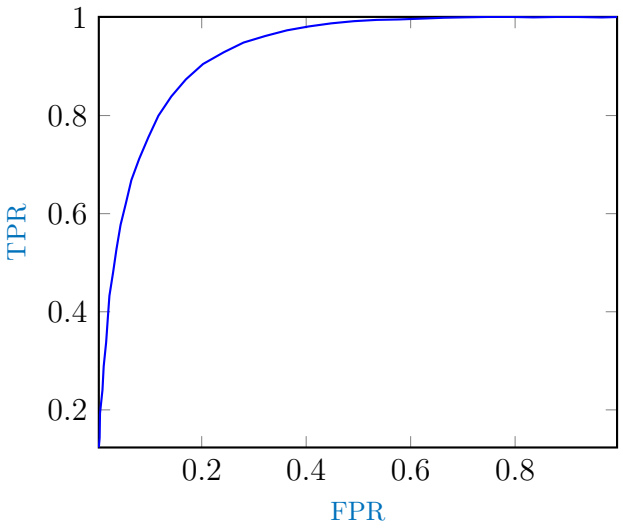


(a) Classification using the proposed method.

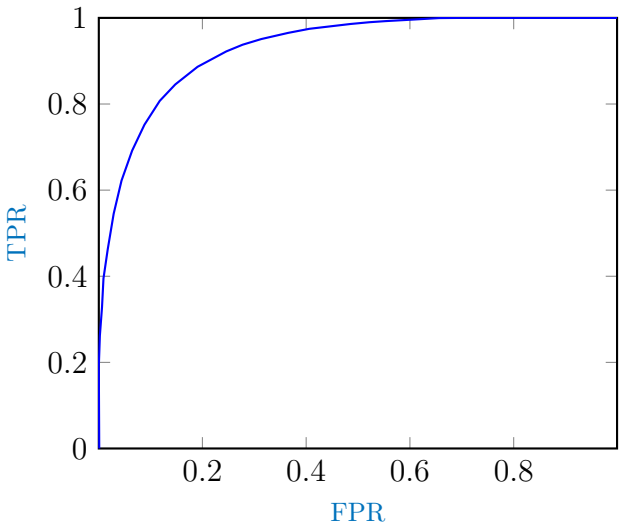


(b) Classification using the work of Alam et. al. [10].

Figure 4.6: Classification using the proposed method and state of the art.

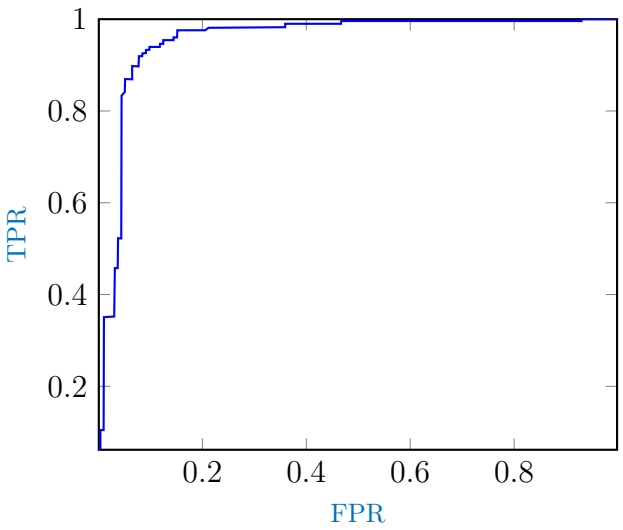


(c) Classification using the work of Tan et. al. [8].

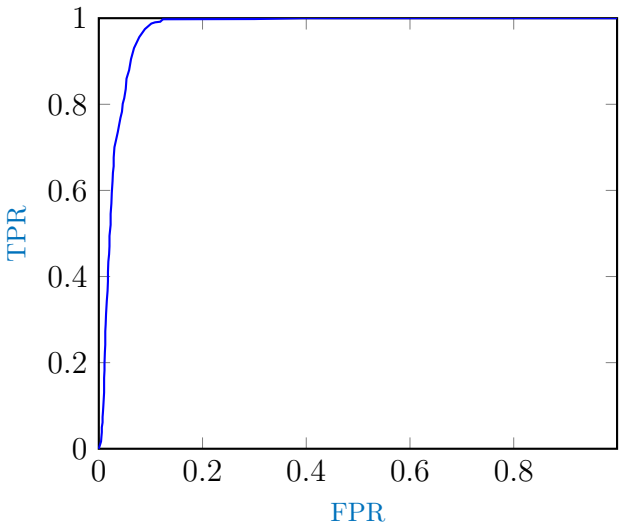


(d) Classification using the work of Shankar et. al. [9].

Figure 4.6 (Continued): Classification using the proposed method and state of the art.



(e) Classification using the work of Joo et. al. [11].



(f) Classification using the work of Huang and Chen [12].

Figure 4.6 (Continued): Classification using the proposed method and state of the art.

Table 4.1: Classification results at different stages.

Name of Feature	AUC
RF	0.5220
Envelope	0.4857
Midband	0.5336
Intercept	0.5681
Slope	0.5550
BI-RADS features using the rank	0.8278
Removal of outliers	0.8674
Majority rule	0.8278
BI-RADS features using the group rank	0.9459
p-value analysis	0.9495
PCA with 15 components	0.9450
PCA with 10 components	0.7646
Optimized BI-RADS features	0.9754

Table 4.2: Classification results of the proposed method and state of the art.

Classifier	AUC
Proposed classifier	0.9754
Tan et. al. [8]	0.93
Shankar et. al. [9]	0.94
Alam et. al. [10]	0.95
Joo et. al. [11]	0.96
Huang and Chen [12]	0.97

Chapter 5

Conclusions and Suggestions for Future Work

5.1 Summary of the Thesis

Sparse Representation-based Classifier (SRC) is introduced for the first time in medical diagnosis from ultrasound images. This work elegantly demonstrates that SRC becomes a powerful tool in discriminating between malignant and benign lesions when the SRs of multiple scans are averaged and feature sets are optimized. Therefore, the judicial use of multiple scans indeed makes difference and it is always a better option compared to the complicated biopsy when the later one is less convenient and less comfortable for a patient and also costlier in several order of magnitudes.

5.2 Future Work

In future, we hope to look into the cases where we would like to investigate how the multiple scans can be combined for the conventional classifiers such as SVM, LOGREG, and LDA for better performance and how they perform compared to the proposed one in this thesis.

References

[1] W. M. Saltzman, *Biomedical Engineering: Bridging Medicine and Technology (Cambridge Texts in Biomedical Engineering)*. Cambridge University Press, Jun. 2012. [Online]. Available: http://amazon.com/o/ASIN/0511802730/

[2] A. C. Society, "Cancer facts & figures 2013," Atlanta: American Cancer Society, Annual Report, 2013.

[3] C. D. Mathers, D. M. Fat, and J. Boerma, *The global burden of disease: 2004 update*. World Health Organization, 2008.

[4] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: a cancer journal for clinicians*, vol. 61, no. 2, pp. 69–90, 2011. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-79952232216&partnerID=40&md5=39a084ccb65a4dbcecf07e612f44f3d9

[5] B. O. Anderson, C.-H. Yip, S. D. Ramsey, R. Bengoa, S. Braun, M. Fitch, M. Groot, H. Sancho-Garnier, and V. D. Tsu, "Breast cancer in limited-resource countries: Health care systems and public

policy," *The Breast Journal*, vol. 12, no. s1, pp. S54–S69, 2006. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-33644890223&partnerID=40&md5=cbcb4d9c687519f286b98f00f2412870

[6] H. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern Recognition*, vol. 43, no. 1, pp. 299–317, 2010. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-68949164775&partnerID=40&md5=0fe1d5af640ae572b7a4b8a51165e022

[7] M. Halliwell, "A tutorial on ultrasonic physics and imaging techniques," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 224, no. 2, pp. 127–142, 2010. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-76849088916&partnerID=40&md5=8c40de3f9a859961956a22cc90df5700

[8] T. Tan, B. Platel, H. Huisman, C. I. Sánchez, R. Mus, and N. Karssemeijer, "Computer-aided lesion diagnosis in automated 3-D breast ultrasound using coronal spiculation," *IEEE Transactions on Medical Imaging*, vol. 31, no. 5, pp. 1034–1042, 2012. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84860688845&partnerID=40&md5=d5f188b2425b23dbee30a9e8559d4a93

[9] P. M. Shankar, V. A. Dumane, C. W. Piccoli, J. M. Reid, F. Forsberg, and B. B. Goldberg, "Computer-aided classification of breast masses in ultrasonic B-scans using a multiparameter approach," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 50, no. 8, pp. 1002–1009, 2003.

[Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0141888745&partnerID=40&md5=72abc5abd5ba5b8989121b72734437e6

[10] S. K. Alam, E. J. Feleppa, M. Rondeau, A. Kalisz, and B. S. Garra, "Ultrasonic multi-feature analysis procedure for computer-aided diagnosis of solid breast lesions," *Ultrasonic Imaging*, vol. 33, no. 1, pp. 17–38, 2011. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-79957634769&partnerID=40&md5=77d95ca9a4ef15efb1e642a6cc4ea6ce

[11] S. Joo, Y. S. Yang, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: Use of an artificial neural network based on multiple sonographic features," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1292–1300, 2004. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-6344221607&partnerID=40&md5=8a20cf813da773bf6d7dd71c5c1fe3a9

[12] Y.-L. Huang and D.-R. Chen, "Support vector machines in sonography: Application to decision making in the diagnosis of breast cancer," *Clinical Imaging*, vol. 29, no. 3, pp. 179–184, 2005. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-18044391075&partnerID=40&md5=8fbe03d9da6c885ce3af5c3435f2a8a9

[13] R.-F. Chang, C.-J. Chen, M.-F. Ho, D.-R. Chen, and W. K. Moon, "Breast ultrasound image classification using fractal analysis," in *Proceedings - Fourth IEEE Symposium on Bioinformatics and Bioengineering, BIBE 2004*. Taichung: IEEE, 2004, pp. 100–107.

[Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-4544328219&partnerID=40&md5=9b23fa7f18e31c682c08e86fca8478d9

[14] Z. Yanjiao, L. Jiangli, C. Ke, and P. Yulan, "Boundary-based feature extraction and recognition of breast tumors using support vector machine," in *Proceedings - 2009 International Forum on Information Technology and Applications, IFITA 2009*, vol. 3, Chengdu, 2009, pp. 89–92. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-70350554606&partnerID=40&md5=43ec490eda976915c12efc11df524419

[15] L. Bocchi, F. Gritti, C. Manfredi, E. Giannotti, and J. Nori, "Semiautomated breast cancer classification from ultrasound video," in *Proceedings - International Symposium on Biomedical Imaging*. Barcelona: IEEE, 2012, pp. 1112–1115. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84864833449&partnerID=40&md5=096053dd3e8a2eace99b2c14f86cc694

[16] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997. [Online]. Available: https://www.scopus.com/inward/record.url?eid=2-s2.0-0031185845&partnerID=40&md5=adcf96602c940aa1d0fa3b13fd6082eb

[17] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0035248924&partnerID=40&md5=aae197cac594a62bfb956a7a6f76f49c

[18] S. Fidler, D. Skočaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 337–350, 2006. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-31644433509&partnerID=40&md5=2f4b01259dd9c0e105ac3c0e09ccbf1d

[19] M. Elad and M. Aharon, "Image denoising via learned dictionaries and sparse representation," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 895–900.

[20] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-61549128441&partnerID=40&md5=3a4fa1856e806f8fbd8c92f09636469f

[21] A. C. of Radiology, Ed., *Breast Imaging Reporting and Data System (BI-RADS) Atlas*. Reston, VA: American College of Radiology, 2003.

[22] E. B. Mendelson, W. A. Berg, and C. R. Merritt, "Toward a standardized breast ultrasound lexicon, BI-RADS: Ultrasound," *Seminars in Roentgenology*, vol. 36, no. 3, pp. 217–225, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0037198X01800195

[23] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, 2003. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0037418225&partnerID=40&md5=afe0333c574ad94b00a9f0e6060289bc

[24] E. Amaldi and V. Kann, "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems," *Theoretical Computer Science*, vol. 209, no. 1–2, pp. 237–260, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304397597001151

[25] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-33646365077&partnerID=40&md5=4fdc4eabac5a7704813788dcc9198eae

[26] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006. [Online]. Available: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-33745604236&partnerID=K84CvKBR&rel=3.0.0&md5=816a82e61c2a7df938c3f519da2b9a95

[27] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. Lanckriet, "A direct formulation for sparse PCA using semidefinite

programming," *SIAM review*, vol. 49, no. 3, pp. 434–448, 2007. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-34548514458&partnerID=40&md5=4ed9a94759ce644c2997e0ab02bde5f6

[28] J. Hanley and B. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0020083498&partnerID=40&md5=cc9b3191569612d45d6c23f16ff76d50