

2005

# Development of computer-aided diagnostic system for breast MRI lesion classification

Lina Arbash Meinel  
*University of Iowa*

Copyright 2005 Lina Arbash Meinel

This dissertation is available at Iowa Research Online: <https://ir.uiowa.edu/etd/101>

---

## Recommended Citation

Meinel, Lina Arbash. "Development of computer-aided diagnostic system for breast MRI lesion classification." PhD (Doctor of Philosophy) thesis, University of Iowa, 2005.  
<https://doi.org/10.17077/etd.ay6sqh7y>

---

Follow this and additional works at: <https://ir.uiowa.edu/etd>

Part of the [Biomedical Engineering and Bioengineering Commons](#)

DEVELOPMENT OF COMPUTER-AIDED DIAGNOSTIC SYSTEM FOR  
BREAST MRI LESION CLASSIFICATION

by

Lina Arbash Meinel

An Abstract

Of a thesis submitted in partial fulfillment of the  
requirements for the Doctor of Philosophy  
degree in Biomedical Engineering  
in the Graduate College of  
The University of Iowa

May 2005

Thesis Supervisor: Associate Professor Joseph M. Reinhardt

## ABSTRACT

Breast cancer is the second leading cause of cancer deaths in women today. Currently, mammography is the primary method of early detection. However, research has shown that many cases (10-30%) missed by mammography can be detected using breast MRI (BMRI). BMRI is more difficult to interpret than mammography because it generates significantly more data. Also, there are fewer people qualified to use it for diagnosis because it is not the standard breast imaging modality.

Our goal is to develop and test a CAD system to aid and improve the performance of radiologists with different levels of experience in reading breast MR images. Part of the CAD system is an image loader and viewer capable of displaying multiple sequences simultaneously, with standard region of interest and high level analysis tools. We propose a semi-automatic segmentation method that identifies significant lesions. Then, 42 shape, texture, and enhancement kinetics based features were computed. The top 13 best features were selected and used as inputs to three artificial classifiers: a backpropagation neural network (BNN), a support vector machine (SVM), and a Bayesian classifier (BC). Each one was trained using pathology results as the gold standard. Five human readers (a BMRI expert, two mammographers, and two body imaging fellows) manually classified 75 BMRI datasets (80 lesions), both with and without CAD system assistance. The performance of the computer classifiers and human readers were compared using ROC curves, and the human readers' performance was also evaluated using MRMC analysis.

The ROC curve analysis showed that the BNN system significantly outper-

formed the other two classifiers with  $A_z = 0.970$ , and  $p < 0.05$ , and a sensitivity of 91.3% with zero false positives. Also, all human readers significantly improved when aided by the CAD system ( $p < 0.05$ ). The MRMC analysis showed that the human reader performance with and without CAD system assistance can be generalized over the population of cases and still maintain a statistically significant improvement ( $F(1, 74) = 6.805$ ,  $p = 0.0110 < 0.05$ ). These results show significant advantages to using CAD systems in classifying BMRI lesions.

Abstract Approved: \_\_\_\_\_

Thesis Supervisor

\_\_\_\_\_

Title and Department

\_\_\_\_\_

Date



DEVELOPMENT OF COMPUTER-AIDED DIAGNOSTIC SYSTEM FOR  
BREAST MRI LESION CLASSIFICATION

by

Lina Arbash Meinel

A thesis submitted in partial fulfillment of the  
requirements for the Doctor of Philosophy  
degree in Biomedical Engineering  
in the Graduate College of  
The University of Iowa

May 2005

Thesis Supervisor: Associate Professor Joseph M. Reinhardt

Graduate College  
The University of Iowa  
Iowa City, Iowa

CERTIFICATE OF APPROVAL

---

PH.D. THESIS

---

This is to certify that the Ph.D. thesis of

Lina Arbash Meinel

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Biomedical Engineering at the May 2005 graduation.

Thesis Committee: \_\_\_\_\_

Joseph M. Reinhardt, Thesis Supervisor

\_\_\_\_\_  
Alan H. Stolpen

\_\_\_\_\_  
Laurie L. Fajardo

\_\_\_\_\_  
Edwin L. Dove

\_\_\_\_\_  
Michael A. Mackey



To all women ... May you never need my system.

Lina

## ACKNOWLEDGEMENTS

I would like to give a special thanks to my adviser Professor Joseph M. Reinhardt for his help and supervision. I would like to thank Dr. Alan Stolpen for helping with all the research stages and participating in a preliminary human reader experiment. I also would like to acknowledge Dr. Laurie L. Fajardo, Radiology Chair, UIHC, Dr. Wendie Berg, breast imaging consultant, Dr. Dorria Salem, Radiology Chair, Cairo University, Dr. Lee Bennett, Radiology, UIHC, Dr. Megha Garg, Radiology, UIHC, and Dr. Ann Schreiber, Radiology, UIHC for participating in the human reader experiments.

I would like to give a special thanks to Dr. Lizann Bolinger for helping me in the start of this research, and to Dr. Mitchell Schnall, University of Pennsylvania for allowing me to use his datasets. Also, thanks to Dr. Kevin Berbaum and Dr. Steve Hillis, Radiology, UI for the statistical analysis discussions and performing the MRMC analysis. I also would like to acknowledge Dr. Steven C. Mitchell who coded the SVM classifier.

I would like to thank University of Iowa Student Government (UISG), Women In Science and Engineering (WISE), and Graduate Student Senate (GSS) for the travel fund awards. This work was supported in part by the research grant from Confirma, Inc.

I really appreciate all the discussions and general support from my husband John Arbash Meinel.

## TABLE OF CONTENTS

LIST OF TABLES . . . . .	vii
LIST OF FIGURES . . . . .	ix
CHAPTER	
1 INTRODUCTION . . . . .	1
1.1 Hypothesis and Specific Aims . . . . .	1
2 SIGNIFICANCE . . . . .	3
2.1 Background . . . . .	3
2.2 Significance . . . . .	4
2.3 Literature Review . . . . .	7
2.3.1 Penn 2001 . . . . .	13
3 FUNDAMENTALS . . . . .	16
3.1 MRI . . . . .	16
3.1.1 Spin . . . . .	17
3.1.2 $T_1$ Processes . . . . .	17
3.1.3 Precession . . . . .	19
3.1.4 $T_2$ Processes . . . . .	20
3.1.5 Imaging Parameters . . . . .	21
3.2 Breast MRI . . . . .	24
3.2.1 Uptake of Contrast Agent . . . . .	26
3.2.2 Optimization of Sequence Protocols . . . . .	26
3.3 Neural Network Design . . . . .	30
3.3.1 Neural Network Learning . . . . .	35
3.4 Support Vector Machines (SVM) . . . . .	40
4 METHODS . . . . .	44
4.1 Image Viewer . . . . .	44
4.2 Datasets . . . . .	46
4.3 Specific Aim 1 . . . . .	47
4.3.1 Skin and Chest Wall Segmentation . . . . .	48
4.3.2 Segmentation of Enhanced Regions . . . . .	52
4.3.3 Segmentation Using Morphological Operations . . . . .	57
4.4 Specific Aim 2 . . . . .	60
4.4.1 Shape-based Features . . . . .	64
4.4.2 Texture-based Features . . . . .	65

4.4.3	BMRI Features . . . . .	66
4.4.4	Features Selection . . . . .	70
4.5	Specific Aim 3 . . . . .	70
4.5.1	Classification of Breast Masses from Mammograms . . . . .	70
4.5.2	BNN Design . . . . .	71
4.5.3	SVM Design . . . . .	83
4.5.4	BC Design . . . . .	87
4.5.5	Data Classification . . . . .	88
4.5.6	Classification by Human Readers . . . . .	89
4.6	Specific Aim 4 . . . . .	90
4.6.1	Chest Wall Detection . . . . .	91
4.6.2	Lesion Classification: CAD Performance . . . . .	91
4.6.3	Reader Performance with and without CAD Assistance . . . . .	91
4.6.4	MRMC Method . . . . .	92
5	RESULTS . . . . .	94
5.1	Segmentation of BMRI . . . . .	94
5.1.1	Skin and Chest Wall Segmentation . . . . .	94
5.1.2	Segmentation of Enhanced Regions . . . . .	96
5.1.3	Segmentation Using Morphological Operations . . . . .	96
5.2	Features Selection . . . . .	105
5.3	Classification . . . . .	107
5.3.1	Classifying Mammographic Masses . . . . .	107
5.3.2	Classifying BMRI Lesions . . . . .	109
5.3.3	Human Reader Experiments . . . . .	112
5.3.4	MRMC Results for Human Readers . . . . .	116
6	DISCUSSION . . . . .	117
6.1	BMRI Segmentation . . . . .	117
6.1.1	Skin and Chest Wall Segmentation . . . . .	117
6.1.2	Enhanced Lesions Segmentation . . . . .	118
6.2	BMRI Classification . . . . .	119
6.2.1	Features Computation . . . . .	119
6.2.2	Artificial Classifiers . . . . .	119
6.2.3	MRMC Analysis . . . . .	134
6.3	Limitation of the Research . . . . .	135
6.4	Summary . . . . .	136
7	FUTURE WORK . . . . .	138
	REFERENCES . . . . .	140

## LIST OF TABLES

### Table

2.1	Summary of section 2.3 Comparison among different studies used artificial neural networks to classify mammographic masses. . . . .	15
3.1	Consequences of adjusting imaging parameters to increase the signal-to-noise ratio. . . . .	29
3.2	Consequences of adjustments in imaging parameters to increase the spatial resolution. . . . .	29
3.3	Consequences of adjustments in imaging parameters to decrease the acquisition time. . . . .	30
5.1	The unsigned difference between the manual and the automatic segmentation of the chest wall in mm. . . . .	100
5.2	The minimum, maximum, average, and RMS of five different features of the benign lesions. . . . .	105
5.3	The minimum, maximum, average, and RMS of five different features of the malignant lesions. . . . .	106
5.4	The area under each of the ROC curves ( $A_z$ ), the standard error (SE), $p$ value, and the 95% confidence interval of the area. . . . .	107
5.5	The area under the ROC curve ( $A_z$ ), the standard error (SE), $p$ value, and the 95% confidence interval of the area for the BNN system with different $N_h$ . . . . .	109
5.6	The area under the ROC curve ( $A_z$ ), the standard error (SE), $p$ value, and the 95% confidence interval of the area for each classifier. . . . .	110
5.7	The area under the ROC curve $A_z$ , the standard error (SE), $p$ value, and 95% of the confidence interval of the area for the BMRI expert. . . . .	112
5.8	The area under the ROC curve ( $A_z$ ), the standard error (SE), $p$ value, and the 95% confidence interval of the area for the two mammographers. . .	113
5.9	The area under the ROC curve ( $A_z$ ), the standard error (SE), $p$ value, and the 95% confidence interval of the area for the two body imaging fellows. . .	116



6.1	The differences in the BNN system performance for different number of hided neurons in terms of ( $A_z$ ) along with the $p$ value associated with each difference. . . . .	124
6.2	The differences among the three classifiers in terms of ( $A_z$ ) along with the $p$ value associated with each difference. . . . .	125
6.3	The difference between area under ROC curve ( $A_z$ ) between the BMRI expert with and without the BNN assistance vs. the BNN system classification. . . . .	130
6.4	The difference between area under ROC curve ( $A_z$ ) between the average mammographer with and without the BNN assistance vs. the BNN system classification. . . . .	130
6.5	The difference between area under ROC curve ( $A_z$ ) between the average for the two fellows with and without the BNN assistance vs. the BNN system classification. . . . .	132

## LIST OF FIGURES

Figure		
3.1	The longitudinal magnetization. . . . .	19
3.2	The spin lattice relaxation time ( $T_1$ ). . . . .	20
3.3	The spin lattice relaxation time ( $T_1$ ) when the net magnetization is placed along the $-Z$ axis. . . . .	20
3.4	The spin-spin relaxation time, $T_2$ . . . . .	21
3.5	General schematic of a three-layer, feed-forward neural network. $p$ is an input node output, $^1b$ is a hidden node bias, $^1n$ is a hidden node input, $h$ is a hidden node output, $^2b$ is an output node bias, $^2n$ is an output node input, and $o$ is the network output. . . . .	31
3.6	2D example of separating data to two classes. . . . .	41
3.7	Maximal margin classifiers. . . . .	42
4.1	The image view showing the pre-contrast and four post-contrast images. Two enhanced regions: The first region is surrounded by a circle, and the second is surrounded by a box (Case: 6922, slice: 18). . . . .	46
4.2	The Hough transform approach starts with marking the two intersection points between the air and the breast skin (the red x marks in the left image), then drawing the line that connects these two points (the red line in the left image), finally applying Hough transform on the right side of the line. (Case: 1606, slice: 3). . . . .	49
4.3	The linear ramp filter that used to highlight the skin air boundary. . . . .	50
4.4	(Case: 1806, slice: 15) a) The original image. b) A blurred image using 2D Gaussian filter. c) The cost function computed from the edge image. d) The skin path found by DP. . . . .	53
4.5	This figure shows Case: 1806, slice: 15. a) The original image with the automatic skin path overlaid on top in the red line. b) The seven regions mask with the breast tissues in yellow and inside the chest in green. . . . .	54
4.6	The pre-contrast and three post-contrast images. The lesion is obvious in the post-contrast images, but is barely visible in the pre-contrast images. ( Case: 5489, slice: 9). . . . .	58

4.7	The difference images between the pre-contrast and each of the three post-contrast images (Case: 5489, slice: 9) . . . . .	59
4.8	The labeled regions using 3D connected component algorithm (Case: 5489, slice: 9) . . . . .	59
4.9	In the top left corner: the original image, followed by the gradient of the three post-contrast sequences computed using Equation 4.3 on page 57. The gradient was able to highlight the blood vessels. (Case: 5489, slice: 26)	61
4.10	In the top left corner: the original image, followed by the three post-contrast sequences after applying the gray-level morphological algorithm. Blood vessels are diminished. (Case: 5489, slice: 26) . . . . .	62
4.11	Spiculation Measurements: $C_m$ is the center of mass, $P$ is a boundary pixel, and $\phi$ is the orientation of $P$ . . . . .	66
4.12	Time/intensity curve of a blood vessel. . . . .	68
4.13	This figure illustrates that some of the datasets we have used do not have sufficient temporal resolution to capture a second wash-in phase. (Case: 5489, slice: 26) . . . . .	69
4.14	Non-linear transfer function: Hyperbolic Tangent Sigmoid used in the hidden and output layers neurons. . . . .	73
4.15	Using two neurons in the hidden layer allow a linear decision boundary as shown in the top, whereas using three neurons allows more complicated decision boundaries such as the bottom case. . . . .	74
4.16	The effect of using L-M algorithm on accelerating network learning. Using L-M methods allows faster movement to the minimum error point. . . . .	80
4.17	The effects of different learning rates $\eta$ on the convergence of a 1-D function. a: $\eta < \eta_{optimal}$ the convergence is guaranteed, but the learning is slow. b: $\eta = \eta_{optimal}$ one learning step is sufficient to reach the global minimum. c: $\eta_{optimal} < \eta < 2\eta_{optimal}$ oscillation, but converging slowly. d: $\eta > 2\eta_{optimal}$ diverge. Adapted from [28]. . . . .	84
4.18	A second order error function, and its linear derivative. The optimal learning rate guarantees the arrival to the optimal weight value, which gives the minimum error in one step. . . . .	85

4.19	Classifying the data using the leaving-one-out method. For each dataset ( $n$ ) in a collection $C$ of $N$ datasets, a new collection $C'$ (of size $N - 1$ ) is created without the $n^{\text{th}}$ dataset. A classifier is trained with $C'$ , and $n$ is evaluated. $\text{Output}_n$ is then the classification of dataset $n$ . . . . .	88
4.20	Every human reader classifies each dataset; for a dataset ( $n$ ), they first evaluate it without CAD assistance ( $\text{HR}_n^-$ ). Then they are shown the results of the CAD system ( $\text{Output}_n$ ), and allowed to re-evaluate their decision ( $\text{HR}_n^+$ ). . . . .	90
4.21	Validation of CAD system. For each dataset ( $n$ ), the classification of the CAD system ( $\text{Output}_n$ ), is compared with the biopsy result ( $\text{Biopsy}_n$ ). This is used to compute the sensitivity (True Positive Rate), and specificity (1-False Positive Rate) of the CAD system. . . . .	92
5.1	The pre-contrast and three post-contrast images. The lesion is marked with a purple box, it is more obvious in the post-contrast images, but is barely visible in the pre-contrast images. (Case: 5828, slice: 20). . . . .	95
5.2	This figure shows a good match between the manual segmentation and the automatic segmentation. (Case: 5520, slice: 12) <i>a</i> ) The blurred image. <i>b</i> ) The cost image. <i>c</i> ) The manual segmentation is the pink line; the automatic segmentation is the cyan line. <i>d</i> ) The automatic segmentation showing the seven regions mask. . . . .	97
5.3	This figure shows what happens when there isn't an obvious boundary. In the upper right portion of the original image, the boundary is very dim. It isn't obvious if it continues on diagonally, or if it switches to a vertical line. Also, if it was a diagonal line, this breaks one of the assumptions that the chest wall boundary is generally vertical. (Case: 5549, slice 20) <i>a</i> ) The blurred image. <i>b</i> ) The cost image. <i>c</i> ) The manual segmentation is the pink line, while the automatic segmentation is the cyan one. <i>d</i> ) The automatic segmentation. . . . .	98
5.4	This figure shows the other common failure in the automatic algorithm for finding the chest wall. When the breast is not entirely within the field of view, the breast mask fails to cover all the breast tissue. Because breast tissue is generally heavily textured, it can appear as a stronger edge in the cost image (edge image) than the chest wall boundary. (Case: 5520, slice 9). <i>a</i> ) The blurred image. <i>b</i> ) The cost image. <i>c</i> ) The manual segmentation is the pink line; the automatic segmentation is the cyan line. <i>d</i> ) The automatic segmentation using DP. . . . .	99

5.5	This graph shows the average boundary segmentation error for each datasets listed in table 5.1 on page 100. The blue bars are the average and standard deviation over all of the datasets. . . . .	101
5.6	The difference images between the pre-contrast and each of the three post-contrast images (Case: 5828, slice: 20). . . . .	101
5.7	The labeled regions using 3D connected component algorithm (Case: 5828, slice: 20). . . . .	102
5.8	The histogram for the individual slice follows a very similar distribution to that of all the slices. Both of them follow a single modal distribution, which makes it difficult to find an optimal threshold. <i>a)</i> The difference image. <i>b)</i> The histogram of that slice. <i>c)</i> The histogram of all the difference images. (Case: 5520, slice: 15). . . . .	102
5.9	The gradient image of a case with a mass. (Case: 5489, slice: 9). . . . .	103
5.10	The morphology image of a case with a mass: The mass is still present, but the blood vessels have been removed. (Case: 5489, slice: 9). . . . .	104
5.11	The time intensity curve of a mass consists of a uptake section and a wash out section. (Case: 5489, slice: 9). . . . .	106
5.12	The ROC curve for classifying the mammographic masses. . . . .	108
5.13	The ROC curves of of the BNN system when four, ten and twenty neurons were used in the hidden layer . . . . .	110
5.14	The ROC curves of the Bayesian Classifier (BC), the support vector machine (SVM), and the backpropagation neural network (BNN) for classifying the BMRI lesions. . . . .	111
5.15	The ROC curve of the breast MRI expert for classifying the BMRI lesions. . . . .	113
5.16	The ROC curves of the two mammographers for classifying the BMRI lesions. . . . .	114
5.17	The ROC curves of the two body imaging fellows for classifying the BMRI lesions. . . . .	115
6.1	MRI signal intensity values are qualitative. This Figure shows a dataset that has an enhanced lesion (red arrow) with intensity of 306. (Case: 1809, Slice: 15). . . . .	120

6.2	MRI signal intensity values are qualitative. This Figure shows a dataset that has an enhanced lesion (red arrow) with intensity of 1179. (Case: 7869, Slice: 8). . . . .	121
6.3	This figure shows that blood vessels (red arrow) enhance very strongly, but it is not a malignancy indication. (Case: 1806, Slice: 3). . . . .	122
6.4	This figure shows a dataset where the heart (red arrow) is visible. (Case: 1610, Slice: 17). . . . .	123
6.5	This figure shows the ROC curves of the BNN system, BMRI expert without using the BNN assistance, and the BMRI expert with using the BNN system assistance for classifying the 80 BMRI lesions. . . . .	129
6.6	This figure shows the ROC curves of the BNN system, the average mammographer performance without using the BNN assistance, and the average mammographer performance with the BNN system assistance for classifying the 80 BMRI lesions. . . . .	131
6.7	This figure shows the ROC curves of the BNN system, the average body imaging fellow performance without using the BNN assistance, and the average fellow performance with the BNN system assistance for classifying the 80 BMRI lesions. . . . .	133

## CHAPTER 1 INTRODUCTION

Breast cancer is the most common cancer among women (excluding non-melanoma skin cancers) and the second leading cause of cancer deaths in women after lung cancer [46]. According to the American cancer society [2], in 2004, approximately 216,000 women and approximately 1,500 men will have breast cancer in the United States of America. Approximately 18% will die from this illness. Breast cancer incidence in women has increased from one in 20 in 1960 to one in eight today. [71]

**The key to surviving breast cancer is early detection and treatment.**

According to the American Cancer Society, when breast cancer is confined to the breast, the five-year survival rate approaches 100%. Not only does early detection improve survival, it also increases the likelihood that the cancer will be amenable to breast conservation therapy, which benefits the patient by reducing the physical and psychological morbidity of therapy. Accurate staging of breast cancer ensures that patients receive the correct treatment. Highly specific breast cancer detection methods minimize the risk of unnecessary biopsy or surgery. Breast MRI (BMRI) has emerged as a promising technique for detecting, diagnosing, and staging breast cancer. Our goal is to help women with breast cancer by improving the performance and confidence of radiologists who interpret BMRI.

### 1.1 Hypothesis and Specific Aims

The overall hypothesis that motivates this work is that **computer-aided diagnosis (CAD) can improve the accuracy of expert and non-expert readers**

## for classifying suspicious lesions on BMRI.

Our plan is to develop and test an automated classification system which improves the performance of radiologists reading breast MR images. We propose these four specific aims:

1. Develop a method to segment BMRI datasets into regions of interest (lesions and ducts) and background.
2. Select and compute the shape, texture, and contrast enhancement pattern of suspicious lesions in BMRI.
3. Devise classifiers to distinguish benign from malignant lesions, using the features identified in aim 2.
4. Assess the performance of the classifiers for expert and non-expert readers by using a set of pathologically-proven lesions from BMRI data.

This thesis contains seven chapters. The first chapter is a general introduction of the work, hypothesis, and specific aims. In the second chapter, the Background and Significance sections discuss the importance of this research and related work. The third chapter explains the fundamentals of magnetic resonance imaging (MRI), breast MRI (BMRI), neural networks, and support vector machines (SVM). In the fourth chapter, we describe our methods for reaching our specific aims; chapter four includes a section for each of the specific aims. Chapter five presents the results of the work, and then we discuss and summarize these results in the sixth chapter. The seventh chapter suggests possible future areas of work related to this research.



## CHAPTER 2 SIGNIFICANCE

The first section of this chapter describes the background, the second section discusses the significance of this research, and the third section is a literature review.

### 2.1 Background

Most breast cancers are diagnosed by mammography, but early cancers can be difficult to detect within the complex architecture of the breast. Research has shown that 10% to 30% of breast MRI (BMRI) visible cancers are overlooked in mammography [14, 38, 70]. Human factors related to perception and interpretation also contribute to false negative mammograms. These factors include a failure to notice a visible cancer, perhaps due to distracting features in the image, and a failure to correctly classify a lesion as a cancer [21]. Classification errors are detrimental to the patient and costly to the health care system.

Liberman [62] showed in a recent study published in The New England Journal of Medicine that the value of mammograms for women at average risk of breast cancer has been hotly debated, though most doctors agree that the test saves lives in women older than 50. Also, this study showed that the benefits of better detection from MRI outweigh its drawbacks for those with a strong family history of breast cancer or faulty genes.

The most useful MRI technique for breast imaging uses a contrast material called gadolinium chelate, which is injected into a vein in the arm before or during the exam to improve the quality of the images. The gadolinium travels through

the vascular system and helps produce images with good contrast that “highlight” any abnormalities [47]. During the examination, the breast is imaged once before and several times after the injection of the contrast agent. The first series is called the pre-contrast sequence, and the other series are called the post-contrast sequences [14]. The BMRI diagnosis is based on the experimental observation that abnormal tissues (both benign and malignant) absorb more contrast agent than normal tissues because of their increased vascular permeability and angiogenesis [15]. This increased absorption yields an increased MR response, which appears as a region of enhanced grayscale intensity on the MR images. This increased enhancement of the abnormal tissues conveys to the observer the amount of and the speed of the diffusion process of the contrast agent, where both the rate of enhancement after the injection and the shape of the enhancement curve are considered important criterion. MRI is inherently an imaging modality with inconsistent signal intensity, and as a result, MRI signal intensity values are qualitative instead of quantitative. Thus, even if you scan the same patient twice on the same day, you may get different MR image intensity values.

## 2.2 Significance

Early experience with BMRI indicates a high sensitivity and moderate specificity for detection of breast cancer [90]. Moreover, 17-34% of cancers that are visible in BMRI are not detected in mammography [99]. The superior detection sensitivity of BMRI has been ascribed to its tomographic properties, greater soft tissue contrast, and larger field-of-view. Tumors located high on the chest wall or in the axillary tail of the breast are better discerned by BMRI [66, 90]. BMRI also depicts the size

and extent of breast cancers more accurately than either mammography or physical examination [99]. These advantages suggest that BMRI might serve as an adjunct to mammography for detecting and characterizing breast masses [33]. Nonetheless, enthusiasm for BMRI has been tempered by its high cost, limited availability, the large size of BMRI data sets ( $\approx 400$  images per patient), and the greater time required to read BMRI. Thus, it is unlikely that BMRI will completely replace mammography, especially for breast cancer screening in the general population. However, for certain high-risk patients (e.g., *BRCA1/2* positive, strong family or personal history of breast cancer, etc.), BMRI may become the preferred screening modality.

Kriege *et al.* [52] reported in a study of clinical breast examination, mammography, and MRI in 1909 women who had a genetic or familial predisposition to breast cancer lifetime risk  $\geq 15\%$  in the Netherlands. Of these women, 19% had *BRCA* mutations. Kriege *et al.* provided data on almost twice as many patients and twice as many mutation carriers as were included in all previously published evaluations of MRI in high-risk patients combined. Those who interpreted the MRIs and mammograms were unaware of the results of the other technique. The investigators analyzed their data in subgroups according to quantified levels of risk. Their study confirms the high sensitivity of MRI in identifying invasive breast cancer in high-risk patients.

Kriege *et al.* found that the breast-cancer detection rate was 9.5 cases per 1000 woman-years of follow-up overall: 7.8 cases per 1000 women with a 15 to 29% lifetime risk, 5.4 per 1000 for those with a 30 to 49% lifetime risk, and 26.5 per 1000 for carriers

of *BRCA1* or *BRCA2* mutations. Among 45 cancers, 49% were identified by MRI but not mammography, 22% were identified by both MRI and mammography, 18% were identified by mammography but not MRI. Of these 45 tumors, 4 were cancers, and 1 was identified by clinical examination only. Certain features appeared in more than half of the cancers in mutation carriers: they were diagnosed in women between the ages of 30 and 39 years; they were invasive cancers; and the tumors were of high nuclear grade, estrogen receptor-negative, and node-negative. Only 17% of cancers in mutation carriers were interval cancers. In their analyses, MRI, as compared with mammography, had higher sensitivity (71% vs. 40%) but lower specificity (90% vs. 95%).

Kriege *et al.* report that short-term follow-up MRI was recommended in 7% of examinations, as compared with 10 to 25% in prior reports [53]. MRI had limited sensitivity (17%) in detecting ductal carcinoma in situ; in prior studies, the sensitivity of MRI for this type of lesion ranged from 0% [98] to 100% [53,77]. Kriege *et al.* also reported that MRI had lower specificity than mammography, but Kuhl *et al.* [53] found that MRI had higher sensitivity and specificity than mammography. But we believe that **refinement and standardization of MRI technique and interpretation may improve specificity while retaining high sensitivity**. Not addressed by Kriege *et al.* is the potential role of ultrasonography in screening high-risk women. In studies that supplemented mammography with both MRI and ultrasonography, MRI had higher sensitivity and specificity than ultrasonography and was superior in detecting ductal carcinoma in situ. [53, 77, 98]

Mammographers seem the best qualified, and certainly the most likely to read BMRI, but they are accustomed to reading 4-6 images per patient as compared with the 200-400 images per patient that BMRI generates. Also, some mammographers may be uncomfortable interpreting MR images. A user-friendly computerized system for BMRI image presentation and lesion classification may help mammographers overcome their objections and anxieties concerning BMRI. Such a system would reduce the time required to read BMRI, making the technique a more practical clinical tool.

The diagnostic accuracy of mammography is increased when two radiologists examine the same mammogram, or when the same radiologist re-reads a mammogram [96]. These strategies decrease the number of missed cancers. The rationale for CAD systems is to replace the second pair of human eyes with a set of “electronic eyes.” The development of dependable, low-cost CAD systems for breast lesion detection and classification is of great practical interest [32,95].

### 2.3 Literature Review

Researchers have developed mammography CAD systems for detecting breast tumors [22, 63, 64, 69, 76, 94, 102] and microcalcifications [34]. Many classification methods have already been developed and applied to characterize mammographic breast masses as benign or malignant. These methods include wavelets [18, 81, 97], fractals [56, 58], statistical methods [17, 48], vision-based methods [79] and, recently, artificial neural networks (ANN) [3,5,7,20,22,25,34,55,59,63,64,69,76,87,94,100–103]. ANN-based methods are especially popular [101] because they can be implemented

using parallel processing to efficiently handle large amounts of information, they can analyze and classify patterns even when the information is noisy, complex or incomplete, and they can adapt their behavior to a variety of training data. ANN have been used to detect and classify microcalcifications [25, 100, 103] and breast masses on mammography [3, 5, 7, 20, 22, 25, 34, 55, 59, 60, 63, 64, 69, 76, 87, 94, 100–103].

Several research groups have reported encouraging results in using neural networks to classify breast masses on mammography [7, 13, 19, 25, 30, 45, 54, 55, 87].

Arbach *et al.* [5] compared mammographic mass classification performance between a backpropagation neural network (BNN), expert radiologists, and residents. The goal was to reduce false negatives during routine reading of mammograms. The database consisted of 160 cases from 3 different institutions. Each case contained at least one mass and had an accompanying biopsy result. Ten texture and shape based features (area, perimeter, compactness, radial length, spiculation, mean/standard deviation of radial length, minimum/maximum axis, and boundary roughness) were used as inputs to a three-layer neural network. 140 cases were used for training the BNN and the remaining 20 cases were used for testing. The testing set was diagnosed by three expert radiologists, three residents, and the BNN.  $A_z = 0.923$  for the BNN, 0.846 for the expert radiologists, and 0.648 for the residents.

Shih-Chung *et al.* [87] designed a multiple circular path convolution neural network system (MCPCNN) for detection of mammographic masses on 91 patients. They first divided each suspected tumor area into sectors and computed the defined mass features for each sector independently. These sector features were used on the

input layer of the neural network. The masses were divided into two categories. The first one is small, well defined, and smoothed edges masses. And the second one is large, poorly defined, and spiculated masses. The overall performance was 0.78 for the area under the ROC curve ( $A_z$ ). This group used an average filter in the pre-processing step, which helped in losing important information about the edges of the masses. Also, they used 125 neurons in the hidden layer of the neural network. The consequence of too many hidden layer neurons results in network memorization so that they do not properly generalize well to the validation and test set environment data.

Land *et al.* [55] developed a neural network for improving the benign/malignant diagnosis of breast cancer using mammogram findings (MLFN). They developed a machine learning paradigm (AB) focuses on finding weak learning algorithm(s) that initially need to provide slightly better than random performance (i.e., approximately 55%) when processing a mammogram training set. Then, by successive development of additional architectures (using the mammogram training set), the adaptive boosting process improved the performance of the basic evolutionary programming derived neural network architectures. The results of these several EP-derived hybrid architectures were then intelligently combined and tested using a similar validation mammogram data set. Optimization focused on improving specificity and positive predictive value at very high sensitivities, where an analysis of the performance of the hybrid would be most meaningful. Using the Duke mammogram database of 500 biopsy proven samples, on average this hybrid was able to achieve an  $A_z=0.85$ . This

group used one neuron in the hidden layer of their neural network. This would reduce the ability of the system to training. Shape features only were used as an input to the Feed-forward neural network. Using Feed-forward training algorithms usually done by training the system with one of the inputs at the time, which forces the network to calculate each input weights independently from the others. This would be in most cases not the best sets of weights to be used.

Bruce *et al.* [13] used a neural network with Wavelet transforms for an automated mammographic mass classification (NNWT). The discrete Wavelet transform was applied to the radial distance measure of the mass shape, and scalar energy feature was computed to form the two layers network inputs. The system was trained on 23 mammograms and tested on a simulated mammogram. The performance of this neural network was  $A_z=0.87$ .

Huo *et al.* [45] developed a computer-aided diagnosis method for differentiating malignant from benign masses from mammograms (CADM). The four computer extracted features they used were spiculation, margin sharpness, and two density-related measures. These features were merged using either an artificial neural network or a hybrid method consisting of a rule-based system followed by an artificial neural network. The method was trained with 95 mammograms containing masses from 65 patients. Various features related to the margin and density of each mass were extracted automatically from the image data and merged into an estimated likelihood of malignancy. The round-robin performance of the computer in distinguishing between benign and malignant masses was evaluated by ROC analysis. This group computer



classification scheme yielded an  $A_z$  value of 0.94, similar to that of an experienced mammographer ( $A_z = 0.91$ ) and statistically significantly higher than the average performance of five radiologists with less mammographic experience ( $A_z = 0.81$ ). With the database they used, the computer scheme achieved, at 100% sensitivity, a positive predictive value of 83%, which was 12% higher than that of the experienced mammographer and 21% higher than that of the average performance of the less experienced mammographers at a  $p$ -value of less than 0.001.

Enderwick *et al.* [30] performed a classification on a biopsy-proven malignant masses using a mix of shape and texture features. Shape features included size, translation, and rotation invariant moments. Texture features included fractal-based features and spatial gray level dependence (SGLD) matrix features. The entropy was also computed for each mass. The type of classifier used was a neural network based on the ALOPEX training algorithm, which depends on stochastic optimization method. The dataset consisted of 90 masses for the training stage and 14 masses for testing stage. Because of lack of storage area, the masses were reduced by half to 256\*256 pixels. The  $A_z$  for this experiment was 0.79.

Kocur *et al.* [54] designed a computer-aided diagnosis system to be used as a second opinion to radiologists, in order to aid in decreasing the number of false readings of mammograms (CADMP). Neural networks were used at both the classification and feature selection stages in the development of a computer-aided breast cancer diagnostic system. The multilayer perceptron was used to classify and contrast three features (angular second moment, eigenmasses, and wavelets) developed

to distinguish benign from malignant lesion in a database of 94 difficult-to-diagnose digitized microcalcification cases. System performance of 74% correct classifications is achieved. Feature selection techniques were presented which further improved performance. Neural and decision boundary-based methods were implemented, compared, and validated to isolate and remove useless features. The contribution from this analysis was an increase to 88% correct classification in system performance.

Cheng *et al.* [19] used a multilayer back-propagation neural network (MBNN) to classify the mammograms into three risk groups. They employed a feature extraction algorithm based on the fuzzy co-occurrence matrix concept, and used it to deal with early and accurate breast cancer diagnosis by analyzing microscope-slide biopsy images. The data base for this study consisted of 90 mammograms; 60 of them were used for training the three layer neural network with four neurons in the hidden layer. This system gave  $A_z=0.762$ .

Dhawan *et al.* [25] classified mammographic microcalcifications using artificial neural network and image structure features (CMMNN). They presented a second-order gray-level histogram based feature extraction approach to extract microcalcification features. These features, called image structure features, were computed from the second-order gray-level histogram statistics, and did not require segmentation of the original image into binary regions. Several image structure features were computed for 100 cases of “difficult to diagnose” microcalcification cases with known biopsy results. These features were analyzed in a correlation study which provided a set of five best image structure features. A feedforward neural network was used to

classify mammographic microcalcifications using the image structure features. The network was trained on 10 cases of mammographic microcalcifications and tested on additional 85 ‘difficult-to-diagnose’ microcalcifications cases using the selected image structure features. The trained network yielded  $A_z=0.74$ . This group used only 10% of the datasets for training, which made the system incapable of learning from enough examples. Also, they used Feed-forward training algorithm which does not give the best results compared with feed-back algorithm.

Table 2.1 on page 15 shows a comparison among the different systems mentioned above.

### 2.3.1 Penn 2001

In 2001, Alan Penn, *et al.* [75] published a paper about the first and second revisions of their BMRI CAD system. They used 95 contrast-enhanced MR images, and had an expert locate the suspicious lesion for each case. The expert chose a representative slice for each case, and manually overlaid an arrow indicating where the lesion was located.

They had 5 human readers score each of the masses on five different features: border shape, rim enhancement, septation, intensity and density. Each reader assessed the likelihood-of-cancer using a 1-5 scale with 1 being definitely benign, and 5 being definitely malignant. The readers also supplied a follow-up action of either biopsy, repeat MR in 3-6 months, or return to routine mammography.

Readers selected a region of interest for each lesion, and this region was used to compute a fractal-dimension feature. This feature was heuristically thresholded into 2

categories, strong-indication-of-malignancy, and not strong-indication-of-malignancy.

After the readers had scored each of the cases, they used logical regression to compute a correlation between each feature and the likelihood of malignancy.

After 30 days, each reader was presented with the cases again. This time, they were also presented with the fractal dimensional feature, and the likelihood-of-malignancy measure. It should be noted that the users were allowed to change their scorings and see how this would affect the likelihood-of-malignancy.

They were able to show a statistically significant improvement in the classification results for 2 out of the 5 readers. However, for one of those readers there was actually a statistically significant decrease in sensitivity. This reader had an overall improvement in their accuracy from a vast improvement in their specificity.

As you can see, there is room for improvement in breast MR CAD systems.

Table 2.1: Summary of section 2.3 Comparison among different studies used artificial neural networks to classify mammographic masses.

Study Name	Reference Number	Number of Mammograms	Network Type	Network Layers	Number of Hidden Layer Neurons	Training Algorithm	Features Used	$A_z$
BNN	[5]	160	Nonlinear	3	3	Backpropagation	10 shape-based texture-based	.923
MPCNN	[87]	91	Nonlinear	3	125	Feed-forward	3 shape-based	.78
MLFN	[55]	500	Nonlinear	3	1	Feed-forward	5 shape-based	.85
NNWT	[13]	23	Nonlinear	2	-	Backpropagation	2 shape and texture-based	.87
CADM	[45]	95	Nonlinear	3	5	Backpropagation	4 shape and texture-based	.94
SGLD	[30]	104	Nonlinear	5	20	ALOPEX	4 shape and texture-based	.79
CADMP	[54]	94	Linear	5	-	Feed-forward	4 shape and texture-based	.88
MBNN	[19]	90	Nonlinear	3	4	Backpropagation	3 shape-based texture-based	.76
CMMNN	[25]	100	Nonlinear	4	-	Feed-forward	5 texture-based	.74

## CHAPTER 3 FUNDAMENTALS

In this chapter, we will discuss the fundamentals related to this research. In the first section, we will describe MRI basics. Breast MRI principles and techniques will be discussed in the second section. In the third section, we will talk about artificial neural networks (ANN) and their basic concepts. And in the last section, we will describe the support vector machine (SVM).

### 3.1 MRI

Magnetic resonance imaging (MRI) is an imaging technique used primarily in medical settings to produce high quality images of the inside of the human body [43]. MRI is based on the principles of nuclear magnetic resonance (NMR), a spectroscopic technique used by scientists to obtain microscopic chemical and physical information about molecules. The technique was called magnetic resonance imaging rather than nuclear magnetic resonance imaging (NMRI) because of the negative connotations associated with the word nuclear in the late 1970's. MRI started out as a tomographic imaging technique, that is, it produced an image of the NMR signal in a thin slice through the human body. MRI has advanced beyond a tomographic imaging technique to a volume imaging technique. The intensity of a pixel is proportional to the NMR signal intensity of the contents of the corresponding volume element or voxel of the object being imaged.

Magnetic resonance imaging is based on the absorption and emission of energy in the radio frequency range of the electromagnetic spectrum. The human body is

primarily fat and water, approximately 63% hydrogen atoms. Magnetic resonance imaging primarily images the NMR signal from the hydrogen nuclei. The proton (in a hydrogen atom) possesses a property called spin that will cause the nucleus to produce an NMR signal.

### 3.1.1 Spin

Spin is a fundamental property of nature for protons, electrons, and neutrons just like electrical charge or mass. Spin comes in multiples of  $1/2$  and can be positive or negative. Individual unpaired electrons, protons, and neutrons each possess a spin of  $1/2$ . In the deuterium atom ( $^2H$ ), with one unpaired electron, one unpaired proton, and one unpaired neutron, the total electronic spin =  $1/2$  and the total nuclear spin = 1. Two or more particles with spins having opposite signs can pair up to eliminate the observable manifestations of spin.

**Properties of Spin** When placed in a magnetic field of strength  $B$ , a particle with a net spin can absorb a photon, of frequency  $\nu$ , which depends on the gyromagnetic ratio.

$$\nu = \gamma B. \quad (3.1)$$

For  $^1H$ ,  $\gamma = 42.58 \text{ MHz / T}$ .

### 3.1.2 $T_1$ Processes

At equilibrium, the net magnetization vector lies along the direction of the applied magnetic field  $B_0$  and is called the equilibrium magnetization  $M_0$ . In this

configuration, the Z component of magnetization  $M_z$  equals  $M_o$ .  $M_z$  is referred to as the longitudinal magnetization. There is no transverse ( $M_x$  or  $M_y$ ) component of magnetization here, as shown in Figure 3.1 on the next page. It is possible to change the net magnetization by exposing the nuclear spin system to energy of a frequency proportional to the energy difference between the spin states ( $E = h\nu$ ,  $h$  is Planck's constant). If enough energy is put into the system, it is possible to change all longitude magnetization into transverse magnetization ( $M_z=0$ ). The time constant which describes how  $M_z$  returns to its equilibrium value is called the spin lattice relaxation time ( $T_1$ ). The equation governing this behavior as a function of the time  $t$  after its displacement is:

$$\frac{dM_z}{dt} = \frac{M_o - M_z}{T_1}. \quad (3.2)$$

When the initial  $M_z = 0$ , the solution to this equation is:

$$M_z = M_o(1 - e^{-\frac{t}{T_1}}). \quad (3.3)$$

See Figure 3.2 on page 20 for illustration of the spin-lattice relaxation.  $T_1$  is therefore defined as the time required to change the Z component of magnetization by a factor of  $1 - \frac{1}{e} = 0.63$ . If the net magnetization is placed along the -Z axis, it will gradually return to its equilibrium position along the +Z axis at a rate governed by  $T_1$ . The equation governing this behavior as a function of the time  $t$  after its displacement is:



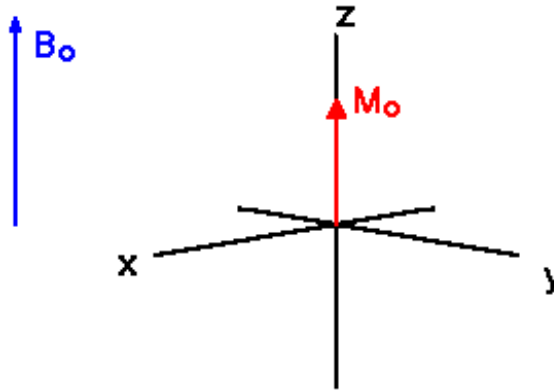


Figure 3.1: The longitudinal magnetization.

Source: [43]

$$M_z = M_o(1 - 2e^{-\frac{t}{T_1}}). \quad (3.4)$$

Figure 3.3 on the next page shows the spin lattice relaxation time ( $T_1$ ) when the net magnetization is placed along the  $-Z$  axis.

The spin-lattice relaxation time ( $T_1$ ) is the time to reduce the difference between the longitudinal magnetization ( $M_z$ ) and its equilibrium value by a factor of  $e^{-1}$ .

### 3.1.3 Precession

If the net magnetization is placed in the  $XY$  plane it will precess about the  $Z$  axis at a frequency equal to the frequency ( $\nu = \frac{\Delta E}{h}$ ) of the photon which would cause a transition between the two energy levels of the spin. This frequency is called the Larmor frequency.

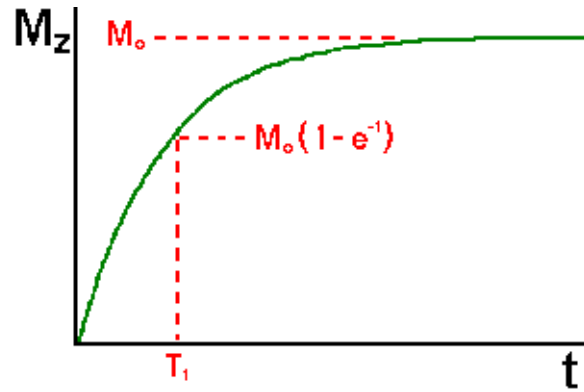


Figure 3.2: The spin lattice relaxation time ( $T_1$ ).

Source: [43]

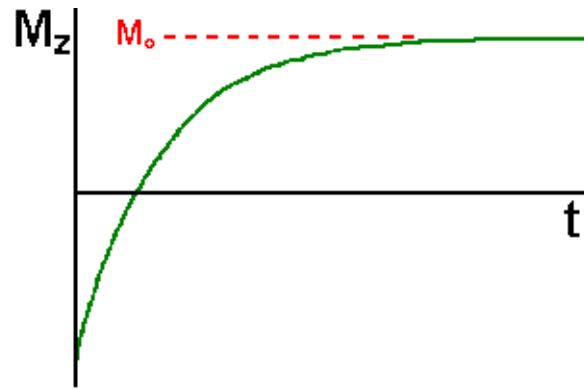


Figure 3.3: The spin lattice relaxation time ( $T_1$ ) when the net magnetization is placed along the  $-Z$  axis.

Source: [43]

#### 3.1.4 $T_2$ Processes

The time constant which describes the return to equilibrium of the transverse magnetization,  $M_{xy}$ , is called the spin-spin relaxation time,  $T_2$  as shown in Figure 3.4 on the following page.

$$M_{xy} = M_{xy0} e^{-\frac{t}{T_2}}. \quad (3.5)$$

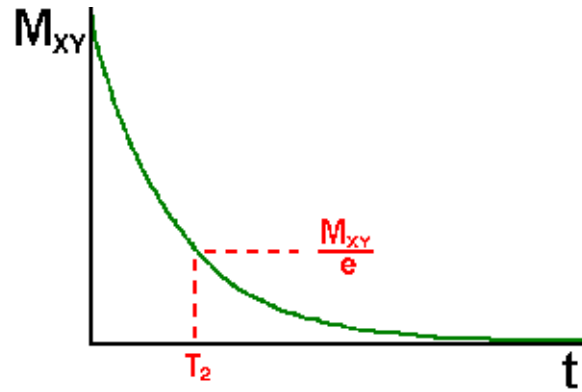


Figure 3.4: The spin-spin relaxation time,  $T_2$ .  
Source: [43]

$T_1$  and  $T_2$  relaxation occur simultaneously, and they are dependant on the physiochemical properties of the tissues. The spin-spin relaxation time  $T_2$  is less than or equal to the spin-lattice relaxation time  $T_1$ . The spin-spin relaxation time,  $T_2$ , is the time to reduce the transverse magnetization by a factor of  $1/e$ . There is another constant called  $T_2^*$ , which is the combination of  $T_2$ , the effect of molecular interaction, and the dephasing effect caused by inhomogeneities in the fixed field  $B_0$ . The relationship is as follows:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{2,inhomo}}. \quad (3.6)$$

### 3.1.5 Imaging Parameters

Imaging parameters are used to control contrast between tissues. The following subheadings are the principals parameters that the operator may change and the effects of each parameter on the image contrast, resolution, etc. It should be noted

that many of these parameters must be considered simultaneously, since their values are often interdependent.

## Definitions

1. **RF Pulse:** A brief burst of electromagnetic energy. The energy is applied at the Larmor frequency, changing the direction of the net magnetization vector (longitudinal magnetization becomes transverse, and transverse magnetization becomes negative longitudinal magnetization). There are two common RF pulses, a  $90^\circ$  pulse is timed to cause a longitudinal magnetization vector to become a transverse magnetization vector, and a  $180^\circ$  pulse, which changes the direction of the magnetization vector.
2. **Dephasing:** The process by which two signals of nearly identical frequencies change from being in-phase (constructive interference) to being out of phase (destructive interference).
3. **Phase Encoding:** One method for localization of MR signals. Once the spins have been excited by a radiofrequency RF pulse, the net xy-magnetization within each voxel precesses according to the Larmor equation which says that precession at higher field strengths is faster. If a gradient field is briefly switched on and then off again at the beginning of the pulse sequence right after the radiofrequency pulse, the voxel magnetizations will either precess faster or slower relative to the ones of the central voxels.

4. **Repetition Time (TR):** Time between successive applications of RF excitation pulses.
5. **Echo Time (TE):** Time between the excitation pulse and the formation of the echo.
6.  **$T_1$  and  $T_2$  Contrasts:**  $T_1$  -weighted image is formed by imaging relatively rapidly, with a  $TR \leq T_1$  of the tissues being imaged. In  $T_1$  weighted images fluids tend to appear dark because of their long  $T_1$  values.  $T_2$  -weighted imaging is achieved by extending TR, such that most tissues do recover most of their z magnetization between successive RF excitation pulses. Tissues with a long  $T_2$  will appear as a bright areas of high signal.  $T_1$  ( $T_2$ ) contrast: is using the difference in tissues'  $T_1$  ( $T_2$ ) constants to create a visual difference in the MR image.

**Repetition Time (TR):** Affects  $T_1$  contrast, scanning time, minimum echo time (TE) and the number of slices in multi-slice imaging. The TR controls the extent of recovery of the longitudinal magnetization. If the TR is long the protons can realign completely between excitations and produce a strong signal. When the TR is short compared with the  $T_1$  relaxation time the equilibrium magnetization is smaller because of its partial recovery. As a result,  $T_1$  contrast is high but the signal is lower than at longer TR, resulting in a reduced signal-to-noise ratio (SNR).

**Echo Time (TE):** Affects  $T_2$  contrast, the minimum possible value for TR, the minimum bandwidth, and the sensitivity to dephasing. A long TE allows more transverse relaxation to occur before measuring the signal. For a given TE, the signal strength of tissues with long  $T_2$  values decreases less than that from tissues with short  $T_2$ . Tissues with long  $T_2$  values appear relatively brighter when TE is longer (signal always drops as TE increases).

**Field of View (FOV):** Horizontal or vertical size of an image. FOV is chosen to match the size of the anatomic area of interest. Its minimum value is determined by the maximum magnetic field gradient strength of the MR system.

Affects the spatial resolution, which increases with decreasing FOV, and SNR ( $SNR \propto FOV^2$ ).

**Number of Excitations (NEX):** Number of times each phase-encoding step is performed. It is also referred to as the number of signal average (NSA).

Affects the SNR and the acquisition time, both of which increase with an increasing NEX.

### 3.2 Breast MRI

Over the last decade we have seen breast MRI develop from a technique appropriate only to a limited number of research institutions, to a clinically useful imaging tool. Breast MRI currently has a role in some clinical situations (e.g. evaluation of breast prostheses, assessment of the post-surgical breast) and is rapidly defining its role in others. As breast MRI expands from a research tool to a clinical situation, it is

clear that the driving force for future development of the technique will lie not with researchers with an interest in a particular imaging modality (MRI), but with clinicians involved in the whole scope of breast diagnosis, supported by the multi-disciplinary breast team.

The MR imaging technique is highly sensitive to physical, chemical and biological characteristics of tissues and fluids. The ability of the radiologists to visualize and detect abnormality may be affected by [36]:

1. Image contrast,
2. Signal-to-noise ratio,
3. Spatial resolution,
4. Temporal resolution,
5. Artifacts.

These factors affecting image quality are in turn determined by a large number of tissue and imaging parameters, in addition to the inherent capabilities of the particular MR imaging system used. Optimization of an MR protocol is therefore complex and will be governed to some extent by the motivation for the examination (for example, detection of disease recurrence, screening) and in part by the nuclear magnetic resonance (NMR) characteristics of the tissues to be imaged [36].

### 3.2.1 Uptake of Contrast Agent

In addition to inherent differences in tissue relaxation times, which are the principal source of signal contrast in MR images, exogenous contrast agents are often administered clinically to enhance the increased conspicuity of certain tissues [91].  $T_1$ -shortening agents such as gadopentetate dimeglumine (Gd-DTPA) are used when the diseased tissue and normal tissue have similar MR properties but respond differently to the contrast agent. The reasons for contrast agent use in the breast include [86]: improvement in sensitivity and specificity, better differentiation of structures, measurement of tissue perfusion, and conspicuity of cancers.

MRI is able to distinguish benign from malignant tissues by exploiting differences in contrast agent kinetics between these tissue types. All clinical breast MRI studies use one of the commercially available extracellular contrast agents. These agents are freely diffusible, that is, they readily pass from the intravascular space into the interstitium and therein give rise to parenchyma enhancement [74].

### 3.2.2 Optimization of Sequence Protocols

This section deals with image quality issues and how imaging factors interrelate to affect image quality. The following image quality characteristics are considered: signal-to-noise ratio (SNR), tissue contrast, spatial resolution, temporal resolution, and artifacts.



### 3.2.2.1 Signal-To-Noise Ratio (SNR)

Table 3.1 on page 29 summarizes methods of increasing image SNR and describes their effects on other image quality factors. It is worth noting that for the same volume of coverage and the same contrast, 3D sequences produce images with a higher SNR than 2D sequences, albeit with a longer acquisition time [39].

### 3.2.2.2 Image Contrast

The imaging parameters determining image contrast are TR, TE, flip angle, and flow compensation. Some priori knowledge of tissue parameters is useful when choosing imaging parameters.

### 3.2.2.3 Spatial Resolution

Methods to increase the spatial resolution are summarized in Table 3.2 on page 29, together with relative disadvantages. In designing a protocol for breast MRI, optimization of the spatial resolution will be determined by the nature of the examination. An increase in spatial resolution will increase acquisition time and therefore reduce temporal resolution, hence there is a need to balance the examination requirements with the capabilities of the MR system [91].

### 3.2.2.4 Temporal Resolution

In general, the time required to obtain data for one image is the acquisition time (TA), is determined by the number of phase-encoding steps  $N_{ph}$ , the repetition

time of the sequence (TR) and the number of excitations (NEX) [36]:

$$TA = TR \cdot N_{ph} \cdot NEX. \quad (3.7)$$

A reduction in the acquisition time is desirable for a number of reasons, including dynamically following a contrast agent as it passes through the vascular system and/or through a tissue, and also for minimizing artifacts due to motion. Each of the parameters in the above equation may be reduced in order to increase the temporal resolution. The consequences of reductions in these parameters on other indicators of image quality are summarized in Table 3.3 on page 30.

### 3.2.2.5 Motion Artifacts

Motion artifacts occur mainly due to:

1. Natural motion of body (heart, lungs, etc),
2. Unintended patient motion.

MR images are acquired over a period of time (hundreds of milliseconds to several minutes) that is too long to “freeze” the movement of anatomic structures. MR examinations, and in particular dynamic contrast-enhanced MR imaging, may require the patient to remain in the scanner for 30 min or longer. During this time, the patient may also move, compromising, for example, the quality of subtracted images.

Motion during image acquisition blurs the images, reducing the apparent spatial and contrast resolution and hampering image intensity calculations. Methods to

Table 3.1: Consequences of adjusting imaging parameters to increase the signal-to-noise ratio.

Parameter Change	Disadvantages
Increase FOV, slice thickness	Decrease spatial resolution
Increase NEX	Increase the acquisition time
Decrease receiver bandwidth	Increase motion artifacts increase minimum possible echo time

Note: FOV - field of view, NEX - number of excitations. Adapted from [36].

Table 3.2: Consequences of adjustments in imaging parameters to increase the spatial resolution.

Parameter Change	Disadvantages
Increase matrix size, phase-encoding direction	Increase acquisition time
Increase matrix size, frequency-encoding direction	Decrease bandwidth Increase artifacts Increase minimum echo time
Decrease field of view	Decrease signal-to-noise ratio Increase artifacts

Adapted from [36].

overcome motion artifacts include breast compression, spatial pre-saturation, ghost re-direction, gating, and flow compensation [44].

**Summary** MRI is a good noninvasive modality for producing images from human or any biological system that travels the structure, metabolism, and function of internal tissues and organs extending the range of human vision [61]. MRI can produce images at higher speed (fast imaging), higher resolution (micro-imaging), and higher

Table 3.3: Consequences of adjustments in imaging parameters to decrease the acquisition time.

Parameter Change	Disadvantages
Decrease TR	Decrease SNR
Decrease number of phase encoding steps	Decrease spatial resolution
Decrease number of slices	Decrease spatial coverage
Increase bandwidth	Decrease the SNR

Adapted from [36].

information content (combined anatomy, metabolism, and function). It will have high impact in biology and neuroscience in the coming years [61].

### 3.3 Neural Network Design

Learning from examples, the problem for which neural networks were designed to solve, is one of the most important research topics in artificial intelligence [37]. A possible way to formalize learning from examples is to assume the existence of a function representing the set of examples and, thus, enabling the function to generalize to yet unseen examples. This can be called function reconstruction from sparse data (or in mathematical terms, depending on the required precision, approximation or interpolation, respectively). Within this general framework, the central issues of interest are the representational power of a given network model (or, in other words, the problem of model selection) and the procedures for obtaining the optimal network parameters. The tasks of parameter estimation and model selection are of crucial importance for the success of real world neural network applications. Model selection or the specification of a network topology is a key methodological issue.

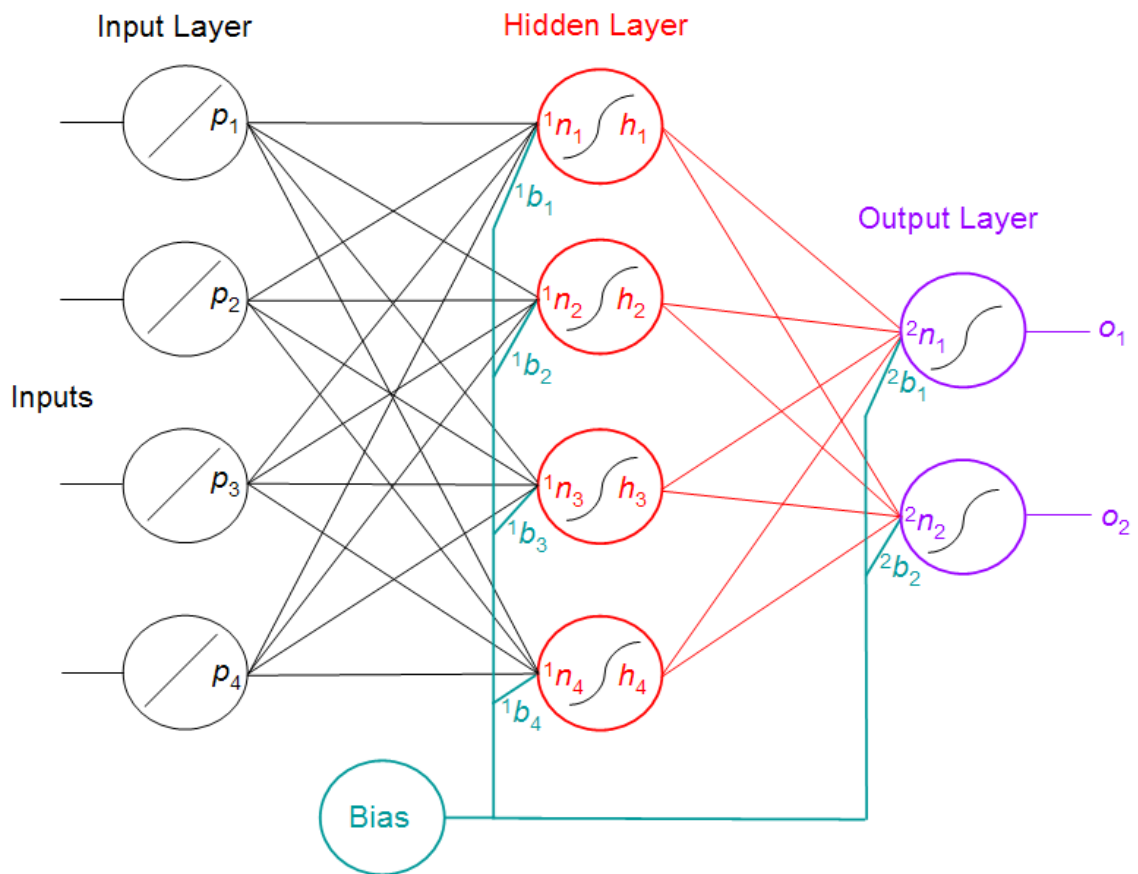


Figure 3.5: General schematic of a three-layer, feed-forward neural network.  $p$  is an input node output,  ${}^1b$  is a hidden node bias,  ${}^1n$  is a hidden node input,  $h$  is a hidden node output,  ${}^2b$  is an output node bias,  ${}^2n$  is an output node input, and  $o$  is the network output.

Figure 3.5 on the preceding page shows a diagram of a three-layer neural network. Each of the circles in the figure is called a unit (or neuron), and represents an input-output transfer function. Neural network researchers have chosen from a variety of transfer functions; among the more popular ones are the logistic-sigmoid and the tangent-sigmoid functions. In this study we use the latter, which, mathematically, is:

$$f(n) = \text{tansig}(n) = \frac{e^n - e^{-n}}{e^n + e^{-n}}, \quad (3.8)$$

where both the input,  $n$ , and the output,  $f(n)$ , are real scalar-values. In vector form,  $\vec{y} = f(\vec{n})$ , each element of  $\vec{n}$  is evaluated, so that  $y_i = f(n_i)$  for the  $i^{\text{th}}$  element of  $\vec{y}$ . The columnar arrangements of the units are called layers. From left to right in Figure 3.5 on the page before, the columns are labeled as the input, hidden, and output layers. Only the hidden and output units perform the nonlinear (tangent-sigmoid) mapping; the input units have identity transfer functions. In our exposition we focus on a three-layer network because the system used in this work was a three-layer system, but in general, any number of layers can be used. Layers are connected to each other by a system of weights, which multiplicatively scale the values traversing the links. In the diagram, we observe that there are two sets of weights: one connecting the input to the hidden layer, and the other from the hidden to the output layer. The values from weights converging on a given unit are added to form  $n$  [49]. If we think of the inputs as a vector  $\vec{p}$  of dimension  $R$ , then the set of weighted sums that form the input to the hidden layer containing  $S$  units is:

$${}^1\vec{n} = {}^1\mathbf{W}\vec{p} + {}^1\vec{b}, \quad (3.9)$$

where  ${}^1\mathbf{W}$  is an  $S \times R$  matrix, the  $i^{\text{th}}$  row of which forms the set of weights that the  $i^{\text{th}}$  hidden unit uses to scale each element of  $\vec{p}$  (the superscript “1” to the left of the symbol indicates this is for the first layer in the network). The elements of the  $R \times 1$  vector  ${}^1\vec{b}$  are called the bias values of each unit; the bias values add an additional degree of freedom by allowing control of a constant offset that is applied to each input. There is a simplification that can be done to make the notation easier to follow. If the input  $\vec{p}$  is augmented with a single new element  $p_{S+1} = 1$ , then the

bias portion can be combined into the weight vector. This gives:

$${}^1\vec{n} = {}^1\mathbf{W}\vec{p} + {}^1\vec{b} \cdot 1 \quad (3.10)$$

$$= \begin{pmatrix} {}^1w_{1,1} & {}^1w_{1,2} & \dots & {}^1w_{1,R} \\ {}^1w_{2,1} & {}^1w_{2,2} & \dots & {}^1w_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ {}^1w_{S,1} & {}^1w_{S,2} & \dots & {}^1w_{S,R} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_R \end{pmatrix} + \begin{pmatrix} {}^1b_1 \\ {}^1b_2 \\ \vdots \\ {}^1b_S \end{pmatrix} \quad (3.11)$$

$$= \begin{pmatrix} {}^1w_{1,1} & {}^1w_{1,2} & \dots & {}^1w_{1,R} & {}^1b_1 \\ {}^1w_{2,1} & {}^1w_{2,2} & \dots & {}^1w_{2,R} & {}^1b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ {}^1w_{S,1} & {}^1w_{S,2} & \dots & {}^1w_{S,R} & {}^1b_S \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_R \\ 1 \end{pmatrix} \quad (3.12)$$

$$= \left( {}^1\mathbf{W} \mid {}^1\vec{b} \right) \begin{pmatrix} \vec{p} \\ - \\ 1 \end{pmatrix} \quad (3.13)$$

$$= {}^1\mathbf{W}^* \vec{p}^*. \quad (3.14)$$

For the rest of the notation, the \* will be dropped to make the equations easier to read, but all other vectors are of this augmented form. The output of the hidden layer is an  $S \times 1$  vector  $\vec{h}$  given by:

$$\vec{h} = f({}^1\vec{n}) = f({}^1\mathbf{W}\vec{p}). \quad (3.15)$$



The outputs of the hidden layer are fed-forward (a characteristic that gives these networks their name) to the output layer, whose outputs are computed as were those of the hidden layer. Specifically, let  ${}^2\mathbf{W}$  be an  $S + 1 \times O$  matrix, the  $i^{\text{th}}$  row being the set of weights which scale each element of  $\vec{h}$  for the input of the  $i^{\text{th}}$  output unit (The  $S + 1$  row is the bias). The output vector  $\vec{o}$  (of dimension  $O$ ) is thus found by:

$$\vec{o} = f({}^2\vec{n}) = f({}^2\mathbf{W}\vec{h}). \quad (3.16)$$

The overall system essentially performs a nonlinear mapping from an input space to an output space with dimensions  $R$  and  $O$  respectively. The weights and biases are the **adaptive parameters** of the system; modifying them appropriately constitutes the network's learning to approximate a desired function with the same domain and range.

### 3.3.1 Neural Network Learning

Learning methods used for adaptive neural networks (*i.e.*, those networks that are able to change their weights) can be classified into two major categories [37]:

1. **Supervised learning** which incorporates an external teacher, so that each output unit is told what the desired response is for specific input signals. During the learning process global information may be required. An important issue concerning supervised learning is the problem of error convergence, *i.e.*, the minimization of the error between the desired and actual unit output values.

The aim is to determine a set of weights which minimizes the error. One well-known method, which is common to many learning paradigms, is least mean square (LMS) convergence.

2. **Unsupervised learning** uses no external teacher and is based upon only local information. It is also referred to as self-organization, in the sense that the network self-organizes the input data and detects any emergent collective properties.

A neural network could learn off-line if the learning phase and the operation phase are distinct. A neural network learns on-line if it learns and operates at the same time. Usually, supervised learning is performed off-line, whereas unsupervised learning is performed on-line.

As mentioned in section 3.3 on page 30, appropriately modifying the adaptive parameters of the system constitutes the network's learning to approximate a desired function with the same domain and range. Next we discuss this procedure, in more detail.

Suppose we have an input set,  $\mathbf{P}$ , and a set of desired targets,  $\mathbf{T}$ , each with  $Q$  elements:

$$\{\vec{p}_1, \vec{t}_1\}, \{\vec{p}_2, \vec{t}_2\}, \dots, \{\vec{p}_Q, \vec{t}_Q\}. \quad (3.17)$$

We want to choose the network parameters so as to minimize the mean-squared error (MSE) between the actual network output  $\mathbf{O}$  and the desired output  $\mathbf{T}$ :

$$\text{MSE} = F = \frac{1}{Q} \sum_{i=1}^Q \|\vec{t}_i - \vec{o}_i\|^2. \quad (3.18)$$

Note that since the output  $\mathbf{O}$  is a function of the weights, the MSE is a nonlinear surface parameterized by the weights. Training the network is equivalent to searching for the minimum value of the MSE. The MSE,  $F$ , is approximated by iteratively presenting the inputs, finding  $\mathbf{O}$ , and at each of the iterations, computing:

$$\hat{F}(k) = \frac{1}{Q} \sum_{i=1}^Q \|\vec{t}_i(k) - \vec{o}_i(k)\|^2, \quad (3.19)$$

where  $\vec{t}_i(k)$ ,  $\vec{o}_i(k)$ , are the desired and actual output of the  $i^{\text{th}}$  element (sample), at the  $k^{\text{th}}$  iteration (epoch). The learning rule is the method used to update the weights after each of the training iterations. The method used here is the **gradient-descent algorithm**, which chooses its “steps” on the error surface to be the direction in weight space that most rapidly decreases the error. In equation forms, the gradient-descent method updates the weights according to:

$${}^m w_{i,j}(k+1) = {}^m w_{i,j}(k) - \alpha \frac{\partial \hat{F}}{\partial {}^m w_{i,j}}, \quad (3.20)$$

where  ${}^m w_{i,j}$  are the elements of the  $m^{\text{th}}$  layer weight matrix. The parameter  $\alpha$  is known as the learning rate; this controls the size of the weight adaptation increments.

Considering just the update of the weights:

$$\Delta {}^m w_{i,j} = -\alpha \frac{\partial \hat{F}}{\partial {}^m w_{i,j}}. \quad (3.21)$$

The challenge of computing the partial derivatives in the above equations is overcome with an algorithm known as backpropagation [35, 67, 85].

**Backpropagation Algorithm** Training the network, or calculating the updates, proceeds as follows for each training input,  $\vec{p}_j$ , with corresponding desired output,  $\vec{t}_j$ :

1. Calculate the output of each layer using the weights and transfer functions of the units. The network output,  $\vec{o}_j$ , is the output of the last layer.
2. For each layer, the unit sensitivities to changes in the summed inputs  $^m n_i$  are calculated.

For the output layer (layer  $M$ , for a three-layer network  $M = 2$ ), the error does not directly depend on the weights, but using the chain rule:

$$\Delta^m w_{i,j} = -\alpha \frac{\partial \hat{F}}{\partial ^M n_i} \frac{\partial ^M n_i}{\partial ^M w_{i,j}} \quad (3.22)$$

$$\frac{\partial \hat{F}}{\partial ^M n_i} = \frac{\partial \hat{F}}{\partial ^M o_i} \frac{\partial ^M o_i}{\partial ^M n_i} \quad (3.23)$$

$$= (o_i - t_i) f'(^M n_i) \quad (3.24)$$

$$\frac{\partial ^M n_i}{\partial ^M w_{i,j}} = h_j. \quad (3.25)$$

Putting these values back into the Equation 3.21:

$$\Delta^M w_{i,j} = -\alpha (o_i - t_i) f'(^M n_i) h_j \quad (3.26)$$

$$= \alpha (t_i - o_i) f'(^M n_i) h_j. \quad (3.27)$$

So the update for the output node is related to the difference between the current output and the target ( $o_i - t_i$ ), the derivative of the transfer function at the current activation point ( $f'(n_i)$ ), the output of the previous node ( $h_j$ ), and the learning rate ( $\alpha$ ).  $(o_i - t_i)f'(n_i)$  can be thought of as the fractional correction of an input, based on the error of the output (*i.e.* it maps the error of the output of the transfer function to an error of the input of the transfer function).

For the hidden layers ( $m = 1, 2, \dots, M - 1$ ), the update is dependent upon the next layer's ( $m + 1$ ) update. This is what gives the term backpropagation. The error at the output is computed, and used to update the weights for the output layer. The error is then propagated back to the previous layer:

$$\Delta^m w_{i,j} = \alpha \left[ \sum_{k=1}^{m+1} w_{k,j} (o_k - t_k) f'(n_k) \right] f'(n_j) p_i. \quad (3.28)$$

To break this into its parts,  $w_{k,j} (o_k - t_k) f'(n_k)$  is the error in the weight from the  $j^{\text{th}}$  hidden node to the  $k^{\text{th}}$  output node. The summation gives the total error of the output of the  $j^{\text{th}}$  hidden node. This is mapped back to the error of the input of the  $j^{\text{th}}$  hidden node by multiplying it by the derivative of the transfer function. Then the weight to the input node is updated based on the learning rate ( $\alpha$ ) and the current value of the input ( $p_i$ ).

The training continues until a criterion has been satisfied (for example a sufficiently low training MSE); Presentation of the whole training set is known as an

epoch.

**Practical Notes** A larger number of neurons allow the system to approximate functions of greater complexity. The caveat, however, is that while more units (and, consequently, more parameters) can fit the training data better, the system will likely be more “whimsical” between the training points as a result of its greater degree of freedom. The upshot is that unseen (test) data will be poorly fit to the function desired. In neural net jargon, this is called poor generalization. On the other hand, too few parameters relative to the training data distribution and even the training set won’t be adequately matched.

The initial conditions of the network also have a strong influence on the training. The steepest descent algorithm is relatively slow and has a tendency to get trapped in local minima of the error surface (where the gradient is zero) [35].

### 3.4 Support Vector Machines (SVM)

SVMs are a type of machine learning algorithm, which have been successfully applied to a wide range of pattern recognition problems. Compared to artificial neural networks (ANNs), they are faster, can be used with larger numbers of features, are easier to interpret, and are deterministic. They find an optimal separating hyperplane between data points of different classes in a (possibly) high dimensional space. The actual support vectors are the points that form the decision boundary between the classes. Here is a simple example in 2D where a user is trying to separate data of two classes of samples (red circles and green squares) as shown in Figure 3.6.

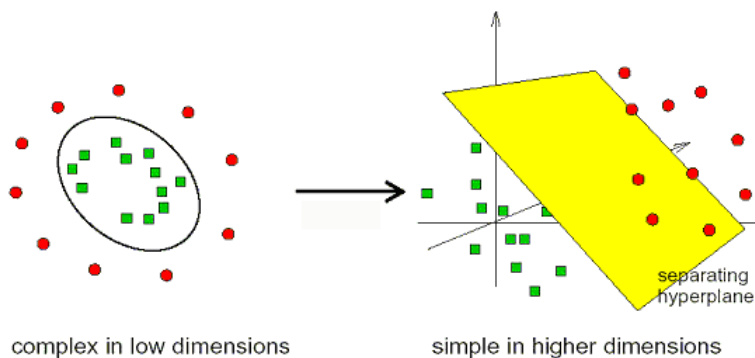


Figure 3.6: 2D example of separating data to two classes.  
Source: [28]

One of the main advantages of SVMs is that they are maximal margin classifiers. For example, in Figure 3.7 on the following page, D is a better separator than A, B or C since it will be more likely to classify new samples that are close to the current decision boundary.

A SVM is a supervised learning technique first discussed by Vapnik [11, 93]. An SVM is a maximum-margin hyperplane that lies in some space. Given training examples labeled either “yes” or “no”, a maximum-margin hyperplane splits the “yes” and “no” training examples, such that the distance from the closest examples (the margin) to the hyperplane is maximized. The use of the maximum-margin hyperplane is motivated by statistical learning theory, which provides a probabilistic test error bound which is minimized when the margin is maximized. If there exists no hyperplane that can split the “yes” and “no” examples, an SVM will choose a hyperplane that splits the examples as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples.

The original SVM was a linear classifier. However, Vapnik suggested using

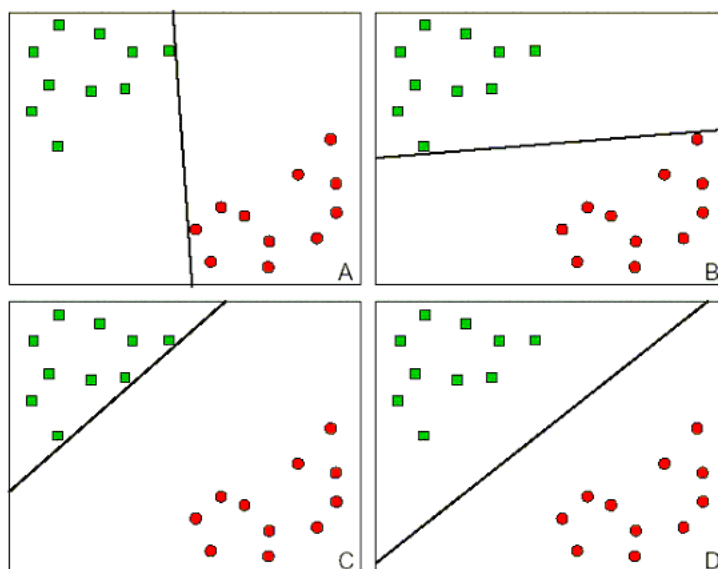


Figure 3.7: Maximal margin classifiers.

Source: [28]

the “kernel trick” [1]. In the kernel trick, every time a linear algorithm uses a dot product, it is replaced by a nonlinear kernel (if  $f$  is any function in any context, then the kernel of  $f$  is a certain equivalence relation on the domain of  $f$  which is defined in terms of  $f$ ) function. This causes a linear algorithm to operate in a different space. For SVMs, using the kernel trick makes the maximum margin hyperplane be fit in a feature space. The feature space is a non-linear map from the original input space, usually of much higher dimensionality than the original input space. In this way, non-linear SVMs can be created [24]. Support Vector Machines non-linearly map their  $n$ -dimensional input space into a high dimensional feature space. In this high dimensional feature space a linear classifier is constructed. Two results make this approach successful:



1. The generalization ability of this learning machine depends on the VC dimension (a measure of the capacity of a learning algorithm) of the set of functions that the machine implements rather than on the dimensionality of the space. A function that describes the data well and belongs to a set with low VC dimension will generalize well regardless of the dimensionality of the space.
2. Construction of the classifier only needs to evaluate an inner product between two vectors of the training data. An explicit mapping into the high dimensional feature space is not necessary.

## CHAPTER 4 METHODS

The general hypothesis that inspires this work is that **computer-aided diagnosis (CAD) can improve the accuracy of expert and non-expert readers for classifying suspicious lesions on BMRI.**

We propose these four specific aims:

1. Develop a method to segment BMRI datasets into regions of interest (lesions and ducts) and background.
2. Select and compute the shape, texture, and contrast enhancement pattern of suspicious lesions in BMRI.
3. Devise classifiers to distinguish benign from malignant lesions, using the features identified in the previous aim.
4. Assess the performance of the classifiers for expert and non-expert readers by using a set of pathologically-proven lesions from BMRI data.

This chapter describes the computer interface that we have developed to load and display BMRI datasets, the datasets we used, and then discusses each of the four specific aims in detail.

### 4.1 Image Viewer

Figure 4.1 on page 46 shows the BMRI image viewer and loader. The image viewer loads a BMRI image sequence and displays orthogonal, 2D cross-sections

through the BMRI data and provides window/level control. Multiple imaging sequences can be loaded simultaneously (eg, T1 and T2 weighted images, pre- and post-contrast images) so that direct comparisons can be made. Standard region of interest tools are available in the image viewer to mark suspicious masses or lesions. After lesions of interest have been marked, they can be saved to a file for subsequent review and analysis. Several analysis tools are also available in the image viewer, e.g., for image subtraction. Radiologists commonly review difference images (i.e., post-contrast minus pre-contrast) to increase the conspicuity of enhancing lesions, which includes most breast cancers. This is especially helpful when the lesion or lesion background exhibits high signal prior to contrast administration. It is also useful for the radiologist to assess the enhancement kinetics of a suspicious mass on BMRI, often called the “time-intensity” curve. Such curves are generated from a series of post-contrast images captured at different times. The viewer also includes tools for image segmentation, such as adaptive thresholding, fixed thresholding, and grayscale morphology, which can be used interactively to identify lesions of interest on BMRI. Although a fully automated segmentation system would be highly desirable, it remains a difficult problem. Allowing some user interaction increases robustness and reduces algorithmic complexity, but still offers real benefits such as increased diagnostic confidence and accuracy. A final layer of tools in the image viewer includes advanced functions for analyzing pre-selected lesions. These functions can extract features (i.e., shape, size, texture) from the selected lesions and classify lesions of interest as benign or malignant. The segmentation and classification processing are

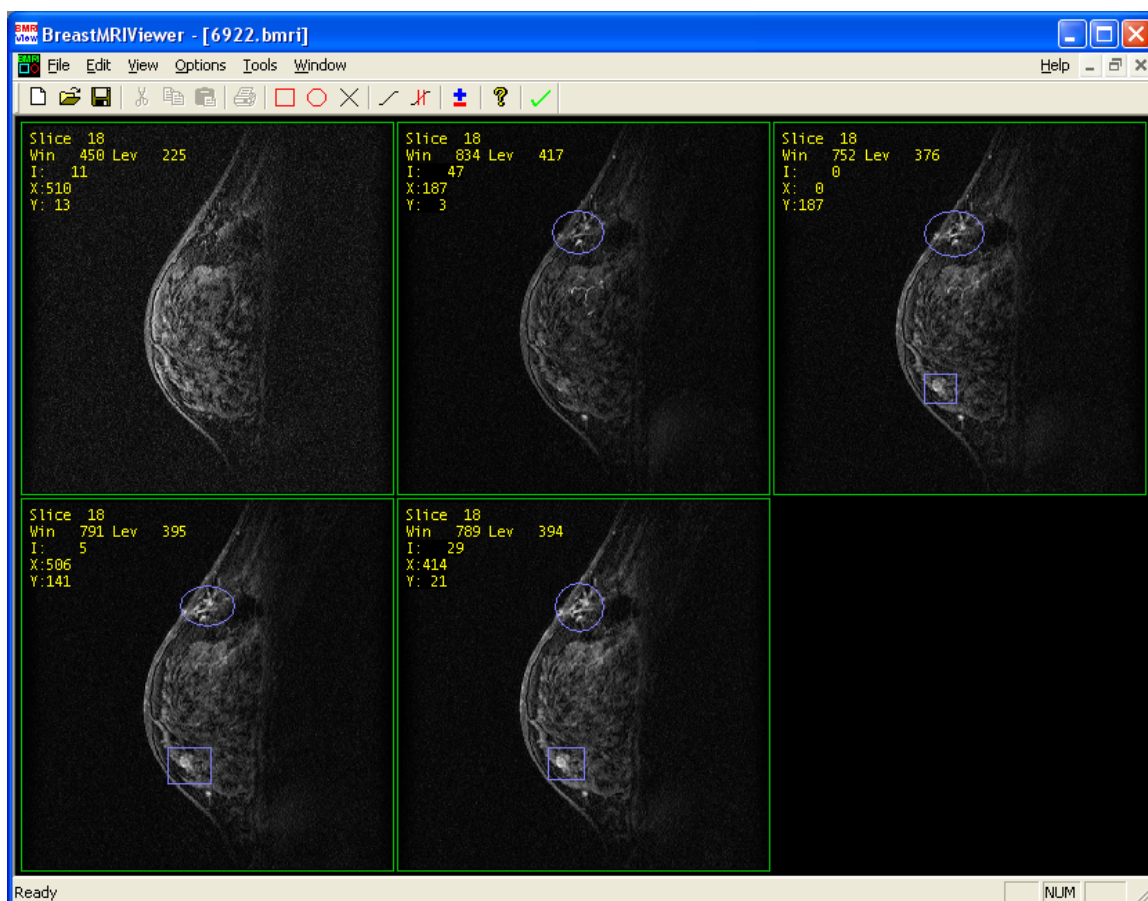


Figure 4.1: The image view showing the pre-contrast and four post-contrast images. Two enhanced regions: The first region is surrounded by a circle, and the second is surrounded by a box (Case: 6922, slice: 18).

described next.

## 4.2 Datasets

Archived, 80 pathologically-proven lesions imaged by BMRI were used to train and test the CAD system collected from the University of Pennsylvania (UPENN). 43 of the lesions were malignant (two ductal carcinoma in situ, DCIS, ten invasive ductal carcinoma, IDC, four invasive lobular carcinoma, ILC, three rim enhancement,

and most malignant lesions has more than 4 lymph nodes involved). 37 were benign lesions (seven fibrocystic, and five fibroadenoma).

For each patient, four different image sequences are acquired: One pre-contrast, and three post-contrast. For each image in the datasets, the in-plane resolution is 0.35mm, and the through-plane resolution is three mm. These studies were preprocessed to remove all patient information. Data are tracked by a unique serial number that is assigned to each dataset. Each set of BMRI data is accompanied by a definitive biopsy result that characterizes the lesion as benign or malignant.

### 4.3 Specific Aim 1

#### Develop a Method to Segment Breast MRI Datasets

Segmentation of suspicious regions is the important first step for a variety of image analysis tasks. There is a wide range of image segmentation techniques in the literature, some considered general purpose and some designed for specific classes of images [42].

For the breast MR images, skin and chest wall segmentation is the first task, then the enhanced regions are extracted.

The segmentation processing is used to identify suspicious lesions and mark them for subsequent classification. The first step is to find the chest wall boundary, since all regions inside of the chest are considered extra-mammary. After chest wall segmentation, the subsequent lesion detection takes advantage of the intensity enhancement in breast lesions when the post-contrast image is compared to the pre-contrast image [65]. It is important to mention that the regions that enhance in

BMRI are breast lesions, ducts, lymph nodes, and blood vessels. However, normal breast parenchyma does enhance progressively, but weakly. Only the first two region enhancements are significant, as it is perfectly normal for blood vessels to enhance.

#### 4.3.1 Skin and Chest Wall Segmentation

Two methods were examined for finding the chest wall. The first method used the Hough transform. The Hough transform approach was eventually abandoned in favor of a method using dynamic programming.

The Hough transform is used to detect lines in image processing applications [10, 27]. The line is a parametric form with two parameters, the angle of the line  $\theta$ , and intercept of the line ( $y = a + bx$ , or alternatively, a point from the line  $p$  is  $p = x \cos(\theta) + y \sin(\theta)$ ). Each pixel in image space contributes to a discrete number of possible lines that could pass through that point. Because of constructive interference, a real line in the image is mapped to a local maxima in Hough parameter space. Figure 4.2 on the following page shows the following four steps for the Hough transform method:

1. The two intersection points between the air and the breast skin (at the top and bottom of the image) are marked.
2. A virtual vertical line is drawn connecting the two points.
3. The Hough transform is applied to the image on the right side of the virtual line.
4. The strongest line is found (on the right side of the virtual line), and chosen as

the chest wall.

The Hough transform method was abandoned because in most datasets the chest wall is not a straight line.

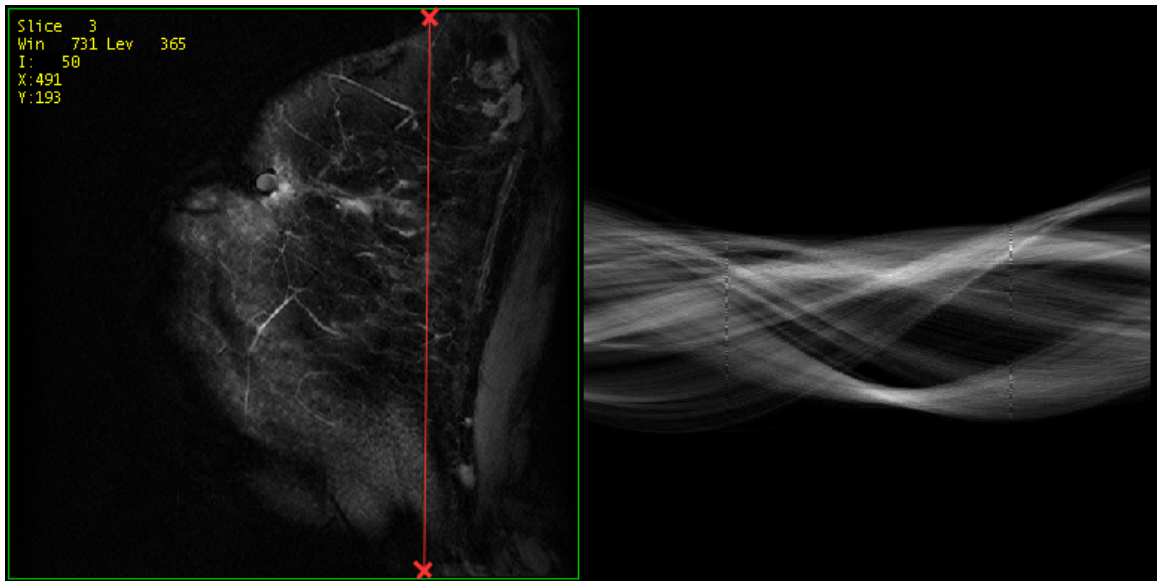


Figure 4.2: The Hough transform approach starts with marking the two intersection points between the air and the breast skin (the red x marks in the left image), then drawing the line that connects these two points (the red line in the left image), finally applying Hough transform on the right side of the line. (Case: 1606, slice: 3).

Dynamic programming (DP) can be used as a path finding algorithm that is capable of finding the optimum path in a tree. [16, 88] Every pixel in an image can be thought of as a node, then the goal is to find the path associated with the lowest/highest cost from the top to the bottom of the image.

For the DP method, we start by combining the pre and post-contrast sequences to decrease the noise and increase the contrast between the chest wall and the sur-

rounding tissue. Then, the images were blurred using 2D Gaussian filter of size  $k \times k$ , where  $k = 2\lceil 2\sigma \rceil + 1 = 9$ , and for this application  $\sigma = 2$ .  $\lceil \cdot \rceil$  represents the “ceiling” operator (i.e., round up to closest integer). This gives an odd sized filter, with two standard deviations in each direction. If a different level of blurring is requested, then  $k$  is automatically recalculated (see Figure 4.4 (b)). After blurring the images, a linear ramp filter ( $5 \times 7$ ), as shown in Figure 4.3 was applied to highlight the skin-air boundary. It was also determined that this was successful in highlighting the chest wall (See Figure 4.4 (c)).

$$Filter = \begin{bmatrix} -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \end{bmatrix}$$

Figure 4.3: The linear ramp filter that used to highlight the skin air boundary.

Then, the skin-air boundary was found using dynamic programming on each slice. Since the edge image (Figure 4.4 (c)) was used as the cost function, the brightest edge was almost always the skin-air boundary. Because the boundary can follow an almost horizontal path, the dynamic programming step searched for the best path within two steps of the current one  $(x+i, y-1)$ ,  $i = -2, -1, 0, 1, 2$ , where  $x, y$  are the location of the current pixel (see Figure 4.4 (d)). Now that the skin has been found, it is possible to mask out the breast tissue and narrow the search space for the chest wall. This is necessary because, in general, the breast tissue can be heavily textured,



and this would appear as bright edges to follow in dynamic programming.

The area to be masked off was computed as follows:

1. The average  $x$  location for the first 10 pixels along the skin boundary was calculated ( $\bar{x}_{top}$ )
2. The same was found for the bottom 10 pixels ( $\bar{x}_{bottom}$ )
3. A fixed offset was applied to each so that more of the breast tissue would be included. (The boundary is only the edge of the breast, not the entire breast tissue).

$$\bar{x}'_{top} = \bar{x}_{top} + o,$$

$$\bar{x}'_{bottom} = \bar{x}_{bottom} + o.$$

For our purposes,  $o = 35$ .

4. A virtual line was drawn between the two points, and all points to the left of that line were considered masked off.

The chest wall was then found using a slightly modified version of dynamic programming. Because the chest wall is a nearly vertical line in the image, the cost search was penalized for moving diagonally. Specifically the cost was computed as

$$C' = C - |i|p.$$

Where  $C$  was the cumulative cost for a pixel,  $i$  is the horizontal distance that was

moved to get to that location (for the chest wall we constrained the search to  $x \pm 1$ ), and  $p$  was a fixed penalty for a diagonal move (we chose  $p = 500$ ).

After segmentation of the skin and chest wall, a new file was created which labeled all of the pixels into one of seven regions:

1. Air,
2. Skin,
3. Breast Tissue,
4. Breast line,
5. Between breast line and chest wall (probably breast tissue),
6. Chest wall,
7. Inside chest.

See Figure 4.5 on page 54 for an example showing the results after region labeling.

#### 4.3.2 Segmentation of Enhanced Regions

First, a difference image is computed by subtracting the pre-contrast and post-contrast image. Two levels of thresholding are then applied. These thresholds are used to identify suspicious lesions and separate them from the background. The first level of thresholding is a more restrictive threshold. It has greater discrimination between ordinary regions and enhancing regions. However, because of this increased discrimination, it often leaves out parts of the enhancing regions that we would like

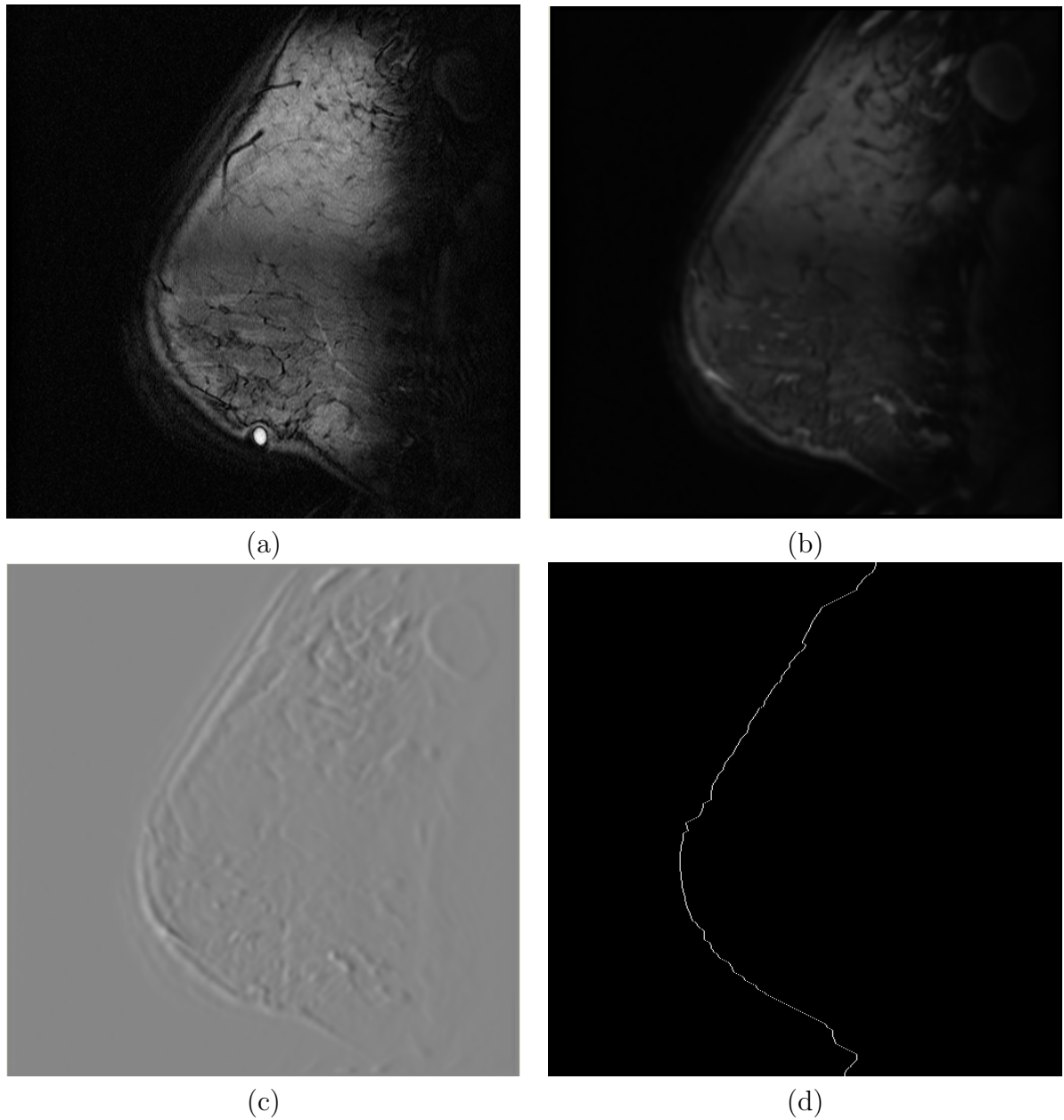


Figure 4.4: (Case: 1806, slice: 15) a) The original image. b) A blurred image using 2D Gaussian filter. c) The cost function computed from the edge image. d) The skin path found by DP.

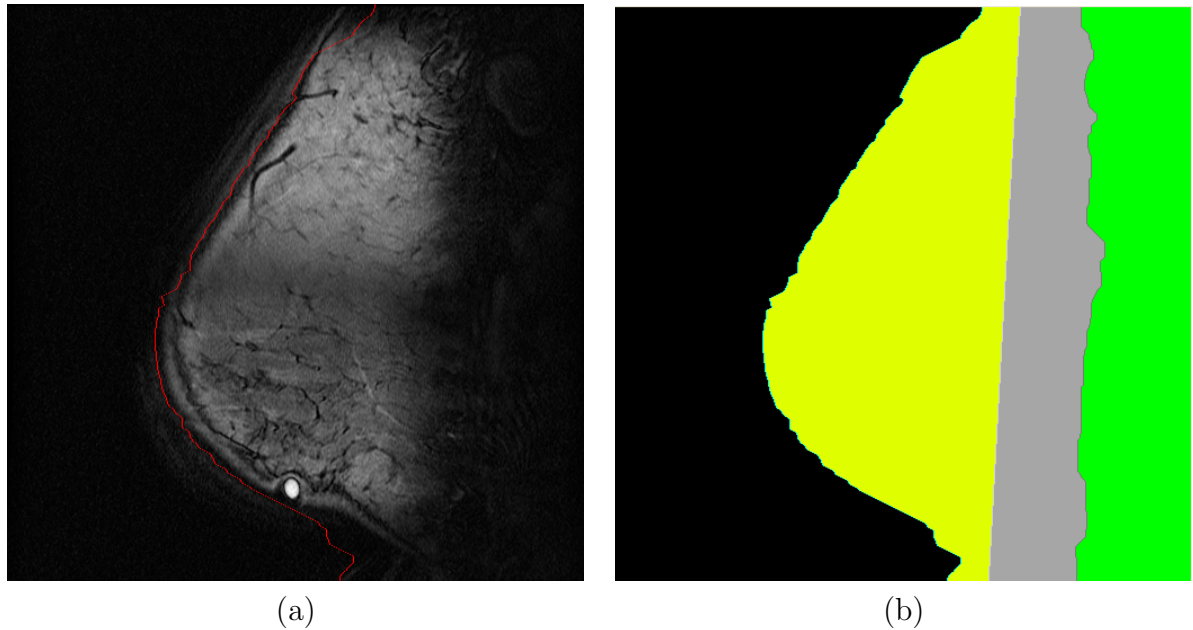


Figure 4.5: This figure shows Case: 1806, slice: 15. *a)* The original image with the automatic skin path overlaid on top in the red line. *b)* The seven regions mask with the breast tissues in yellow and inside the chest in green.

to keep. The second threshold is less restrictive, but it is constrained such that only regions that were present in the first segmentation are kept in the second.

Four methods were examined for automatic selection of the threshold levels. First, a constant threshold was applied to all the datasets. This method might be acceptable for CT and mammography, but not for BMRI, because there are no absolute values for the BMRI signal intensities. As a result, there was no single threshold value that performed adequately on all datasets.

Second, the grayscale histogram was plotted in the hope that an obvious distinction could be made between enhanced regions and non-enhanced regions. However, the histogram was typically unimodal, so no obvious threshold value could be

selected.

Third, the thresholds were computed as some percentage of the maximum value in the data. This approach was ineffective because blood vessels enhance very brightly, and confound the maximum value.

Finally, an adaptive region growing algorithm was applied [78]. Region growing was used for this task, because it is simple, effective, and widely used method for image segmentation. The region growing algorithm starts with a threshold value specified by the user. This is used to find where the lesions should be. In general, this is a fairly high threshold, used to find a subset of the true lesion. Then, connected component labeling is used to find all the initial lesions. Some of these may end up merging together. After labeling the lesions, the algorithm continues to find the minimum and maximum values for each lesion. The minimum and maximum values are found for each lesion  $(V_{i,max}, V_{i,min})$ , and at the same time, a table of lesions is created. This is only for the lesions as they stand, no adaptation occurs at this stage. Then, for all lesions in the table we use adaptive region growing to expand the lesion boundaries [9]:

1. A stack-based 3D region growing algorithm is applied to the volume. The volume is rastered through, and each pixel that matches the current lesion label is pushed onto the stack. If this pixel has not been accessed before, it's neighbors are also pushed onto the stack to be checked.

2. A pixel is added if it falls within the range of  $[T_{min}, T_{max}]$ .

$$T_{min} = (1 - \tau) \frac{V_{i,min} + V_{i,max}}{2}, \quad (4.1)$$

$$T_{max} = (1 + \tau) \frac{V_{i,min} + V_{i,max}}{2}, \quad (4.2)$$

Where  $\tau$  is a tolerance parameter.  $\tau$  has a range of  $[0,1]$ . A value of 1 would allow all pixels below the average (of max and min, not a global average) to be included, and pixels less than twice the average  $T_{min} = 0, T_{max} = V_{i,min} + V_{i,max}$ . A value of 0 would force all pixels to be equal to the average value  $T_{min} = T_{max} = \frac{V_{i,min} + V_{i,max}}{2}$ . Neither situation is generally desirable. More common values are between  $[0.1, 0.5]$ . (Currently, we ignore the  $T_{max}$  threshold, as it tends to remove the center of bright regions.)

3. As pixels are added to the region,  $V_{i,min}$  and  $V_{i,max}$  along with  $T_{min}$  and  $T_{max}$  are updated. So if you started with a bright region with a large amount of moderate valued pixels next to it, the threshold range should decrease to accommodate the change.
4. Also, as regions grow, it is checked to see if they become adjacent with another region. If they do, then the region stops growing, and the other region is considered equivalent to the current region.

Now that all regions have been grown, equivalences are sorted out, i.e., regions that grew together are merged into one region.

The next step used 3D connected components analysis is to label the enhanced

regions. Figure 4.6 on the next page shows the pre-contrast and the three post-contrast of slice 9 of a dataset, Figure 4.7 on page 59 shows the difference images between the pre-contrast and each of the post-contrast, and Figure 4.8 on page 59 shows the labeled regions in the thresholded images.

### 4.3.3 Segmentation Using Morphological Operations

A gray-level morphological algorithm was used to improve the enhanced region segmentation. It has been shown that a connectivity algorithm combined with gray-level morphological filter improved the segmentation of tortuous coronary arteries from 3D MRI [23].

The morphological filter starts by computing the gradient of the image:

$$G(x, y, z) = \sqrt{dx^2 + dy^2 + dz^2}, \quad (4.3)$$

where  $I(x, y, z)$  is the image voxel intensity at location  $(x, y, z)$ , and

$$dx = I(x + 1, y, z) - I(x - 1, y, z), \quad (4.4)$$

$$dy = I(x, y + 1, z) - I(x, y - 1, z), \quad (4.5)$$

$$dz = \frac{S_{x,y}}{S_z} (I(x, y, z + 1) - I(x, y, z - 1)). \quad (4.6)$$

$S_{x,y}$  and  $S_z$  are the spacing in the  $x, y$  and  $z$  dimensions, respectively. The normalization for  $dz$  is necessary because in our datasets the in-plane pixel spacing is isotropic, but there may be an order of magnitude difference between the in-plane and longi-

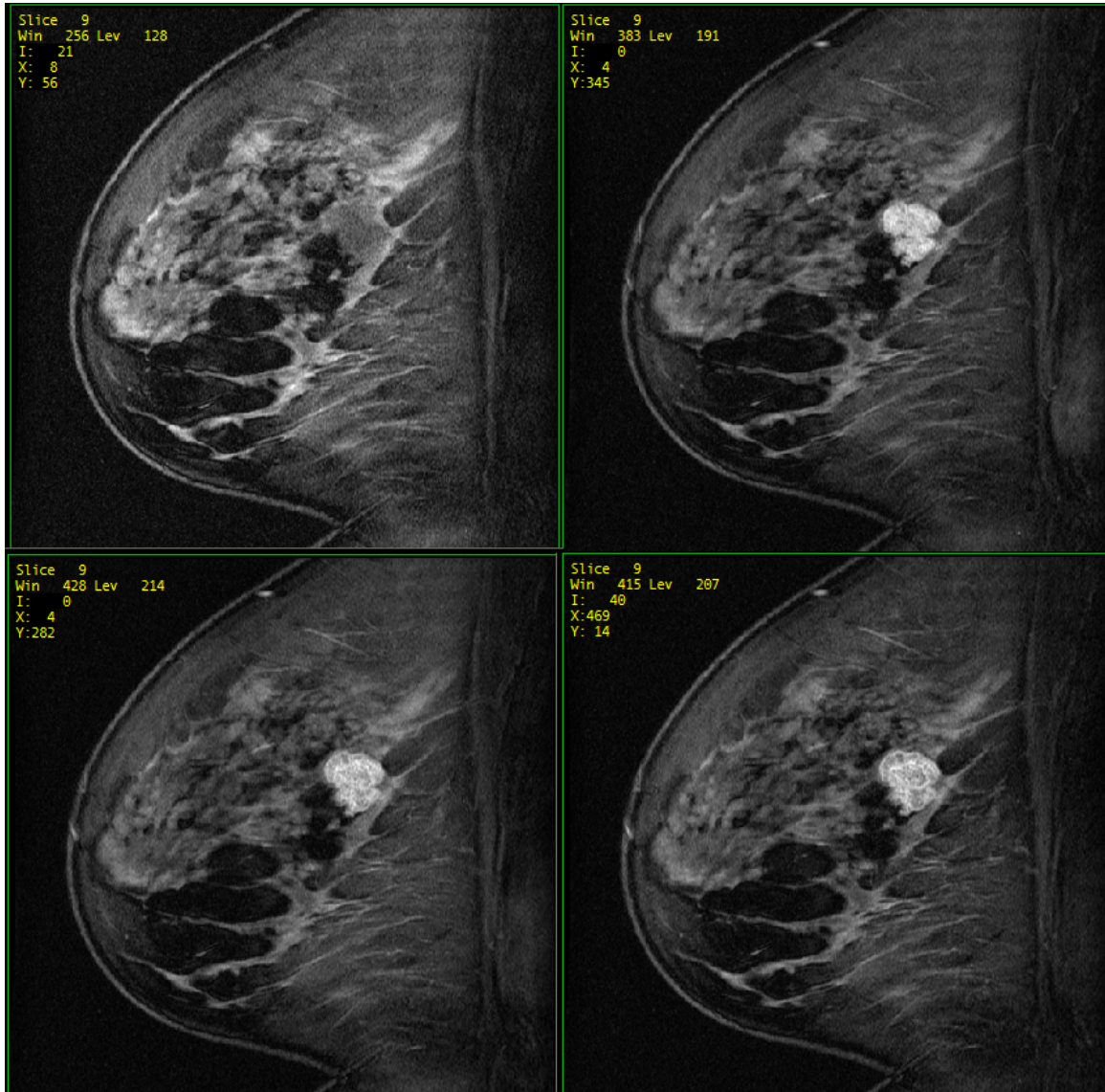


Figure 4.6: The pre-contrast and three post-contrast images. The lesion is obvious in the post-contrast images, but is barely visible in the pre-contrast images. ( Case: 5489, slice: 9).





Figure 4.7: The difference images between the pre-contrast and each of the three post-contrast images (Case: 5489, slice: 9)

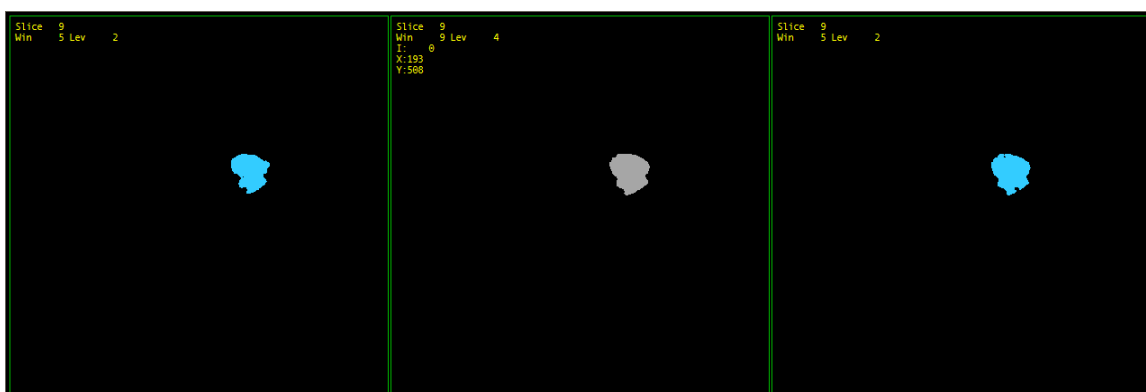


Figure 4.8: The labeled regions using 3D connected component algorithm (Case: 5489, slice: 9)

tudinal spacings. This normalization effectively assumes a linear interpolation in the  $Z$  direction. Also, the gradient is assumed to be 0 across image boundaries (e.g.,  $I(-1, 0, 0) = I(0, 0, 0)$ ). Then, for  $n$  iterations, a graylevel morphological erosion is applied;  $I'(x, y, z) = I(x, y, z) - G(x, y, z)$ , if  $I'(x, y, z) \leq 0$ ,  $I'(x, y, z) = 0$ . Finally, for  $n$  iterations, a graylevel morphological dilation is applied; If  $I'(x, y, z) > 0$ , then  $I'(x, y, z) = I(x, y, z) + G(x, y, z)$ . This method does not effect regions of homogeneity, so large, approximately constant lesions are not changed. Only regions with a large gradient are affected. Thus, objects like blood vessels, which are very narrow, are diminished.

In Figure 4.9 on the following page, The gradient was able to highlight the blood vessels. In Figure 4.10 on page 62, in the morphology images tubal objects, like blood vessels, are diminished.

#### 4.4 Specific Aim 2

**Select and compute the shape, texture, and contrast enhancement pattern of suspicious lesions in BMRI.**

Region shape, texture, and intensity kinetics features were used to characterize the enhanced lesions [4, 45, 73]. The features computed from the BMRI lesions follow three types: shape-based features (volume, area, radial length, spiculation, and perimeter length), texture-based features (compactness), and the special BMRI features (time/intensity curve, and relative intensity). The first two types were proven to be useful in mammographic masses classification [4, 45], whereas the third type was added to use the extra information existed in BMRI as a result of using a contrast-

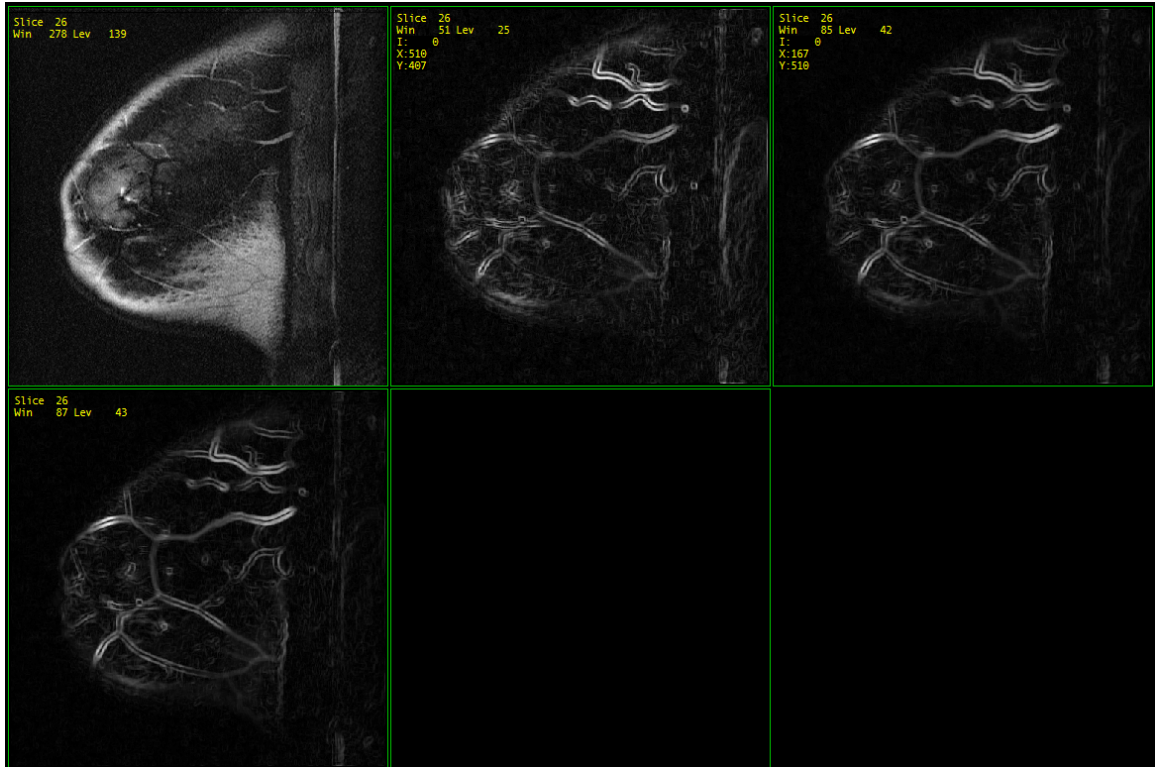


Figure 4.9: In the top left corner: the original image, followed by the gradient of the three post-contrast sequences computed using Equation 4.3 on page 57. The gradient was able to highlight the blood vessels. (Case: 5489, slice: 26)

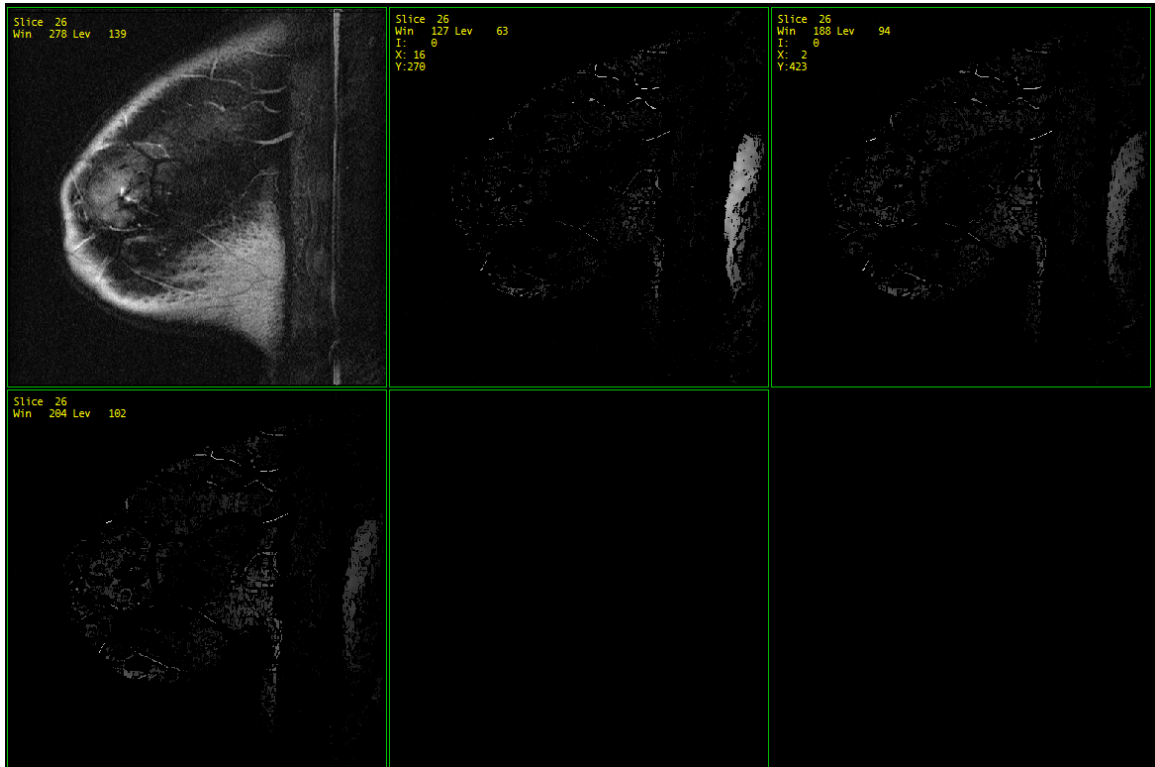


Figure 4.10: In the top left corner: the original image, followed by the three post-contrast sequences after applying the gray-level morphological algorithm. Blood vessels are diminished. (Case: 5489, slice: 26)

enhanced agent [73,90].

The following features were computed from each lesion. In all these equations,  $N$  is the number of voxels in the lesion, and  $n_k$  is the number of voxels in the  $k$ th slice of the lesion. (Some lesions may only be present on one slice.)

Because of the non-isotropic voxels, many features must be computed on a slice by slice basis, and then the value for the lesion was computed based on the minimum, maximum, summation, average, and root mean square of the values for each slice.

The center of mass of the lesion ( $\vec{C}$ ) is computed from

$$\vec{C} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix}, \quad (4.7)$$

where

$$\begin{aligned} \bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i, \\ \bar{y} &= \frac{1}{N} \sum_{i=1}^N y_i, \\ \bar{z} &= \frac{1}{N} \sum_{i=1}^N z_i, \end{aligned} \quad (4.8)$$

where  $(x_i, y_i, z_i)$  is the location of the  $i$ th voxel in the lesion.

For each slice  $k$ , the center of mass of the lesion ( $C_k$ ) is computed using

$$\vec{C}_k = \begin{bmatrix} \bar{x}_k \\ \bar{y}_k \\ z_k \end{bmatrix}, \quad (4.9)$$

where

$$\begin{aligned} \bar{x}_k &= \frac{1}{n_k} \sum_{i=1}^{n_k} x_i, \\ \bar{y}_k &= \frac{1}{n_k} \sum_{i=1}^{n_k} y_i. \end{aligned} \quad (4.10)$$

#### 4.4.1 Shape-based Features

The volume of the lesion is computed from:

$$V = N \, dx \, dy \, dz, \quad (4.11)$$

where  $dx$ ,  $dy$ ,  $dz$  is the spacing in the  $x$ ,  $y$ ,  $z$  directions, respectively, and  $N$  is the number of voxels in the lesion. For most of the sequences  $dx = dy = 0.35\text{mm}$ , and  $dz = 3\text{mm}$ .

For each slice  $k$ , the area is computed using

$$A_k = n_k \, dx \, dy. \quad (4.12)$$

The radial length for each boundary point is computed using

$$R_{j,k} = \sqrt{(x_{j,k} - \bar{x}_k)^2 + (y_{j,k} - \bar{y}_k)^2}, \quad (4.13)$$

where  $(x_{j,k}, y_{j,k})$  is the location of the  $j$ th boundary voxel on slice  $k$ .

Spiculation is computed as the standard deviation of the radial length as shown in Figure 4.11 on the following page

$$\text{Spiculation}_k = \sqrt{\frac{\sum_{j=1}^{m_k} (R_{j,k} - \bar{R}_k)^2}{m_k - 1}}, \quad (4.14)$$

where  $m_k$  is the number of boundary voxels on slice  $k$

The perimeter length for a given slice is the number of boundary pixels  $l_{\text{perimeter}} = m_k$  on that slice.

#### 4.4.2 Texture-based Features

The compactness of a lesion is computed using

$$\text{Compactness}_k = \frac{m_k^2}{4\pi n_k}. \quad (4.15)$$

All of the previous features have been proven to be very good at characterizing breast lesions in mammography [5, 45].

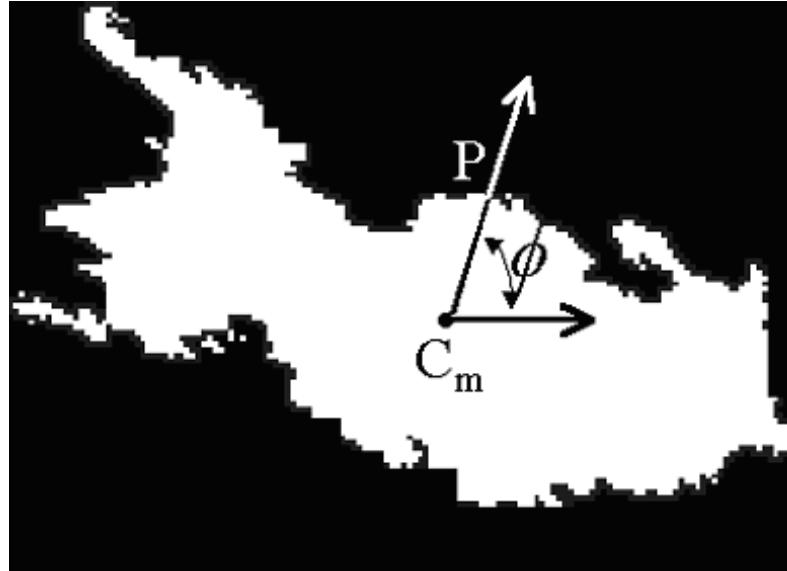


Figure 4.11: Spiculation Measurements:  $C_m$  is the center of mass,  $P$  is a boundary pixel, and  $\phi$  is the orientation of  $P$

#### 4.4.3 BMRI Features

In BMRI, there is a need to distinguish between blood vessels (non-suspicious enhancing lesions) and other enhancing lesions that should be further evaluated. To accomplish this, additional feature was computed.

For each lesion the time/intensity curve was calculated using:

$$\frac{\bar{I}_{\text{post},i} - \bar{I}_{\text{pre}}}{\bar{I}_{\text{pre}}} \times 100\%, \quad (4.16)$$

where  $\bar{I}_{\text{pre}}$  is the average value of the signal intensity in the pre-contrast image, and  $\bar{I}_{\text{post},i}$  is the average value of the signal intensity in the  $i$ th post-contrast image [90]. Previous studies have shown that blood vessels exhibit a second wash-in phase. Assuming BMRI sequences have sufficient temporal resolution to capture a second wash-



in phase, the time-intensity curve of a blood vessel might look like the curve shown in Figure 4.12 on the next page. [90]. Vascular enhancement kinetics were used to identify and reject blood vessels.

Since some of the datasets we have used do not have sufficient temporal resolution to capture a second wash-in phase, the time-intensity curve of a blood vessel looked like the curve shown in Figure 4.13 on page 69.

There are 6 features computed directly from the grayscale intensity of the voxels in the difference image.  $P_i$  is the pixel intensity of the  $i$ th voxel of the lesion. For all the following equations when not specified  $i = 1 \dots N$ .

1. Minimum Intensity

$$I_{min} = \min_i P_i, \quad (4.17)$$

2. Maximum Intensity

$$I_{max} = \max_i P_i, \quad (4.18)$$

3. Sum of Intensity

$$I_{sum} = \sum_{i=1}^N P_i, \quad (4.19)$$

4. Average Intensity

$$I_{avg} = \frac{1}{N} \sum_{i=1}^N P_i, \quad (4.20)$$

## 5. Root Mean Square Intensity

$$I_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N P_i^2}, \quad (4.21)$$

## 6. Standard Deviation of Intensity

$$I_{stddev} = \sqrt{\frac{\sum_{i=1}^N (P_i - I_{avg})^2}{N - 1}}. \quad (4.22)$$

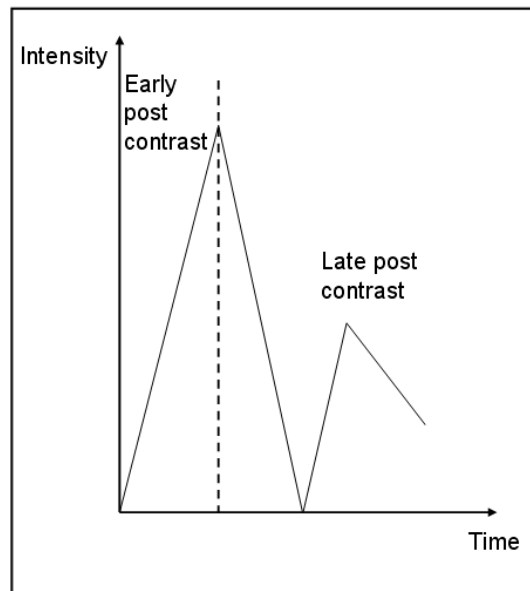


Figure 4.12: Time/intensity curve of a blood vessel.

Source: [90]

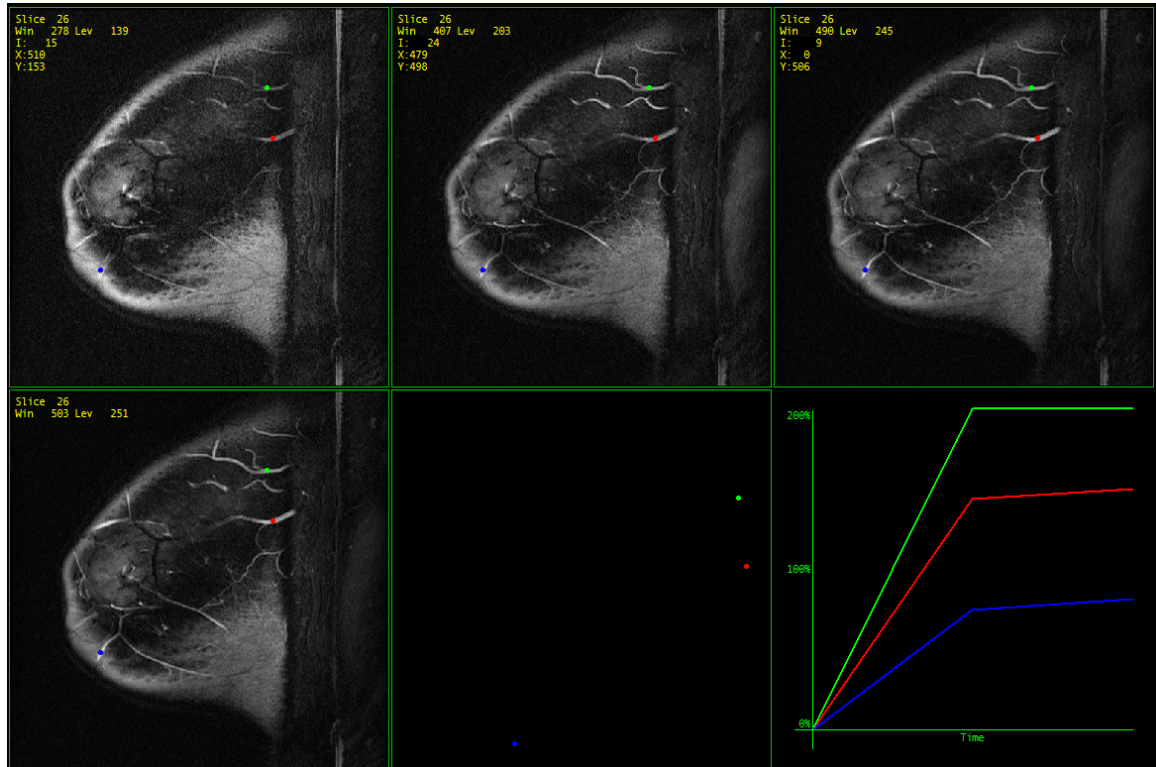


Figure 4.13: This figure illustrates that some of the datasets we have used do not have sufficient temporal resolution to capture a second wash-in phase. (Case: 5489, slice: 26)

#### 4.4.4 Features Selection

The TOOLDIAG package [84] was used for selecting the top best features to be used as inputs to different classifiers. A subset of 19 datasets was used to compute all the shape, texture and BMRI features. These features were computed as minimum, maximum, summation, average, standard deviation, and RMS, which yielded to 42 features used as an input to the TOOLDIAG package. The best features were found using the estimated minimal error probability, 1 nearest neighbor, and leave-one out methods.

The best features were: average relative intensity, RMS relative intensity, average spiculation, RMS spiculation, standard deviation of spiculation, average radius, RMS radius, average perimeter length, RMS perimeter length, average compactness, standard deviation of compactness, average area, and RMS area. **The above best 13 features were used as inputs of the classifiers.**

### 4.5 Specific Aim 3

#### Devise Classifiers to Distinguish Benign from Malignant Lesions

Breast lesions in MRI were classified as benign or malignant using three classifiers: the backpropagation neural network (BNN), the support vector machines (SVM), and the Bayesian classifier (BC).

##### 4.5.1 Classification of Breast Masses from Mammograms

In this section, we briefly describe related previous work comparing the performance of expert radiologists, radiology residents, and a backpropagation neural

network (BNN) for classifying breast masses on mammography. Our goal was to reduce the false negative rate during routine reading of mammograms. 160 mammograms from 3 different institutions were used. There was biopsy proof for at least one mass in each mammogram. The masses were identified by an expert radiologist and extracted using region growing with a seed. Ten texture and shape based features (area, perimeter, compactness, radial length, spiculation, mean and standard deviation of radial length, minimum/maximum axis, and boundary roughness) were used as inputs to a three-layer BNN. Shape features were computed based on the boundary of the mass; texture features were computed from the pixel values inside the mass. 140 cases were used for training the BNN and the remaining 20 cases were used for testing. The testing set was read by three expert radiologists, three radiology residents, and the BNN. The performance of the human readers and the BNN was evaluated by computing the area under the ROC curve ( $A_z$ ) [6,8].

#### 4.5.2 BNN Design

Artificial neural network design includes decision on the number of the network layers, the number of the hidden layer neurons, the neurons type of transfer functions, and the neural network learning method. The design is very dependent on the application.

We used a three layer neural network with three neurons in the hidden layer and one neuron in the output layer. The inputs to the network were the selected best shape, texture, and contrast enhancement features as described in Specific Aim 2 (section 4.4 on page 60). Thus the input space had 13 dimensions, and the input

layer consisted of 13 neurons. The backpropagation algorithm was used for training [28], and Levenberg-Marquardt (L-M) optimization method [51, 57, 89] was used for accelerating the network learning. During training, output from the backpropagation neural network (BNN) was compared with the biopsy results (the targets). The mean-square difference was used to estimate the error.

We used a linear transfer function for the input layer, and Hyperbolic Tangent Sigmoid transfer function for the hidden and output layers neurons as shown in Figure 4.14 on the following page. The BNN system output is a floating number in the range of  $[-1 +1]$ , where -1 indicates a classification decision of a benign lesion, and +1 indicates a classification decision of a malignant lesion.

#### 4.5.2.1 Number of Neurons

For each input of the BNN system  $v_1, v_2, \dots, v_{13}$ , a random weight  $w_1, w_2, \dots, w_{13}$  is assigned. The input of each neuron in the hidden layer would be:

$$x = \sum_{i=1}^{13} (v_i w_i). \quad (4.23)$$

It is important to use more than two neurons in the hidden layer, in order to allow non-linear decision boundary as shown in Figure 4.15 on page 74. The number of the input layer neurons is decided by the number of the features used as inputs, and the number of the output layer neurons is dependant on the number of classes used to classify the patterns. However, selecting the number of the hidden layer neurons ( $N_h$ ) is not as easy. This number controls the complexity of the decision boundary.

When there are linear patterns, one or two neurons are sufficient to do the task, but when there are non-linear patterns you must use more than two neurons in the hidden layer.  $N_h$  also determines the number of network weights, so it is important that the number of weights not exceed the number of the training samples. There is no formula that calculates the optimal  $N_h$ , because every application is a special case. We used trial and error method to choose  $N_h$ . Taking into account the previous studies in chapter two, we started with a large  $N_h$  and reduced  $N_h$  until a good result was reached to prevent unnecessary complexity as shown in Figure 5.13 on page 110 and Table 5.5 on page 109. We ended up with  $N_h = 3$ . Thus, the total number of weights is  $13 \text{ inputs} \times 3 \text{ neurons} = 39 \text{ weights}$ .

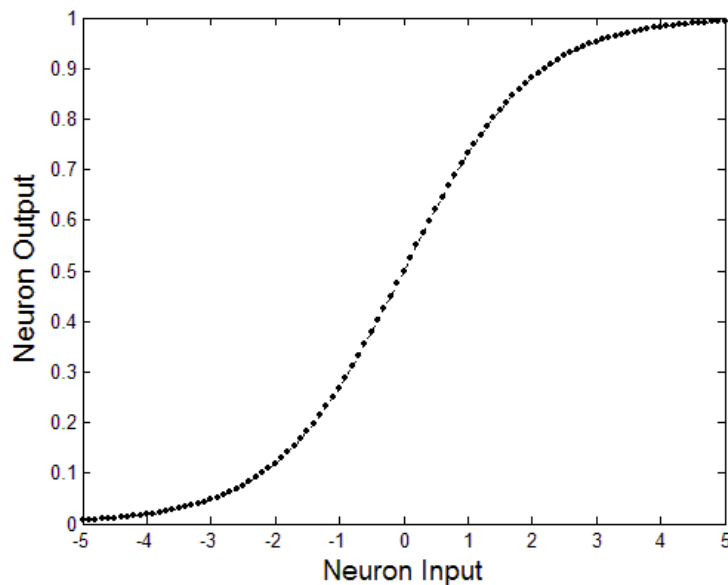


Figure 4.14: Non-linear transfer function: Hyperbolic Tangent Sigmoid used in the hidden and output layers neurons.

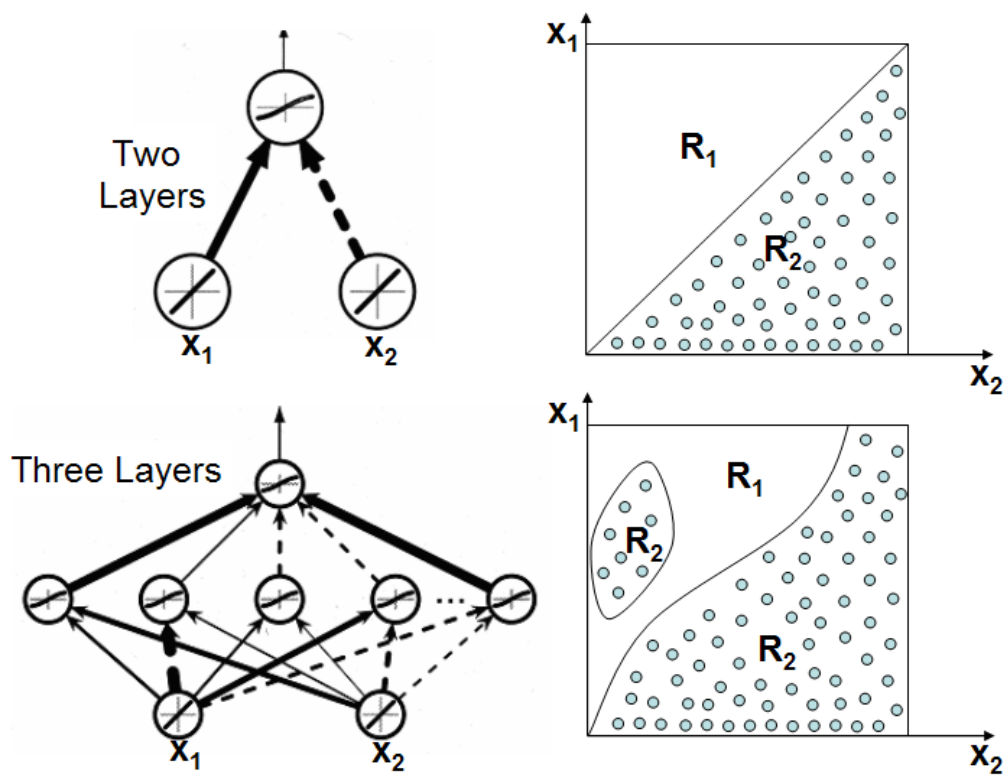


Figure 4.15: Using two neurons in the hidden layer allow a linear decision boundary as shown in the top, whereas using three neurons allows more complicated decision boundaries such as the bottom case.

Source: [28]



#### 4.5.2.2 Transfer Function

A transfer function is a function that maps the input of a neuron to its output.

A good transfer function should be:

- Continuous, and has a none-zero derivative over a wide range, this is important when using backpropagation algorithm for training.
- Asymptotic, otherwise, it is possible during training that one or more of the weights would be adjusted to minimize the network error, which might force the derivative to reach zero (saturation stage).
- Easy and fast to compute.

After studying the above requirement we chose the hyperbolic tangent sigmoid function (see Figure 4.14 on page 73) to be used in the hidden and output layers neurons, which calculates its output according to the following equation:

$$f(x) = 2/(1 + \exp^{-2x} - 1). \quad (4.24)$$

#### 4.5.2.3 Backpropagation Algorithm

The backpropagation learning algorithm is a supervised learning method that can be used with multilayer networks and nonlinear differentiable transfer functions.

Backpropagation is a gradient descent algorithm, in which the network weights are moved along the negative of the gradient of the performance function. The term

backpropagation refers to the manner in which the gradient is computed for nonlinear multilayer networks.

The power of backpropagation is that it allows us to calculate an effective error for each hidden unit, and thus adjust the hidden weights. As in all supervised learning methods, the backpropagation algorithm consists of presenting an input pattern and changing the network parameters to bring the actual outputs closer to the desired target values. These outputs are compared to the target values, any difference corresponds to an error. This error is some scalar function of the weights, thus the weights are adjusted to reduce the error. This error function is the sum of square differences of the outputs and targets. Let  $v^i$  be the training set elements,  $y^i$  are the outputs, and  $w^i$  are the desired output or targets, then the error function is:

$$E = \sum_i \sum_j (y_j^i - w_j^i)^2. \quad (4.25)$$

The backpropagation weights adjustment is:

$$w_{ij}(k+1) = w_{ij}(k) - \eta \frac{\partial E}{\partial w_{ij}}. \quad (4.26)$$

where,  $k$  is the iteration number, and  $\eta$  is the learning rate. The weights are adjusted continuously until a “good performance” is reached.

The steps of backpropagation algorithm are:

1. Initialize  $w_{ij}$  with random small numbers, and set  $k = 0$ .
2. Start with a randomly chosen pattern  $v$  from the training set, and evaluate  $y$ .

3. If  $y \neq w$ , adjust the weights:

$$w_{ij}(k+1) = w_{ij}(k) - \eta \delta_j z_i(k), \quad (4.27)$$

Where,  $z_i(k)$  is node  $i$  output, and  $\delta_j$  is the error of the  $j^{th}$  node.

$$\delta_j = y_j(1 - y_j)(w_j - y_j). \quad (4.28)$$

4. Repeat from the second step with another training pattern.

5.  $k = k + 1$ .

6. Repeat steps 2-4 until  $\|\Delta E(w) < \theta\|$  for all training patterns.

#### 4.5.2.4 Learning Acceleration

Levenberg-Marquardt (L-M) Optimization algorithm was used to accelerate the network learning [51, 57, 89]. This method was designed to reach a second order learning by estimating the Hessian matrix using the sum of outer products of the gradients. The reason of using the Hessian matrix is to include additional information about the local topography of the performance function of the network in order to improve, accelerate, and stabilize finding the minimum error.

For example, when descending the walls of a very steep local minimum bowl we must use a very small step size to avoid “rattling out” of the bowl [28]. On the other hand when we are moving along a gently sloping part of the error surface we

want to take large steps otherwise it will take forever to get anywhere. This problem is compounded by the implementation problem where in steep regions (where we have to be careful not to make our steps too large) we move quickly and in shallow regions (where we need to move in big steps) we move slowly. Another issue is that the curvature of the error surface may not be the same in all directions. For example if there is a long and narrow valley in the error surface the component of the gradient in the direction that points along base of the valley is very small while the component perpendicular to the valley walls is quite large even though we have to move a long distance along the base and a small distance perpendicular to the walls. So we would like to use slightly more sophisticated gradient descent algorithms than simple steepest descent.

Since that we used the sum of squares as a performance function, we were able to approximate computing the Hessian matrix by:

$$H = J^T J, \quad (4.29)$$

Where  $J$  is the Jacobean matrix, which contains the first order derivatives of the network error in terms of the weights and biases. Let  $N_i$  be the number of the network inputs,  $N_h, B_h$  be the number of the hidden layer neurons and biases, and  $N_o, B_o$  the number of the output layer neurons and biases, then the Hessian matrix size would be:

$$H_{size} = [((N_i + B_h)N_h) + ((N_h + B_o)N_o)]^2. \quad (4.30)$$

So, for  $N_h = 13$ :

$$H_{size} = [((13 + 1)3) + ((3 + 1)1)]^2 = (46)^2 = 2116. \quad (4.31)$$

So the Hessian matrix is the second derivative of the network output in terms of 46 variables and with a size of 2116, which is not easy to compute. This is why the Hessian matrix computation was approximated to equation 4.29.

The network error is calculated from the steepest gradient descent:

$$g = J^T e, \quad (4.32)$$

Where  $e$  is the network errors vector.

L-M algorithm uses this approximation when updating the weights:

$$w_{k+1} = w_k - (J^T J + \mu I)^{-1} J^T e, \quad (4.33)$$

Where  $I$  is the identity matrix,  $\mu$  is a constant. When  $\mu = 0$  the last equation becomes:

$$w_{k+1} = w_k - H^{-1} g. \quad (4.34)$$

And when  $\mu$  is big, we get the steepest gradient descent with a small step. With every iteration,  $\mu$  gets a smaller value as the network performance function is being minimized, so the network learns faster.

Figure 4.16 shows the 1-D case of using L-M algorithm to accelerate the network learning.

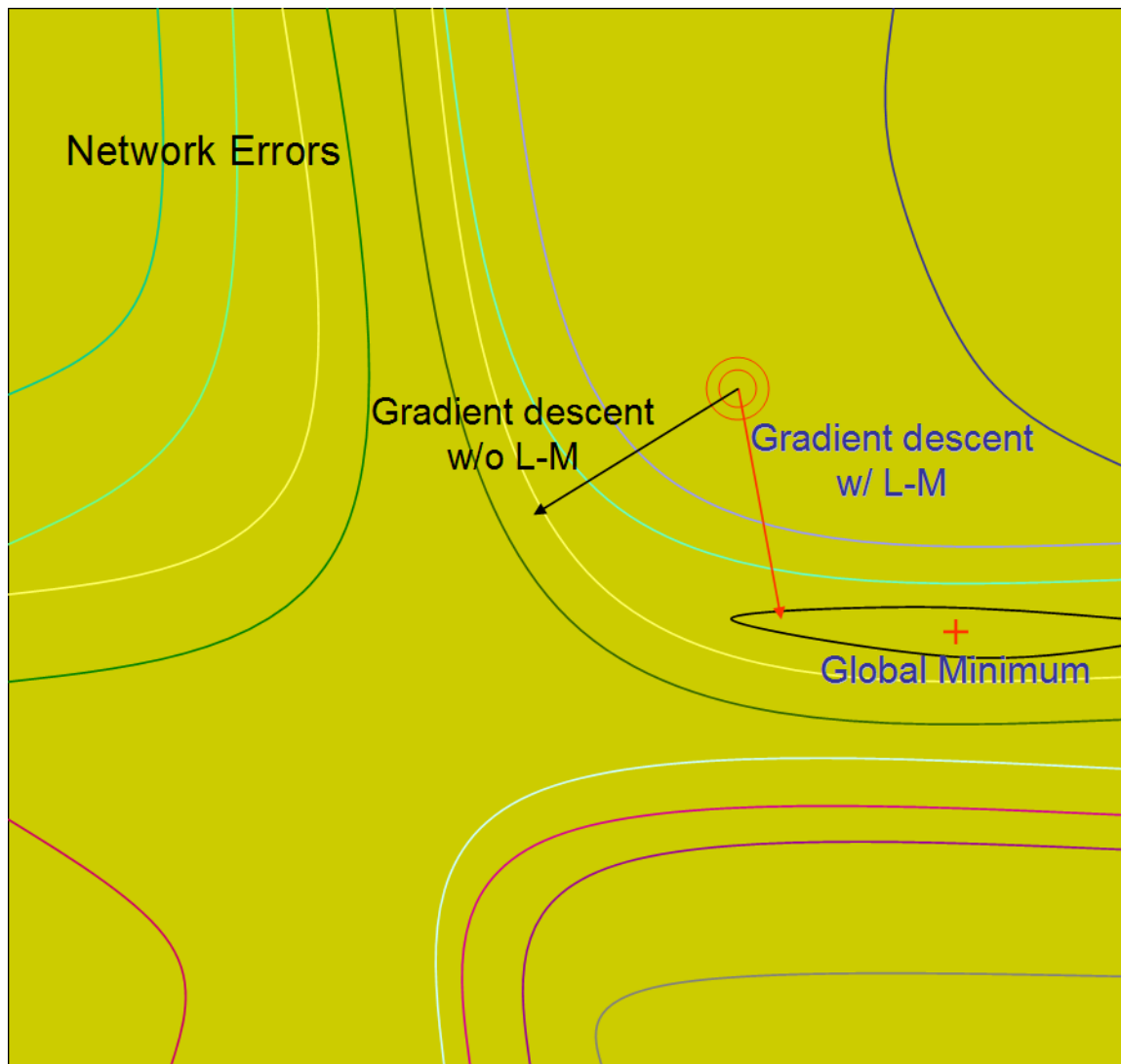


Figure 4.16: The effect of using L-M algorithm on accelerating network learning. Using L-M methods allows faster movement to the minimum error point.

#### 4.5.2.5 Accelerating the Convergence of the Backpropagation Algorithm

The curvatures of the error surface might often contain hills, where the slope of the error function is too small. If  $J(w)$  is the network error function, which is a function of both weights and biases, but for simplification purpose, we will assume that it is a function of merely the weights. Then the slope of the error function would be:

$$E_{slope} = dJ(w)/d(w) \quad (4.35)$$

The small slope areas are created when a large number of weights are exist, where the error function becomes dependent on one of these weights and independently from the others.

One way to accelerate the convergence of the backpropagation algorithm is momentum method, where the equations 4.25, and 4.26 on page 76 are updated to [28]:

$$\Delta w_{ij} = \eta \frac{\partial E}{\partial w_{ij}}, \quad (4.36)$$

$$\Delta w_{ij} = \eta \frac{\partial E}{\partial w_{ij}} + \eta \Delta w_{ij}. \quad (4.37)$$

This gives us:

$$w_{ij}(k+1) = w_{ij}(k) + \eta \delta_j z_i(k) + \alpha \Delta [w_{ij}(k) - w_{ij}(k-1)]. \quad (4.38)$$

$\alpha$  is the momentum constant, and it is in the range of  $[0 \ 1]$ . The role of the momentum constant is to distribute the update in the previous iteration to the current iteration, forcing the weight optimization to proceed through local minima.

#### 4.5.2.6 Learning Rate

The learning rate is the only factor that determines how fast the network will reach the global minimum error point, and not the final values of the weights. The learning rate significantly affects the quality of the net. If some of the weights converged faster than the rest, then the network performance might not be the same for the whole input range. Or even might not be the same for all the classes. Figure 4.17 on page 84 shows the effects of different learning rates on the convergence of a 1-D function.

The optimal learning rate is the one that leads to the global minimum in one step as shown in Figure 4.17, b. The second order error function (mean square root, MSR) is giving by:

$$\frac{\partial^2 E}{\partial w^2} \Delta w = \frac{\partial E}{\partial w}, \quad (4.39)$$

$$\eta_{optimal} = \left( \frac{\partial^2 E}{\partial w^2} \right)^{-1}. \quad (4.40)$$



As shown in Figure 4.18 on page 85, the maximum value of the learning rate that allows convergence is:  $\eta_{max} = 2\eta_{optimal}$ , as shown in Figure 4.17 (c). For the BNN system, we used  $\eta \approx 0.1$  with a learning time of 100 epochs.

### 4.5.3 SVM Design

We also compared the performance of BNN with that of another classification tool: the support vector machine (SVM). SVM learning has many real world applications including object recognition [80]. In SVM learning, instead of minimizing an object function based on the training samples (e.g., mean square error), the SVM attempts to minimize a bound of the generalization error, i.e., the error made when testing on data not seen during training. Consequently, SVM tends to perform better when applied to data outside the training set when compared to other types of classifiers. The SVM achieves good performance by focusing on the training examples that are most difficult to classify, the so-called support vectors [29].

In SVM the input is mapped by a nonlinear function  $\varphi(\cdot)$  to a high dimensional space, and the optimal hyperplane (for the two classes case) is the one that has the largest margin. The support vectors are those (transformed) patterns that determine the margin; they are the hardest patterns to classify, and the most informative ones for designing the classifier. Each input pattern  $v_i$ , ( $i = 1, 2, \dots, n$ ) has been transformed to  $y_i = \varphi(v_i)$ , we let the output  $z_i = \pm 1$ , according to whether pattern  $i$  is in class 1 or class 2. Then a linear discriminant in the augmented  $y$  space is:

$$g(y) = w^t y, \quad (4.41)$$

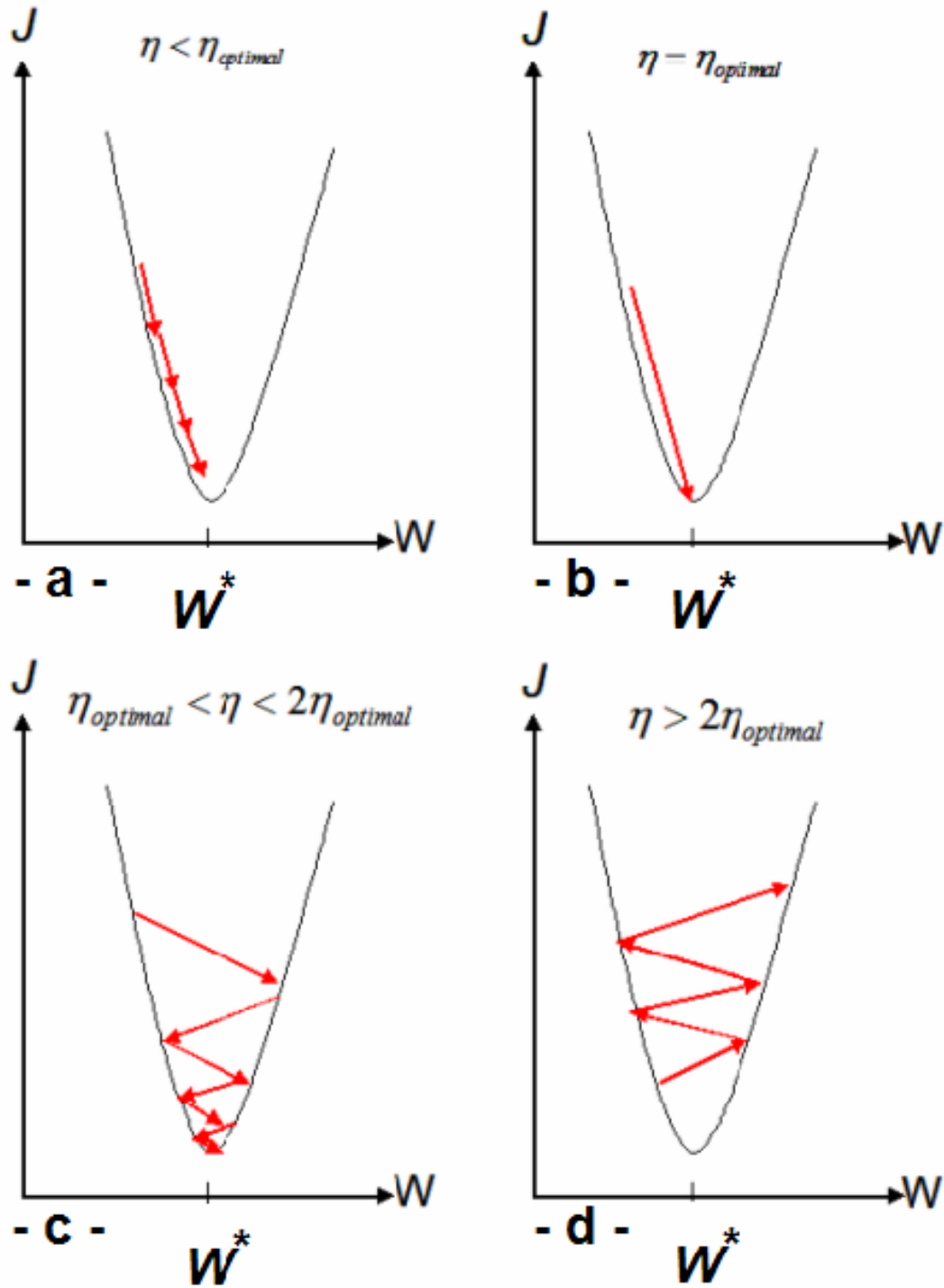


Figure 4.17: The effects of different learning rates  $\eta$  on the convergence of a 1-D function. a:  $\eta < \eta_{optimal}$  the convergence is guaranteed, but the learning is slow. b:  $\eta = \eta_{optimal}$  one learning step is sufficient to reach the global minimum. c:  $\eta_{optimal} < \eta < 2\eta_{optimal}$  oscillation, but converging slowly. d:  $\eta > 2\eta_{optimal}$  diverge. Adapted from [28].

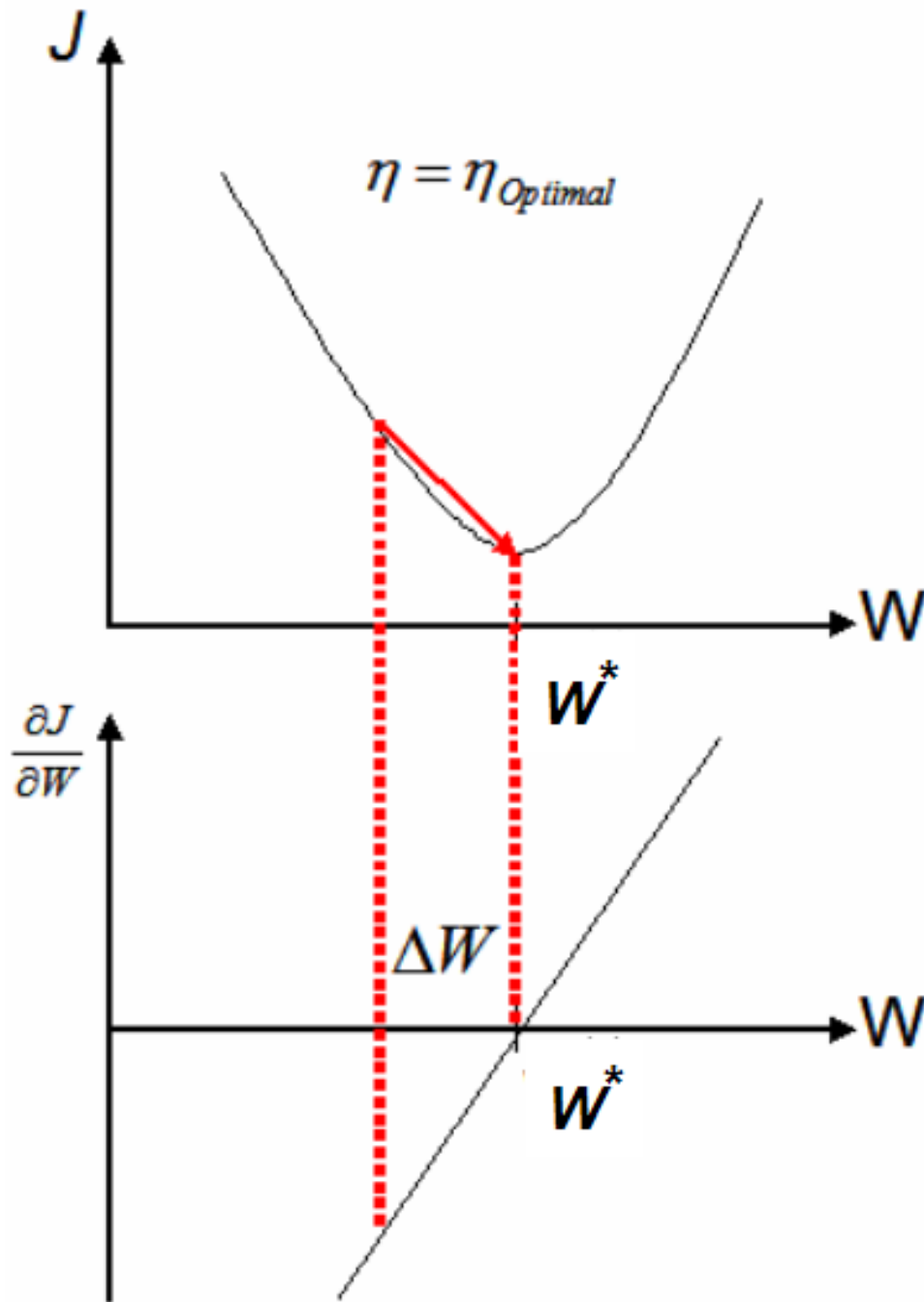


Figure 4.18: A second order error function, and its linear derivative. The optimal learning rate guarantees the arrival to the optimal weight value, which gives the minimum error in one step.

Source: [28]

where both the weight vector and the transformed pattern vector are augmented, thus a separating hyperplane ensures:

$$z_i g(y_i) \geq 1. \quad (4.42)$$

The goal in training a SVM is to find the separating hyperplane with the largest margin (the largest the margin, the better generalization of the classifier). The distance from any hyperplane to a pattern  $y$  is  $|g(y)|/\|w\|$ , and assuming that a positive margin  $b$  exists, the previous equation implies:

$$\frac{z_i g(y_i)}{\|w\|} \geq b. \quad (4.43)$$

The goal here is to find the weight  $w$  that maximizes the margin  $b$ .

#### 4.5.3.1 SVM Learning

Let  $y_1, y_2, \dots, y_n$  be the training  $n$  samples with some labeled as class 1, and some labeled as class 2. These samples were used to determine the weights  $w$  in a linear discriminant function as described in equation 4.41 on page 83. The assumption in this case would be that the samples are linearly separable. A sample  $y_k$  is classified correctly if  $w^t y_k > 0$  and  $y_k$  is labeled class 1, or if  $w^t y_k < 0$  and  $y_k$  is labeled class 2. This suggests a replacement of all samples labeled class 2 by their negatives to simplify the treatment of the two-category case. With this simplification we can forget the labels and look for a weight vector  $w$  such that  $w^t y_k > 0$  for all the samples.

Let  $y_k$  be the  $k$ th sample, and  $y^k$  is the  $k$ th misclassified sample, then the steps of the training algorithm [68] are as follows:

1. begin initialize  $w$  with random numbers,  $k = 0, j = 0$ .
2. do  $j = (j \bmod n) + 1$ ,
3. If  $w^t y_j < 0$  (misclassified) then
  - (a)  $k = k + 1$ ,
  - (b)  $y^k = y_j$ ,
  - (c)  $w = w + y^k$ .
4. until  $w^t y_j > 0$  for all samples  $y_1, y_2, \dots, y_n$ .

SVM learning was used as implemented by Mitchell *et. al.* [68]

#### 4.5.4 BC Design

The Bayesian classifier was trained and tested using the leave-one-out method. Its output is a floating point value for the classification, using the probability of class 1 for benign lesions versus the probability of class 2 for malignant lesions.

The Bayesian classifier assumed a 13-dimensional Gaussian distribution as the inputs. This Gaussian distribution was computed using two parameters: the mean vector and covariance matrix of the distribution.

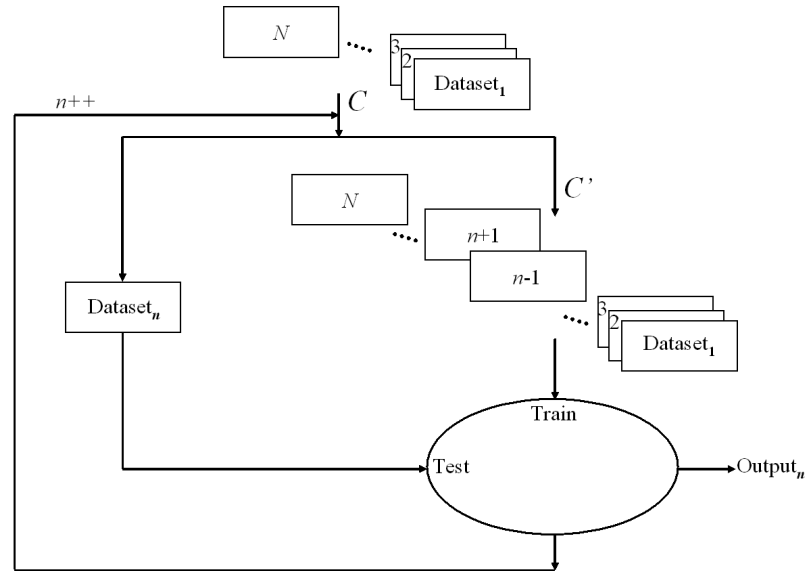


Figure 4.19: Classifying the data using the leaving-one-out method. For each dataset ( $n$ ) in a collection  $C$  of  $N$  datasets, a new collection  $C'$  (of size  $N - 1$ ) is created without the  $n^{\text{th}}$  dataset. A classifier is trained with  $C'$ , and  $n$  is evaluated.  $\text{Output}_n$  is then the classification of dataset  $n$ .

#### 4.5.5 Data Classification

The three classifiers were trained using the leave-one-out method. For each lesion, an experienced reader identified suspicious lesions and mark the lesion using the CAD system. For each of the marked lesions, we extracted the lesion features. The best 13 features, combined with the “ground truth” (pathology result), served as inputs during each of the classifiers training. Figure 4.19 shows how the datasets are classified using the leave-one-out method.

**Summary** Three artificial classifiers were used to distinguish the 80 BMRI lesions into benign and malignant. The backpropagation neural network (BNN), the support

vector machine (SVM), and the Bayesian classifier (BC). For each lesion, 42 shape, texture, and enhancement kinetics features were computed. The best 13 features were selected using the TOOLDIAG package, and used as inputs of the three classifiers. Only the BNN classifications were used in the human reader experiments.

#### 4.5.5.1 Pilot Study

Before the start of the human reader experiments, a pilot study was performed using two readers who were not involved in the final experiments. We used a subset of 37 lesions, and did not give any information about the system performance. We gave the system output as a percentage with 0% as benign, and 100% as malignant.

The BNN significantly outperformed both readers ( $p < 0.001$ ), but since they had not been informed as to its performance they did not trust the system and did not show significant improvement in their classification. Furthermore, the readers had some difficulty understanding the percentage rating. After seeing these results, we decided to modify the human reader experiments as follows:

1. Show the ROC curve of the BNN system to the readers to familiarizing them with the system performance;
2. Before presenting the results to the readers, we transformed the output of the BNN system from a percentage to a binary (benign/malignant) decision.

#### 4.5.6 Classification by Human Readers

Five readers (one breast MR expert, two mammographers, and two body imaging fellows) were used to evaluate system performance. All readers were familiar with

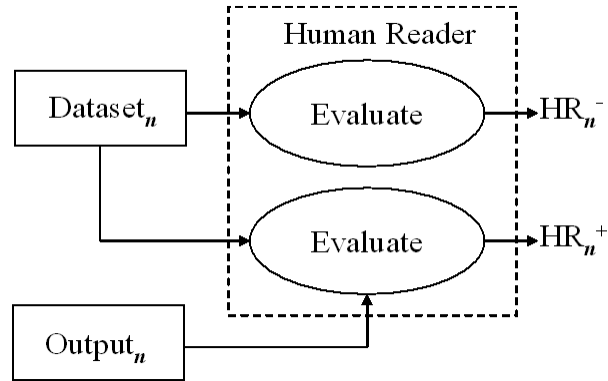


Figure 4.20: Every human reader classifies each dataset; for a dataset ( $n$ ), they first evaluate it without CAD assistance ( $HR_n^-$ ). Then they are shown the results of the CAD system ( $Output_n$ ), and allowed to re-evaluate their decision ( $HR_n^+$ ).

lesion descriptors including margin, shape, contrast enhancement pattern and kinetics, as previously described in [72]. Lesion classification criteria followed the model of Nunes et al. [75]. Figure 4.20 shows how each human reader evaluated the datasets.

#### 4.6 Specific Aim 4

##### Validation

A validation protocol was used to determine the average boundary error in the skin and chest wall segmentation, and sensitivity and specificity of the three classifiers and the five readers with and without the CAD system. ROC curve analysis [12] and multi-reader multi-case (MRMC) analysis [26] were used to evaluate the human readers performance.



#### 4.6.1 Chest Wall Detection

To validate the chest wall detection, the chest wall was manually identified for each of the 19 randomly selected datasets (a subset of the 75 datasets). For each point on the automatic segmentation, the distance to the nearest point on the manual segmentation was determined. Then for each dataset the average distance, the maximum distance and the RMS distance were computed. Each of these is then reported as the average plus and minus the standard deviation across all the datasets.

#### 4.6.2 Lesion Classification: CAD Performance

The performance of the CAD classification was assessed on 80 BMRI lesions. For each data set, the lesion was be pre-identified by an expert reader. Lesion biopsy results served as the standard of reference. The three classifiers were trained using leave-one-out method [10,27]. Sensitivity, specificity, and ROC analysis were used to evaluate performance. Figure 4.21 on the following page shows how the CAD system is validated.

#### 4.6.3 Reader Performance with and without CAD Assistance

The 80 BMRI lesions were randomly shown to five readers in one session. These five readers (one breast MR expert, two mammographers, and two body imaging fellows) were used to evaluate system performance. Each reader was asked to read each image first without using the CAD system, then with the CAD system for guidance. The CAD system provided the locations of suspicious lesions, a summary of the features and the classification for each lesion as either benign or malignant. The

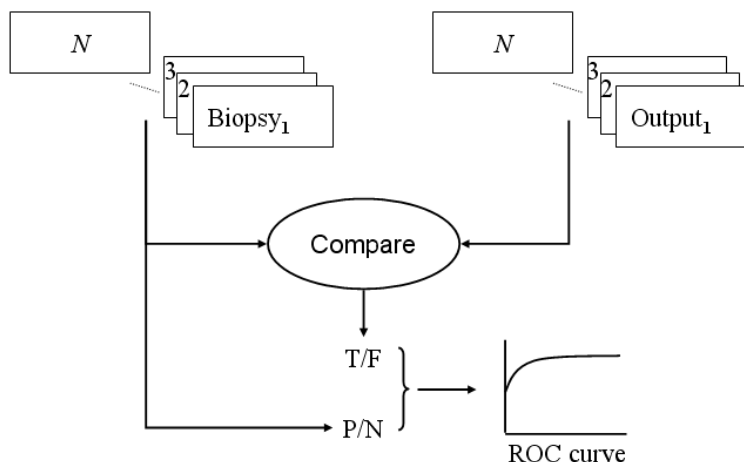


Figure 4.21: Validation of CAD system. For each dataset ( $n$ ), the classification of the CAD system ( $\text{Output}_n$ ), is compared with the biopsy result ( $\text{Biopsy}_n$ ). This is used to compute the sensitivity (True Positive Rate), and specificity (1-False Positive Rate) of the CAD system.

reader made the final determination concerning the classification of the breast lesion on a scale of [1 to 5], with 1 being “negative or no finding”, 2 meaning “definitely benign”, 3 is “probably benign”, 4 being “suspicious”, and 5 meaning “definitely malignant”. The reader may choose to use all, some or none of the information provided by the CAD system. Reader performance was evaluated using the biopsy result as the standard of reference. Performance was evaluated using ROC curve analysis [12] and multi-reader multi-case (MRMC) method [26, 40, 41].

#### 4.6.4 MRMC Method

In 1992, Dorfman *et. al.* proposed a method allowing generalization of ROC curves to the population of readers and cases [26]. Traditional ROC curve analysis focuses on the specific performance of the readers and cases in the study, but the real

goal is to understand the performance on any case, and for the population of readers. The multi-reader multi-case (MRMC) method generalizes the results by applying the jackknife method to the ROC curve statistics [82, 83, 92].

The jackknife method is a general method for reducing statistical bias and performing statistical tests. It involves creating pseudovalues which can be treated as independent and identically distributed. Pseudovalues are generated for a particular statistic under study (*e.g.* mean,  $A_z$ ). First, the statistic is computed using all cases for a particular reader ( $S_{\text{global}}$ ). Then, the statistic is re-computed once for each case ( $S_i$ , for  $i = 1, 2, \dots, Q$ ). The  $S_i$  statistic is computed by ignoring the  $i^{\text{th}}$  case. Finally, the pseudovalue is generated as  $P_i = QS_{\text{global}} - (Q - 1)S_i$ .

In MRMC, the statistic of interest is the ROC area under the curve  $A_z$ . For each reader,  $A_z$  is computed  $Q$  times ( $A_{z(i)}$  for  $i = 1, 2, \dots, Q$ ).  $A_z$  is computed by fitting the standard binormal model to the ROC rating data. For the  $i^{\text{th}}$  case,  $A_{z(i)}$  is the area under the curve for that reader, as though they did not score the  $i^{\text{th}}$  case. The  $Q$  pseudovalues ( $P_i$ ) are  $P_i = QA_z - (Q - 1)A_{z(i)}$ .

For our study, there were only five readers. An absolute minimum of ten readers is necessary to allow for generalization to the population, with 20 being a reasonable number [26]. Therefore, in the MRMC analysis, the readers were considered a fixed factor. The MRMC analysis was performed twice; the first time only the first lesion in each case was considered ( $Q = 75$ ), the second time all lesions were considered ( $Q = 80$ ).

## CHAPTER 5 RESULTS

In this chapter, we show the results of the four specific aims from the previous chapter in detail. First, we present the BMRI segmentation results. Second, we show the classification results for both mammography and BMRI. Then finally we show the classification results of each classifier, and the ROC curves for the human reader experiments.

### 5.1 Segmentation of BMRI

In this section, we present the chest wall segmentation results, then the segmentation of enhanced lesions using the various methods discussed in the previous chapter.

#### 5.1.1 Skin and Chest Wall Segmentation

Figure 5.1 on the next page shows what the data from a breast MRI looks like before processing. We show the pre-contrast and three post-contrast images. The lesion is more obvious in the post-contrast images, but is barely visible in the pre-contrast images.

Figure 5.2 on page 97 shows the result from a good agreement between the manual segmentation and the automatic segmentation. It is difficult to determine which one is “better”, because the smoothing that occurs manually may not be desirable, depending on the application. Figure 5.3 on page 98 shows one case where the matching does not agree. In this instance the chest wall boundary was very weak in the upper portion of the slice, and the manual segmentation chose a different path

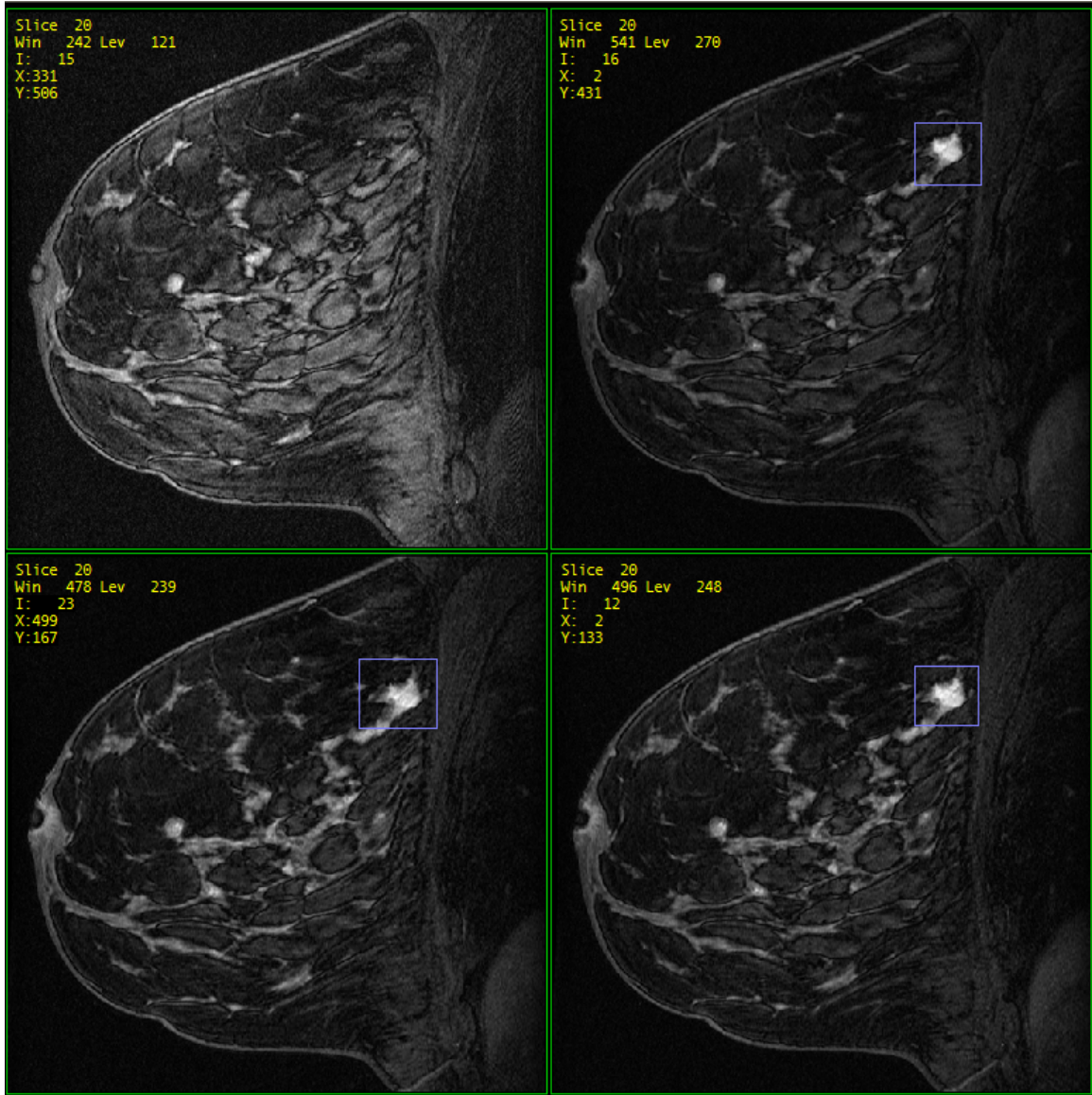


Figure 5.1: The pre-contrast and three post-contrast images. The lesion is marked with a purple box, it is more obvious in the post-contrast images, but is barely visible in the pre-contrast images. (Case: 5828, slice: 20).

than the automatic one. Figure 5.4 on page 99 shows a case where the automatic segmentation does poorly. Not enough of the breast tissue was removed, and the texture in the breast created a nice edge for dynamic programming to follow.

Table 5.1 on page 100 shows the average, maximum, and RMS difference between the automatic segmentation and the manual segmentation for the 19 datasets. The results show large variation in the average error (2.4-23.1 mm), because there are some datasets which have a very strong chest boundary, while in others it is almost impossible to distinguish. Also, because the manual segmentation wasn't done by an expert radiologist, it may not be a perfect gold standard. These data are also summarized in Figure 5.5 on page 101.

### 5.1.2 Segmentation of Enhanced Regions

Figure 5.6 on page 101 shows the difference images obtained by subtracting the pre-contrast sequence from the post-contrast sequences. We can see that the lesion enhanced more than the rest of the tissues. Figure 5.7 on page 102 shows the results after segmentation and connected components labeling.

In Figure 5.8 on page 102, the histogram follows a single modal distribution, so it is difficult to find the correct location to threshold. This was true in both the single slice case, as well as looking at the histogram for the entire dataset.

### 5.1.3 Segmentation Using Morphological Operations

Figure 5.9 on page 103 shows the gradient for a case with a mass, and Figure 5.10 on page 104 shows the morphology for the same case, where the mass was not

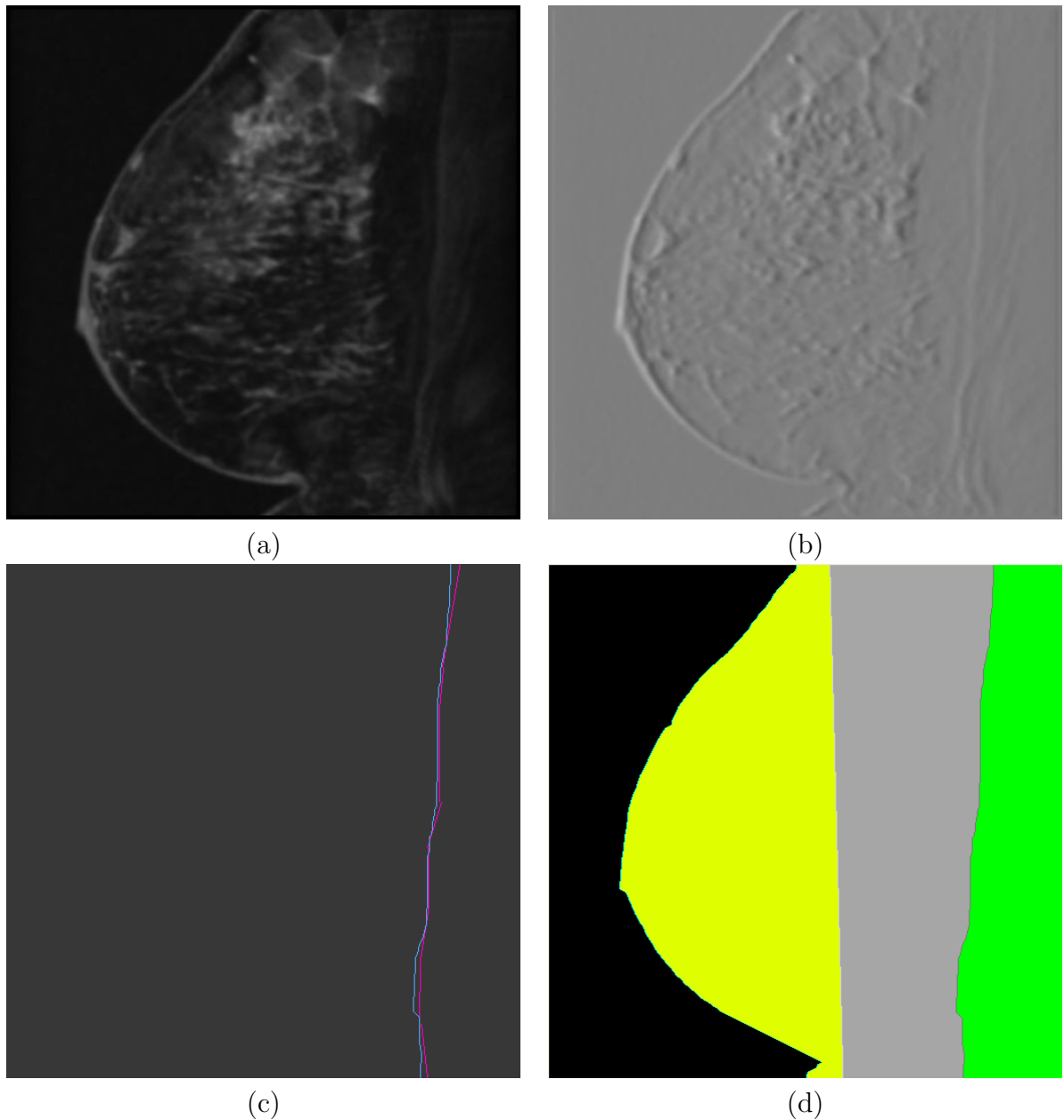


Figure 5.2: This figure shows a good match between the manual segmentation and the automatic segmentation. (Case: 5520, slice: 12) *a*) The blurred image. *b*) The cost image. *c*) The manual segmentation is the pink line; the automatic segmentation is the cyan line. *d*) The automatic segmentation showing the seven regions mask.

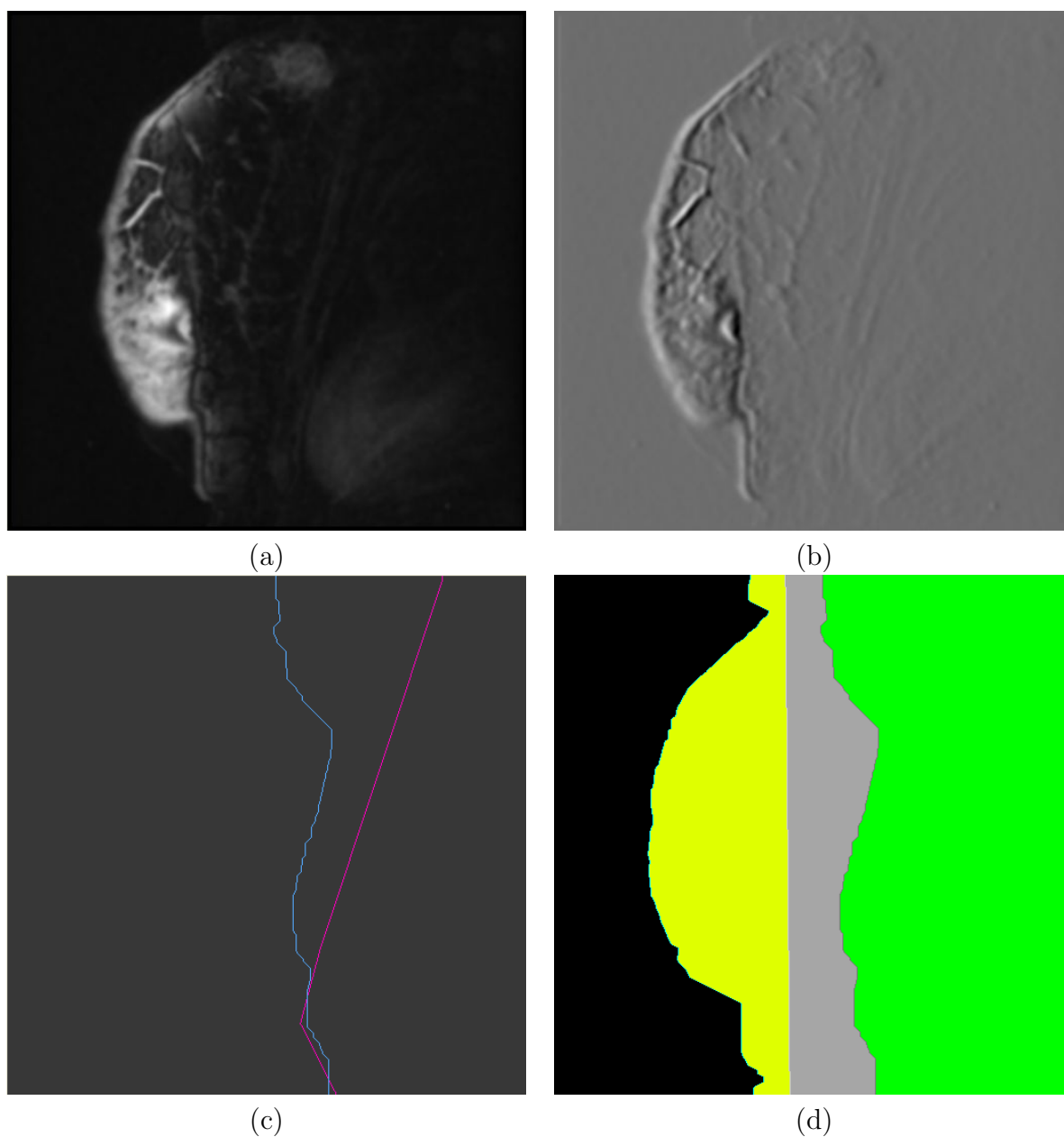


Figure 5.3: This figure shows what happens when there isn't an obvious boundary. In the upper right portion of the original image, the boundary is very dim. It isn't obvious if it continues on diagonally, or if it switches to a vertical line. Also, if it was a diagonal line, this breaks one of the assumptions that the chest wall boundary is generally vertical. (Case: 5549, slice 20) *a*) The blurred image. *b*) The cost image. *c*) The manual segmentation is the pink line, while the automatic segmentation is the cyan one. *d*) The automatic segmentation.



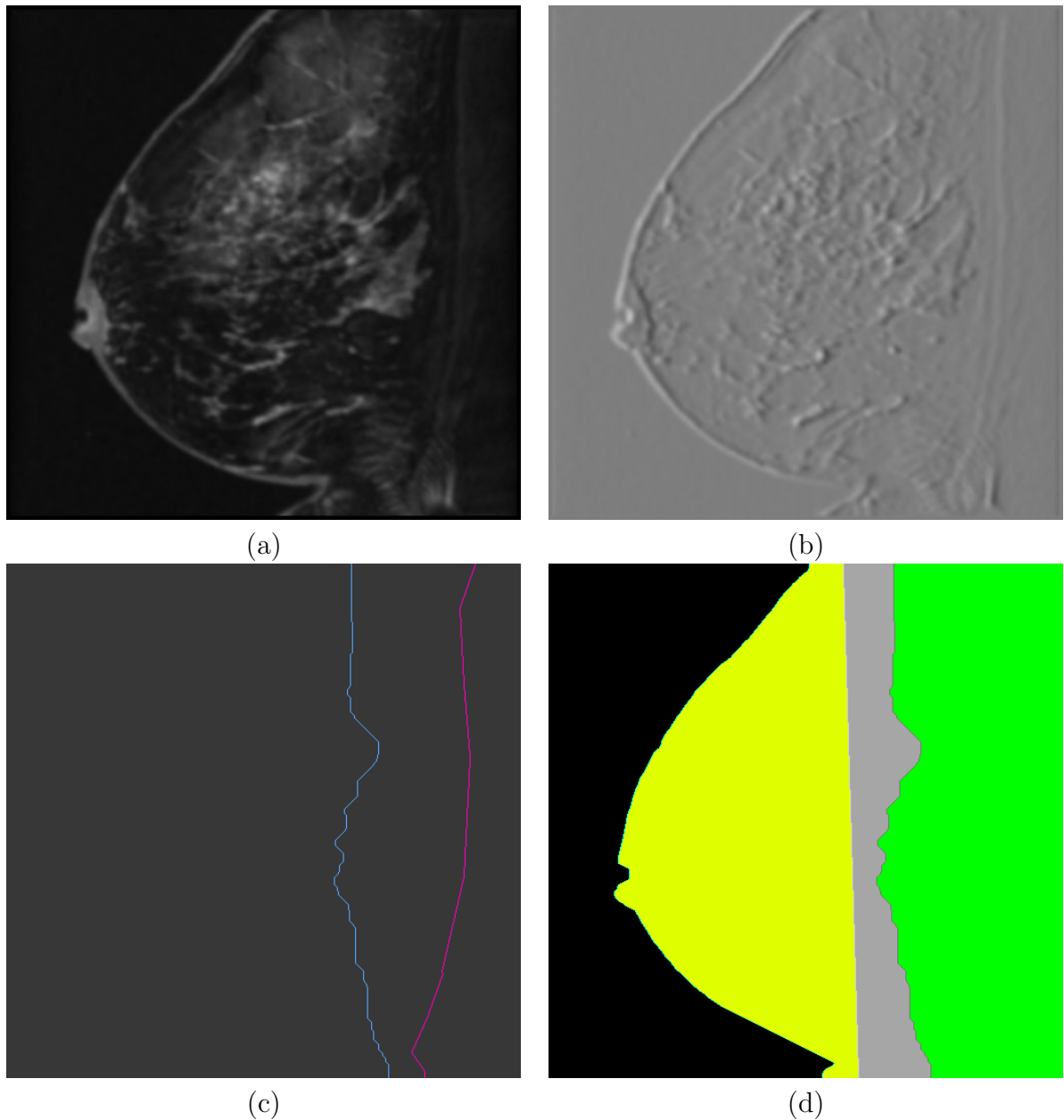


Figure 5.4: This figure shows the other common failure in the automatic algorithm for finding the chest wall. When the breast is not entirely within the field of view, the breast mask fails to cover all the breast tissue. Because breast tissue is generally heavily textured, it can appear as a stronger edge in the cost image (edge image) than the chest wall boundary. (Case: 5520, slice 9). *a*) The blurred image. *b*) The cost image. *c*) The manual segmentation is the pink line; the automatic segmentation is the cyan line. *d*) The automatic segmentation using DP.

Table 5.1: The unsigned difference between the manual and the automatic segmentation of the chest wall in mm.

Sequence	Average	St Dev	Max	RMS
1610	14.1	10.7	24.8	15.6
1806	12.0	5.0	21.9	12.8
5301	6.4	3.7	17.8	7.0
5419	10.7	5.4	16.0	11.9
5438	9.1	3.0	14.5	9.2
5489	2.7	3.3	10.4	4.3
5520	19.5	5.7	35.1	20.6
5525	9.0	6.9	25.8	9.8
5549	3.6	2.5	11.1	4.4
5552	17.5	6.2	22.7	18.4
5596	7.6	3.2	16.7	8.9
5616	23.1	7.7	30.8	26.0
6705	3.0	2.1	12.1	5.0
6819	16.3	10.5	29.4	18.2
6851	16.8	8.9	33.3	18.6
6869	3.5	1.5	13.8	5.3
6872	3.9	3.1	12.2	5.7
6873	2.4	2.0	14.6	5.5
6885	7.7	3.3	16.3	9.1
Total Avg	9.9	5.0	19.9	11.4
Total StDev	6.4	2.8	7.9	6.4

Note: Average is the average unsigned difference for each point between the two segmentations. St Dev is the standard deviation of the unsigned difference. Max is the maximum unsigned difference, and RMS is the Root Mean Square unsigned difference.

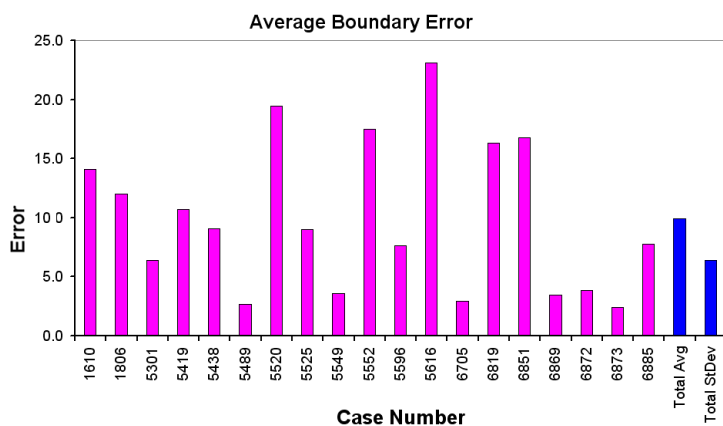


Figure 5.5: This graph shows the average boundary segmentation error for each datasets listed in table 5.1 on the preceding page. The blue bars are the average and standard deviation over all of the datasets.

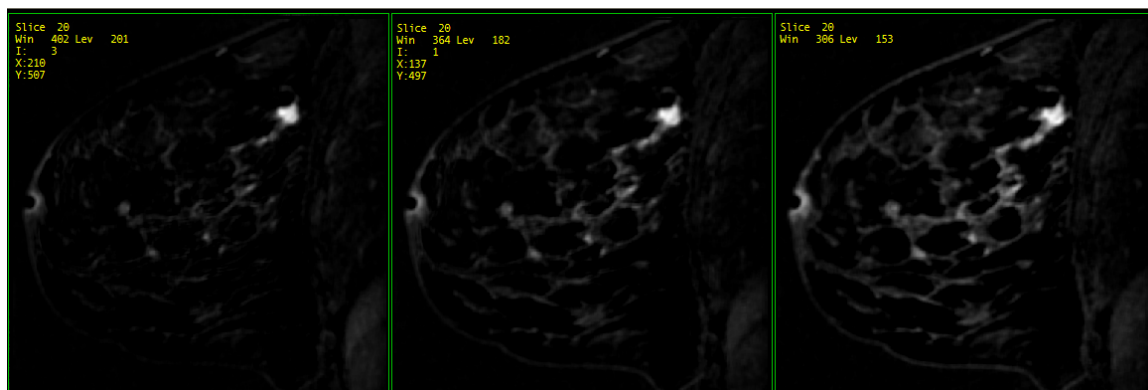


Figure 5.6: The difference images between the pre-contrast and each of the three post-contrast images (Case: 5828, slice: 20).

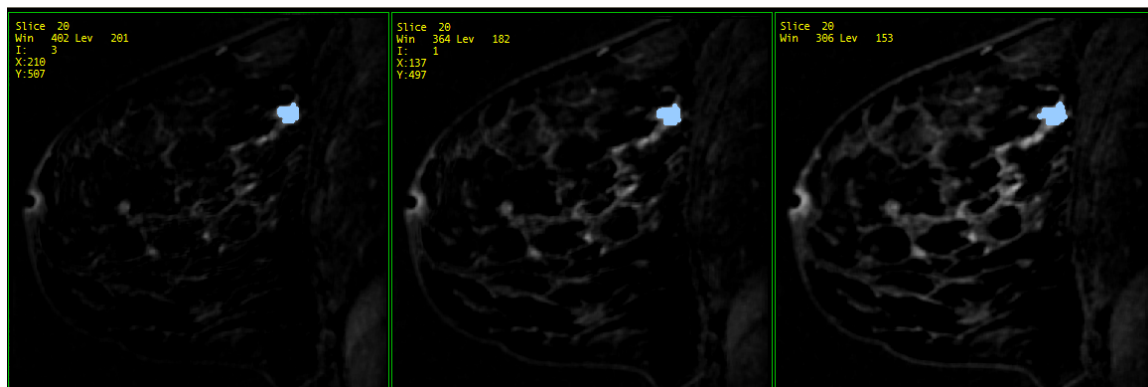


Figure 5.7: The labeled regions using 3D connected component algorithm (Case: 5828, slice: 20).

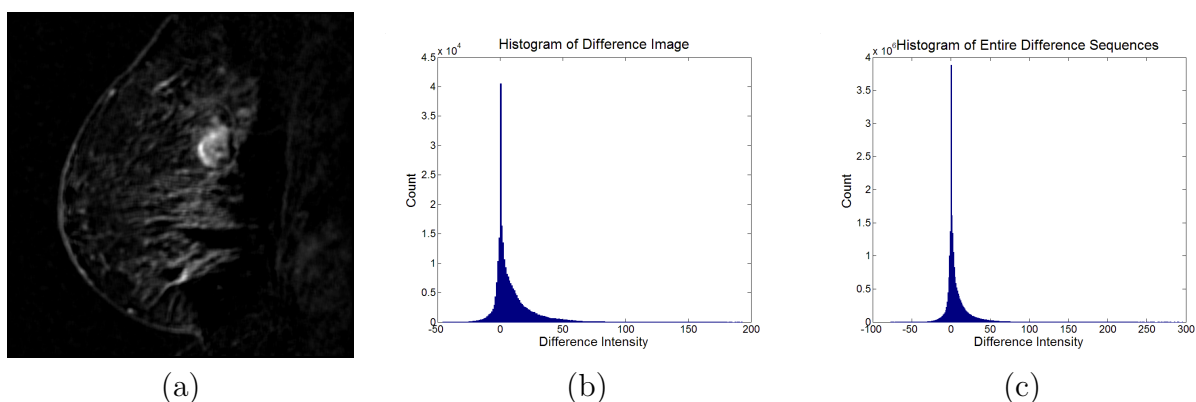


Figure 5.8: The histogram for the individual slice follows a very similar distribution to that of all the slices. Both of them follow a single modal distribution, which makes it difficult to find an optimal threshold. *a)* The difference image. *b)* The histogram of that slice. *c)* The histogram of all the difference images. (Case: 5520, slice: 15).

effected by applying morphology.

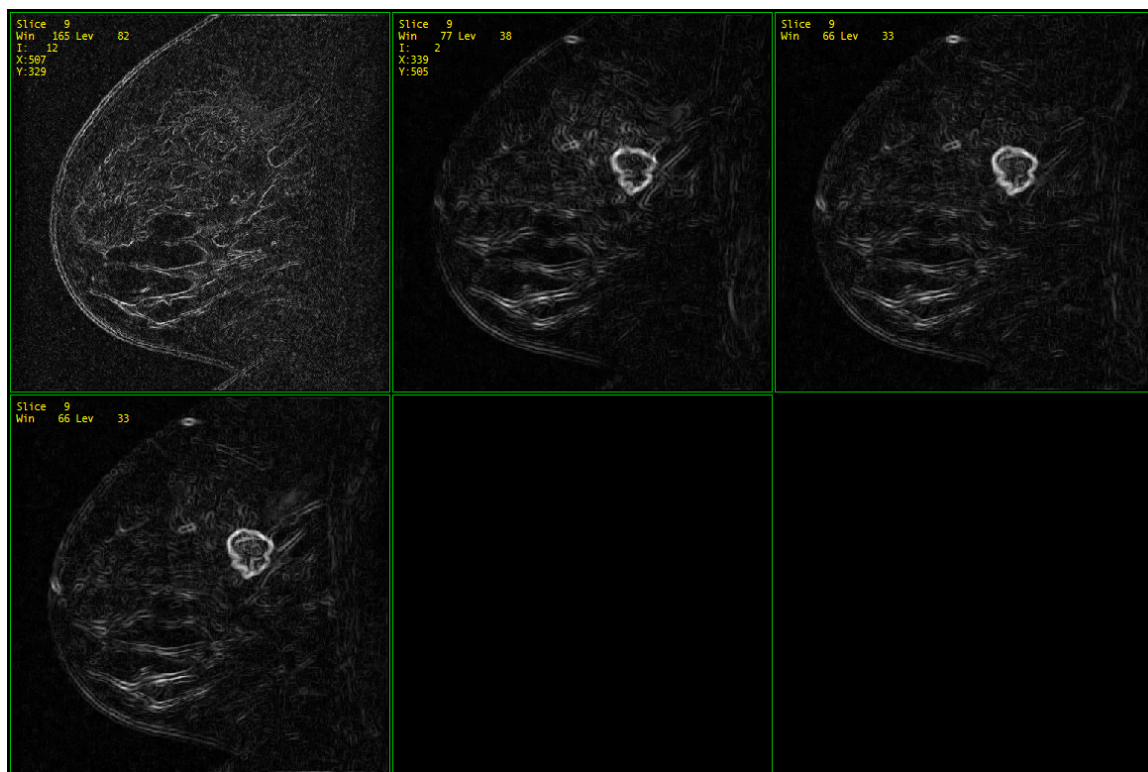


Figure 5.9: The gradient image of a case with a mass. (Case: 5489, slice: 9).

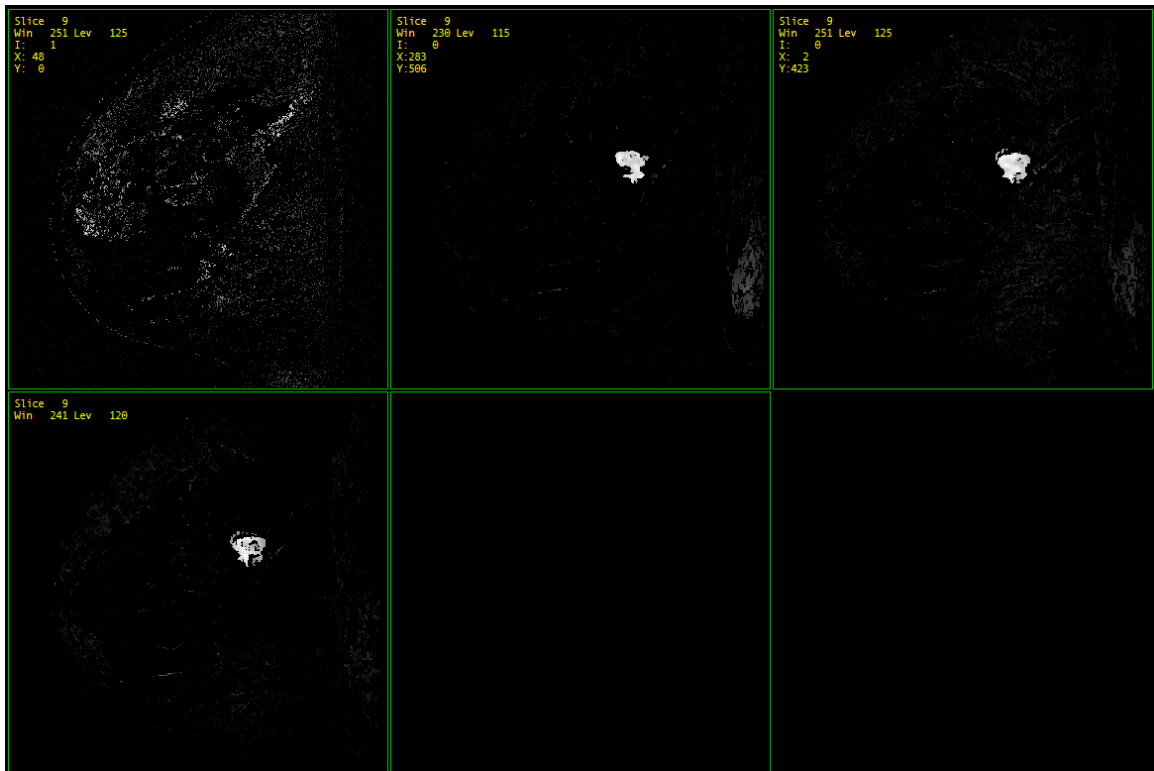


Figure 5.10: The morphology image of a case with a mass: The mass is still present, but the blood vessels have been removed. (Case: 5489, slice: 9).

## 5.2 Features Selection

Using the TOOLDIAG package to select the best features, we found the following 13 features: average relative intensity; RMS relative intensity; average spiculation; RMS spiculation; standard deviation of spiculation; average radius; RMS radius; average perimeter length; RMS perimeter length; average compactness; standard deviation of compactness; average area; and RMS area.

Figure 5.11 on the following page shows the time intensity curve for a mass, illustrating two sections: uptake and wash out. Table 5.2 shows the minimum, maximum, average, and RMS of five different features of the benign lesions, and Table 5.3 on the following page shows the minimum, maximum, average, and RMS of five different features of the malignant lesions

Table 5.2: The minimum, maximum, average, and RMS of five different features of the benign lesions.

Feature	Radius	Spiculation	Relative Intensity	Perimeter Length	Area
Minimum	1.88	3.36	0.00	1076.47	11.65
Maximum	18.65	5.77	176.64	6713.72	172.89
Average	9.81	4.97	94.33	4564.61	95.13
RMS	10.30	5.61	102.98	4599.01	95.65

The unit of radius, spiculation, perimeter length, and area is mm. Relative intensity is unitless.

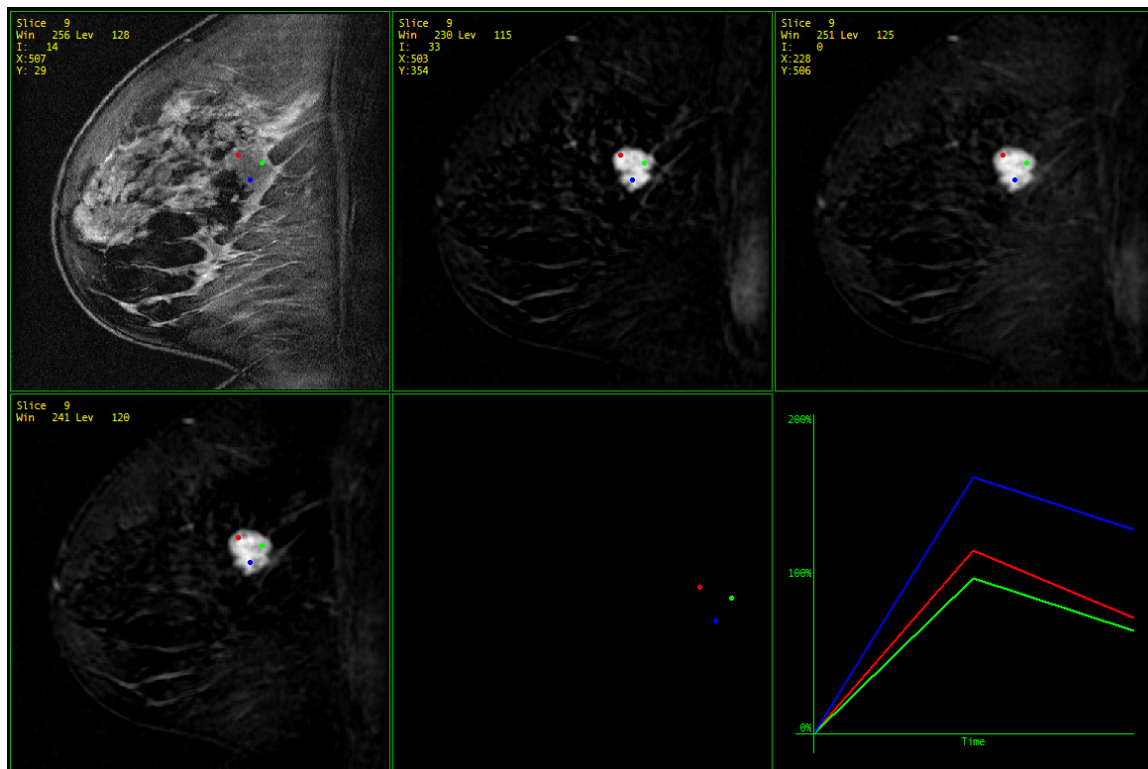


Figure 5.11: The time intensity curve of a mass consists of a uptake section and a wash out section. (Case: 5489, slice: 9).

Table 5.3: The minimum, maximum, average, and RMS of five different features of the malignant lesions.

Feature	Radius	Spiculation	Relative Intensity	Perimeter Length	Area
Minimum	3.34	3.40	3.42	2471.67	15.65
Maximum	23.17	7.72	276.64	8949.36	258.43
Average	12.27	6.11	195.10	6876.12	140.34
RMS	13.53	6.38	196.25	6984.18	143.96

The unit of radius, spiculation, perimeter length, and area is mm. Relative intensity is unitless.



### 5.3 Classification

In this section we give the results of the BMRI datasets classification. First, we show the results from our previous related work on classifying mammographic masses using neural networks. Second, we present the results of classifying the BMRI datasets using three classifiers: Bayesian classifier (BC), support vector machine (SVM), and backpropagation neural network (BNN). Finally, we show the results of the human reader experiments.

#### 5.3.1 Classifying Mammographic Masses

Figure 5.12 on the next page shows the ROC curve of using the BNN system for classifying the mammographic masses, the radiologists average performance, and the residents average performance. Table 5.4 shows the area under each of the ROC curves ( $A_z$ ), the standard error,  $p$  value, and the 95% confidence interval of the area.

$A_z$  was 0.923 for the BNN, 0.846 for the expert radiologists, and 0.648 for the residents. The BNN classifier output for malignant and benign masses was significantly different in the testing stage ( $p < 0.001$ ) [5]. These results illustrate the promise of BNN-based CAD systems for classifying breast masses in mammography.

Table 5.4: The area under each of the ROC curves ( $A_z$ ), the standard error (SE),  $p$  value, and the 95% confidence interval of the area.

Method	$A_z$	SE	$p$	95% CI
BNN	0.923	0.0760	$< 0.0001$	0.774 to 1.000
Radiologists Average	0.846	0.1317	0.0043	0.588 to 1.000
Residents Average	0.648	0.1321	0.1308	0.389 to 0.907

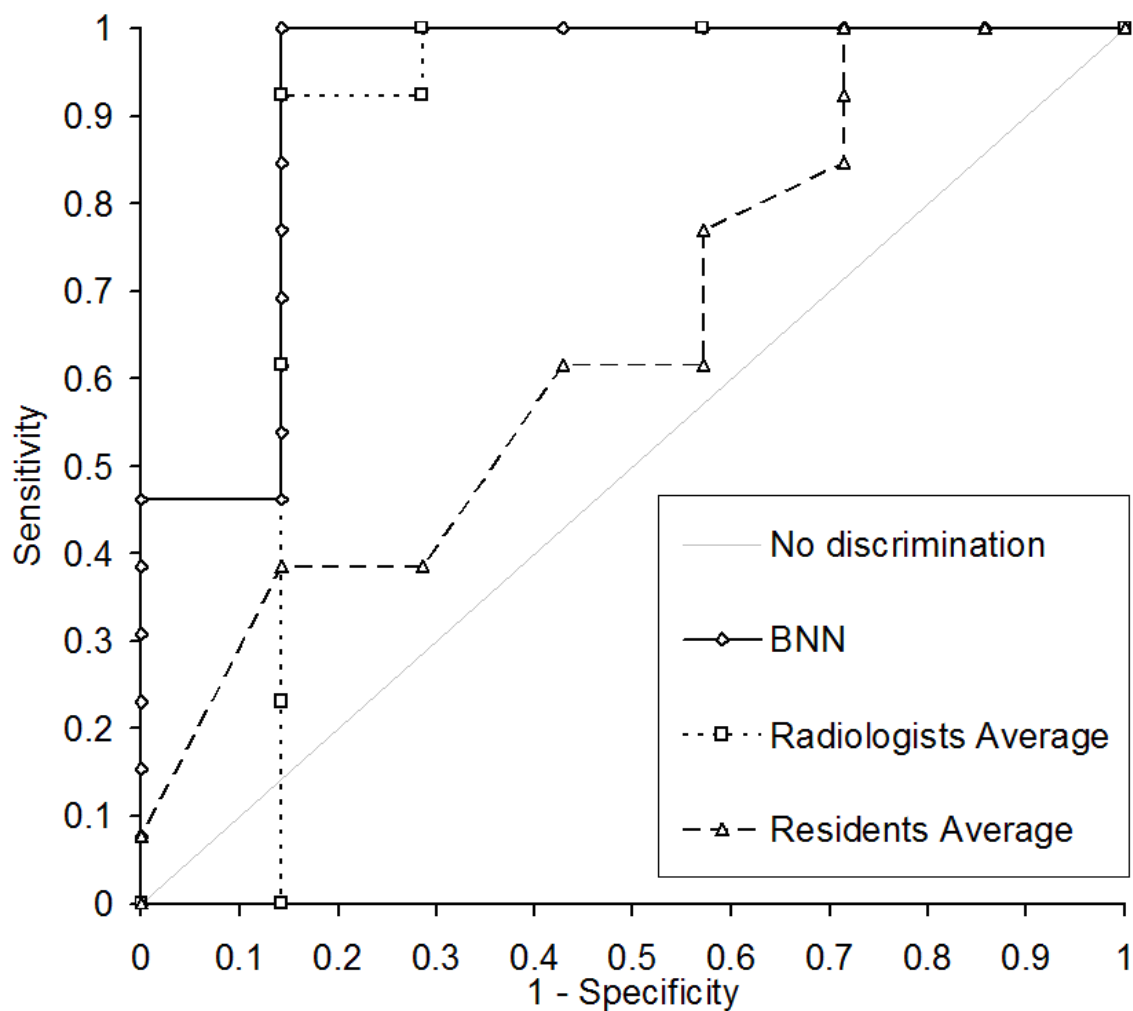


Figure 5.12: The ROC curve for classifying the mammographic masses.

### 5.3.2 Classifying BMRI Lesions

In this section we show the results of using trial and error method to choose  $N_h$ , and the results of classifying the BMRI datasets using three classifiers: the Bayesian classifier (BC), the support vector machine (SVM), and the backpropagation neural network (BMM).

Figure 5.13 on the next page shows the ROC curves of the BNN system when four, ten and twenty neurons were used in the hidden layer. Table 5.5 shows the area under the ROC curve ( $A_z$ ), the standard error,  $p$  value, and the 95% confidence interval of the area for each curve of Figure 5.13 on the next page. Although there was no significant difference in the BNN system output when three hidden neurons vs. four neurons were used, we choose to use the three neurons because  $A_z$  was slightly higher.

Table 5.5: The area under the ROC curve ( $A_z$ ), the standard error (SE),  $p$  value, and the 95% confidence interval of the area for the BNN system with different  $N_h$ .

BNN	$A_z$	SE	$p$	95% CI
$N_h = 3$	0.970	0.0237	< 0.0001	0.924 to 1.000
$N_h = 4$	0.964	0.0243	< 0.0001	0.917 to 1.000
$N_h = 10$	0.905	0.0362	< 0.0001	0.834 to 0.976
$N_h = 20$	0.717	0.0590	0.0001	0.602 to 0.833

Figure 5.14 on page 111 shows the ROC curves of each classifier. Table 5.6 on the next page shows the area under the ROC curve ( $A_z$ ), the standard error,  $p$  value, and the 95% confidence interval of the area.

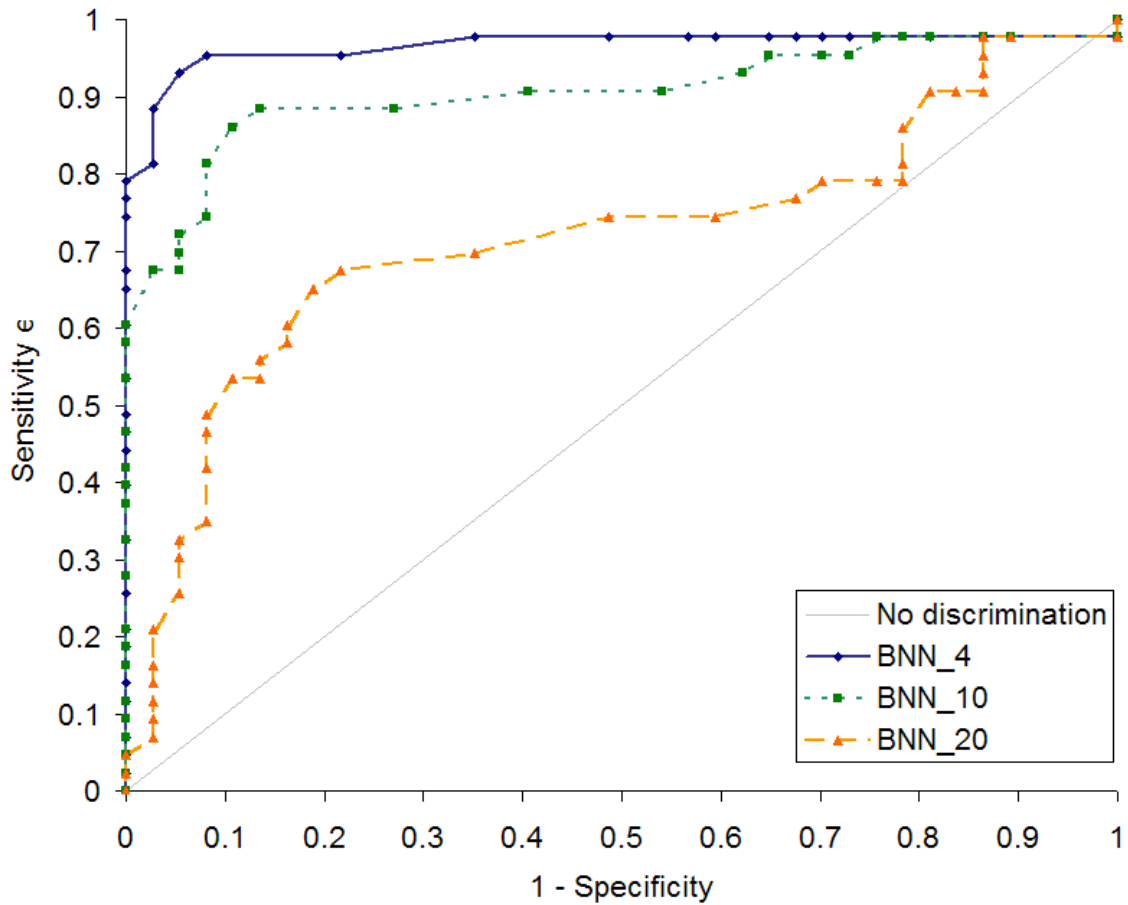


Figure 5.13: The ROC curves of of the BNN system when four, ten and twenty neurons were used in the hidden layer

Table 5.6: The area under the ROC curve ( $A_z$ ), the standard error (SE),  $p$  value, and the 95% confidence interval of the area for each classifier.

Method	$A_z$	SE	$p$	95% CI
BC	0.706	0.0586	0.0002	0.591 to 0.821
SVM	0.856	0.0485	< 0.0001	0.761 to 0.951
BNN	0.970	0.0237	< 0.0001	0.924 to 1.000

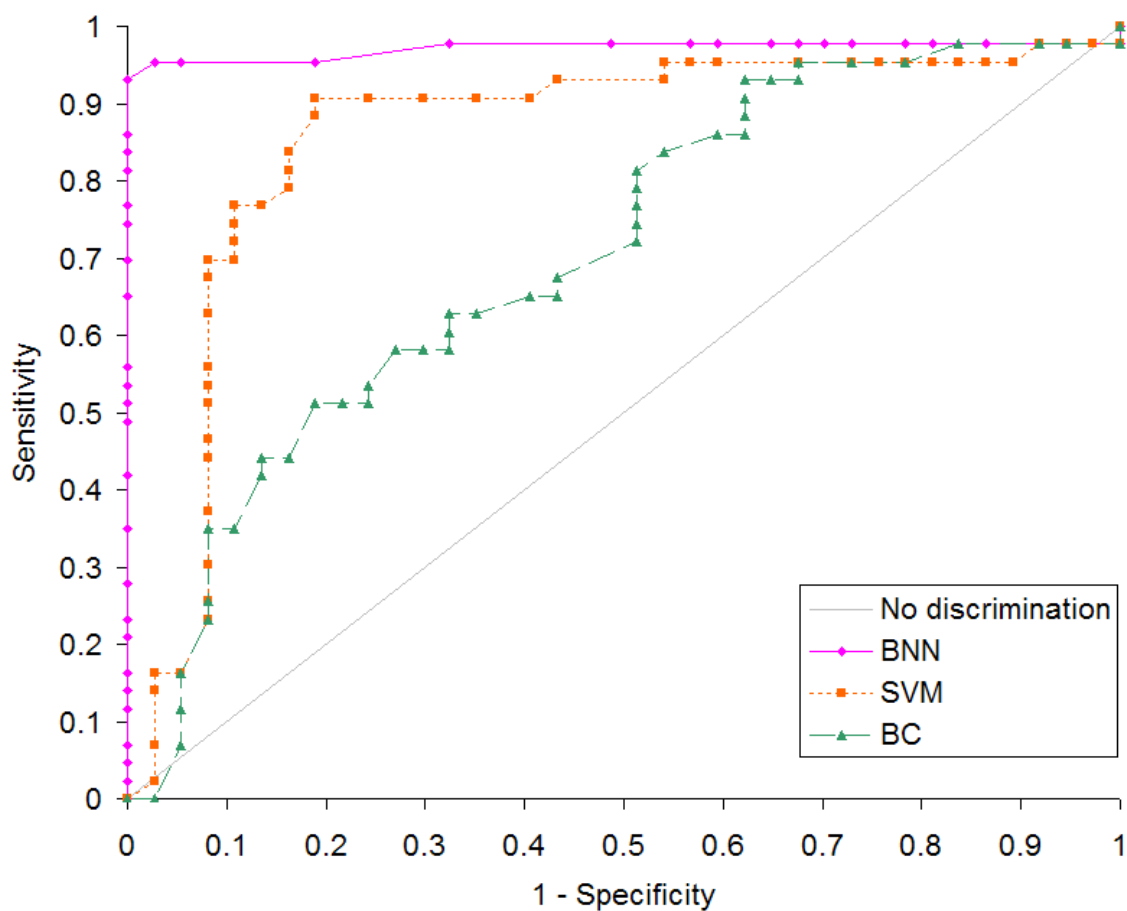


Figure 5.14: The ROC curves of the Bayesian Classifier (BC), the support vector machine (SVM), and the backpropagation neural network (BNN) for classifying the BMRI lesions.

### 5.3.3 Human Reader Experiments

In this section we present the results of the human reader experiments. For each human reader, we show the results of the reader with and without the CAD system assistance, the ROC curves of each reader, and the MRMC results.

#### 5.3.3.1 BMRI Expert

Figure 5.15 on the following page shows the ROC curve of the breast MRI expert for classifying the BMRI lesions. Table 5.7 shows the area under the ROC curve  $A_z$ , the standard error,  $p$  value, and 95% of the confidence interval of the area.

Table 5.7: The area under the ROC curve  $A_z$ , the standard error (SE),  $p$  value, and 95% of the confidence interval of the area for the BMRI expert.

Method	$A_z$	SE	$p$	95% CI
BMRI expert w/o	0.817	0.0498	< 0.0001	0.720 to 0.915
BMRI expert w/	0.933	0.0335	< 0.0001	0.868 to 0.999

#### 5.3.3.2 Mammographers

Figure 5.16 on page 114 shows the ROC curves of the two mammographers for classifying the BMRI lesions. Table 5.8 on the next page shows the area under the ROC curve ( $A_z$ ), the standard error,  $p$  value, and the 95% confidence interval of the area.

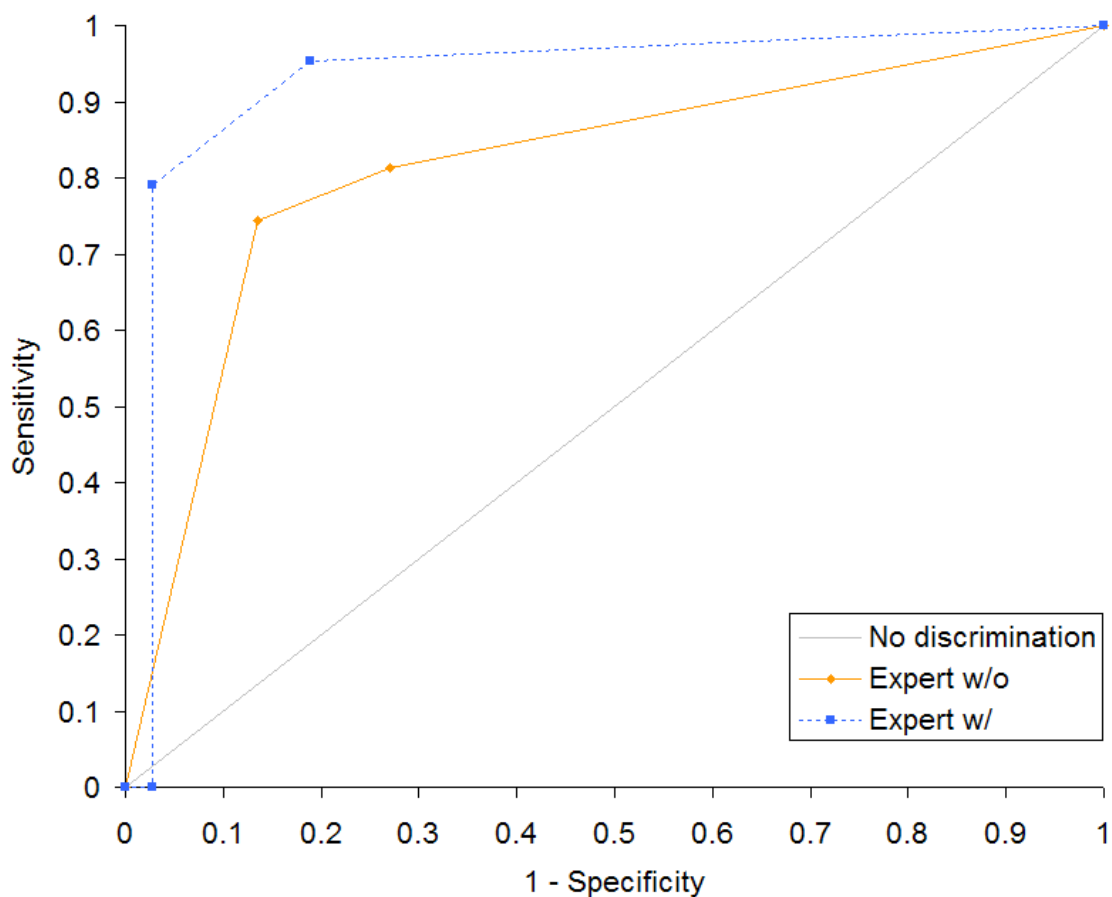


Figure 5.15: The ROC curve of the breast MRI expert for classifying the BMRI lesions.

Table 5.8: The area under the ROC curve ( $A_z$ ), the standard error (SE),  $p$  value, and the 95% confidence interval of the area for the two mammographers.

Method	$A_z$	SE	$p$	95% CI
Mam. 1 w/o	0.774	0.0559	< 0.0001	0.665 to 0.884
Mam. 1 w/	0.896	0.0422	< 0.0001	0.813 to 0.978
Mam. 2 w/o	0.788	0.0534	< 0.0001	0.684 to 0.893
Mam. 2 w/	0.906	0.0386	< 0.0001	0.830 to 0.982

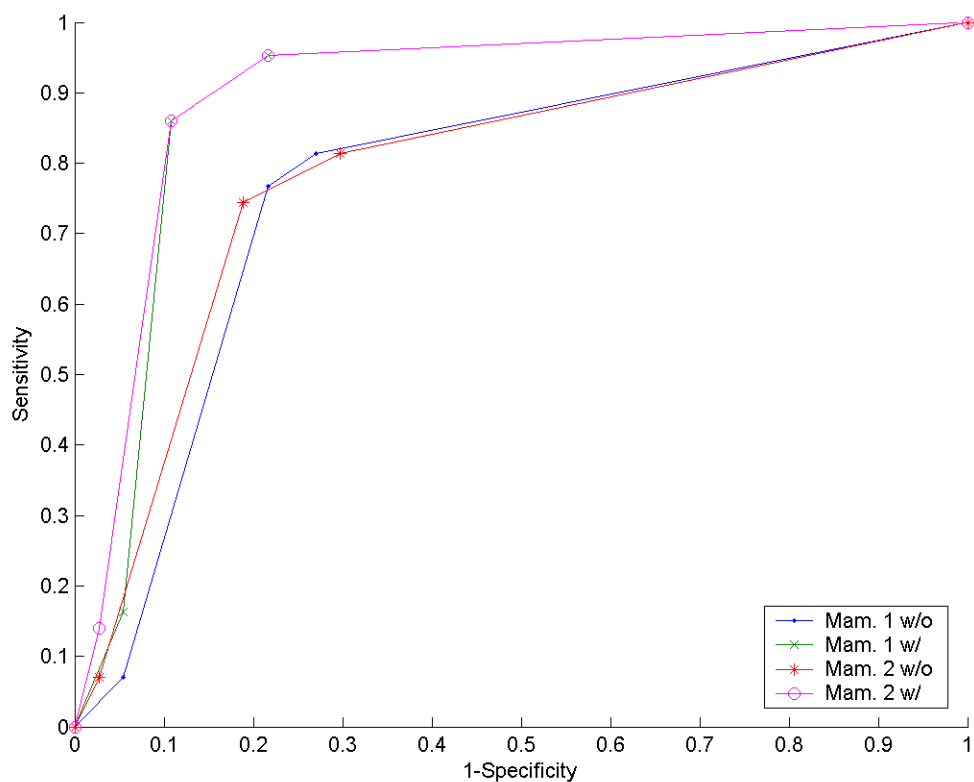


Figure 5.16: The ROC curves of the two mammographers for classifying the BMRI lesions.



### 5.3.3.3 Body Imaging Fellows

Figure 5.17 shows the ROC curves of the two body imaging fellows for classifying the BMRI lesions. Table 5.9 on the next page shows the area under the ROC curve ( $A_z$ ), the standard error,  $p$  value, and the 95% confidence interval of the area.

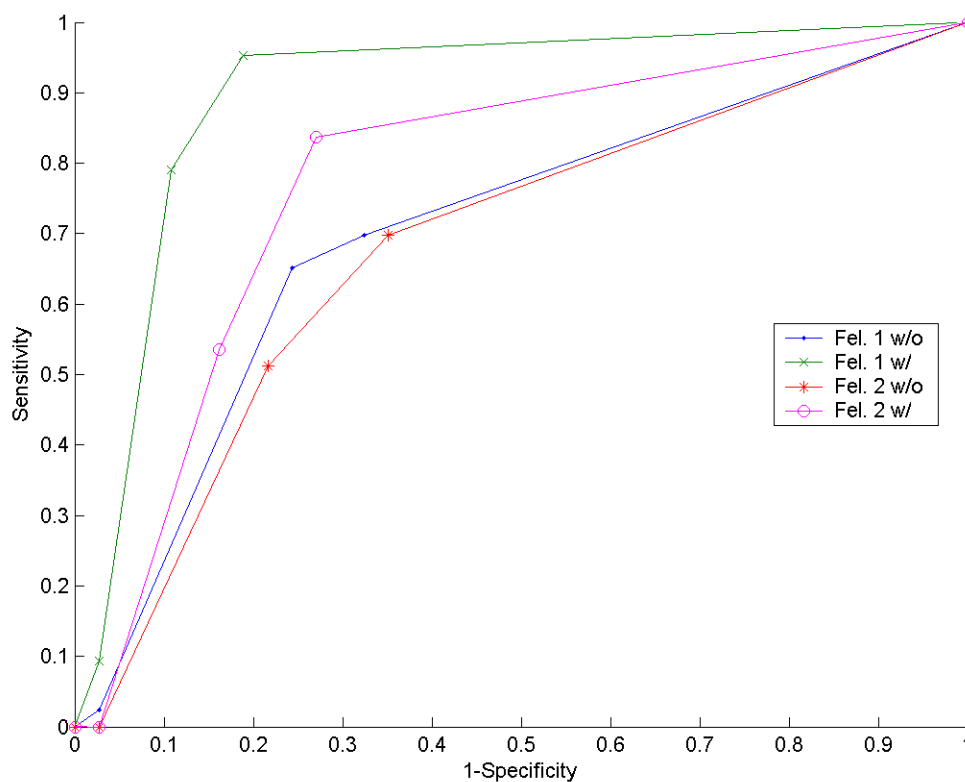


Figure 5.17: The ROC curves of the two body imaging fellows for classifying the BMRI lesions.

Table 5.9: The area under the ROC curve ( $A_z$ ), the standard error (SE),  $p$  value, and the 95% confidence interval of the area for the two body imaging fellows.

Method	$A_z$	SE	$p$	95% CI
Fel. 1 w/o	0.701	0.0599	0.0004	0.584 to 0.819
Fel. 1 w/	0.900	0.0402	< 0.0001	0.821 to 0.979
Fel. 2 w/o	0.681	0.0611	0.0015	0.561 to 0.800
Fel. 2 w/	0.781	0.0551	< 0.0001	0.673 to 0.889

#### 5.3.4 MRMC Results for Human Readers

When only the first abnormality was included for each case, the average area under the ROC curve for interpretation without CAD system assistance was  $A_z = 0.816$ , and with CAD system assistance was  $A_z = 0.907$ . The difference was statistically significant ( $F(1, 74) = 6.805$ ,  $p = 0.0110 < 0.05$ ).

When all abnormalities were included  $A_z = 0.812$  for interpretation without CAD assistance, and  $A_z = 0.912$  for interpretation with CAD assistance. The difference was also statistically significant ( $F(1, 79) = 8.94$ ,  $p = 0.0037 < 0.01$ ).

## CHAPTER 6 DISCUSSION

This chapter has three sections; in the first two sections we discuss our results. The first section is the BMRI segmentation discussion, and the second is the classification of the BMRI lesions. The classification discussion includes the ROC curve and the MRMC analyses. In the third section, we summarize this study.

### 6.1 BMRI Segmentation

In this section we discuss the BMRI segmentation results. This includes the skin and chest wall segmentation, and the enhanced lesions segmentation.

#### 6.1.1 Skin and Chest Wall Segmentation

There are two factors that confound the search for the chest wall. First, the chest wall is not always very visible in an image. This can be seen in Figure 5.3 on page 98. Second, if the breast tissue contains a lot of texture, the textural features may appear as edges to the dynamic programming step, and thus can be mistaken for the chest wall as seen in Figure 5.4 on page 99. Both of these features are caused by the way the images were acquired. Because these are fat suppressed images, the ducts appear enhanced, and this causes the parenchymal texture. If another pulse sequence, such as a  $T_1$  weighting, were used, breast fat would appear bright and the parenchyma and most ducts would appear dark and it would be easier to find the chest wall. However, this sequence wouldn't be as effective for finding disease, but it might be a suitable sequence to acquire to aid the post processing steps.

### 6.1.2 Enhanced Lesions Segmentation

When attempting automatic segmentation, there were two primary factors that we were unable to compensate for. First, MRI is inherently an imaging modality with inconsistent signal intensity. MRI signal intensity values are qualitative instead of quantitative. So, even if you scan the same patient twice on the same day, you will get different MR image intensity values. In our study, the range of pixel values could be 300 units in one case, and 1000 units in another. Figure 6.1 on page 120 shows a dataset with enhanced lesion that has intensity of approximately 300 units. Figure 6.2 on page 121 shows another dataset with an enhanced lesion that has intensity of approximately 1000 units.

The second significant factor is the overabundance of enhancing regions, many of which are not significant. Blood vessels enhance very strongly (as shown in Figure 6.3 on page 122), but are in no way indicative of malignancy. In most scans, the heart is visible. Since it is also a highly perfused region, it also enhances strongly (as shown in Figure 6.4 on page 123).

Because of the challenges of implementing fully automatic segmentation, we selected an alternative strategy. Instead of having the computer sort through all enhancing regions, trying to determine which ones are significant, we allow the user to select seed points for the computer to focus on.

**Segmentation by Histogram** In all the datasets, the voxel histogram was unimodal as shown in Figure 5.8 on page 102. The histograms of individual slices, as well as the histogram for the entire dataset, did not reveal an obvious threshold value

to separate the breast lesion from the background.

**Segmentation by Morphology** This segmentation method can be improved by applying a low pass filter to blur the gradient image. This is because a blood vessel might be constant on the inside, but have a sharp edge where it meets other tissue. By blurring the gradient, you can spread out the effects of a large gradient, and overcome small areas of homogeneity.

## 6.2 BMRI Classification

In this section we discuss features computation, and BMRI lesion classification for the three artificial classifiers, and for the human reader experiments.

### 6.2.1 Features Computation

Because scans are not isotropic and our features were calculated in the image plane with the highest resolution, changing the perspective (CC vs. MLO) can effect the measurement of lesions. However, most lesions are fairly isomorphic (radically symmetric) and thus would appear similar in both views.

### 6.2.2 Artificial Classifiers

The classification used 75 pathology-proven datasets. Thirteen shape, texture, and enhancement kinetics based features computed for 80 lesions pre-identified by the pathology locations. Lesion classification is applied to distinguish between benign and malignant lesions. The features for each lesion were used as inputs to the backpropagation neural network (BNN), the support vector machine (SVM), and the

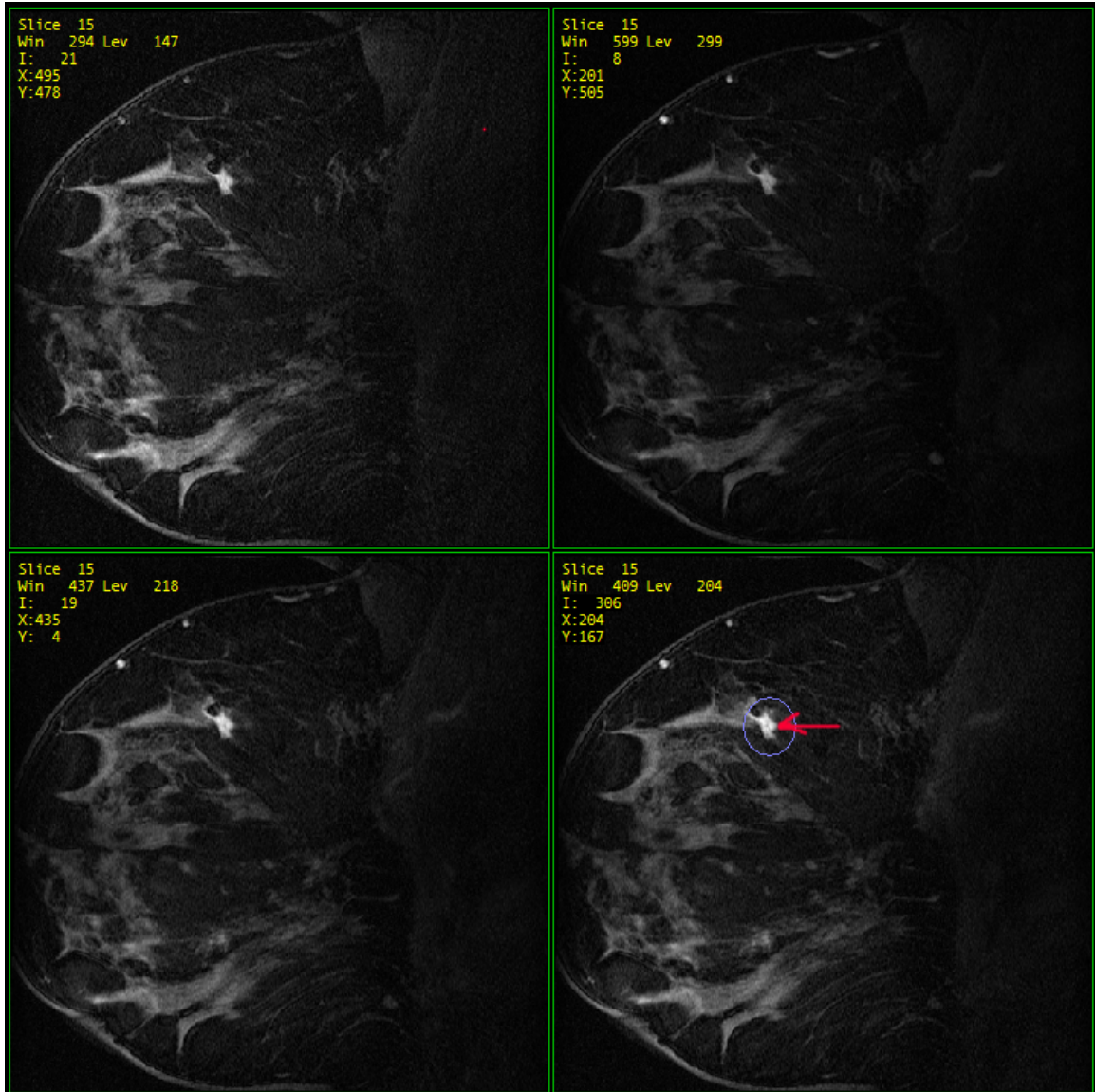


Figure 6.1: MRI signal intensity values are qualitative. This Figure shows a dataset that has an enhanced lesion (red arrow) with intensity of 306. (Case: 1809, Slice: 15).

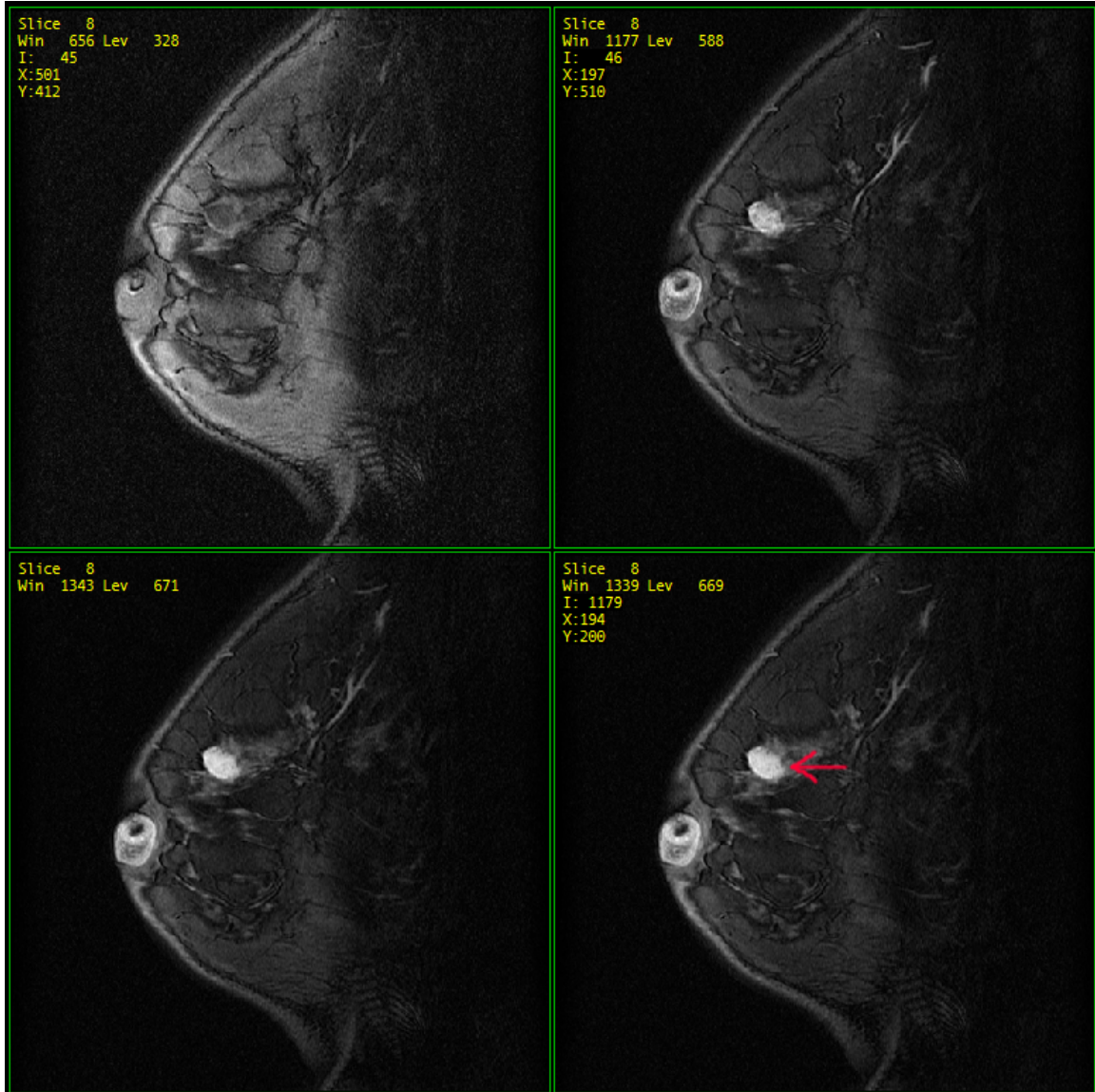


Figure 6.2: MRI signal intensity values are qualitative. This Figure shows a dataset that has an enhanced lesion (red arrow) with intensity of 1179. (Case: 7869, Slice: 8).



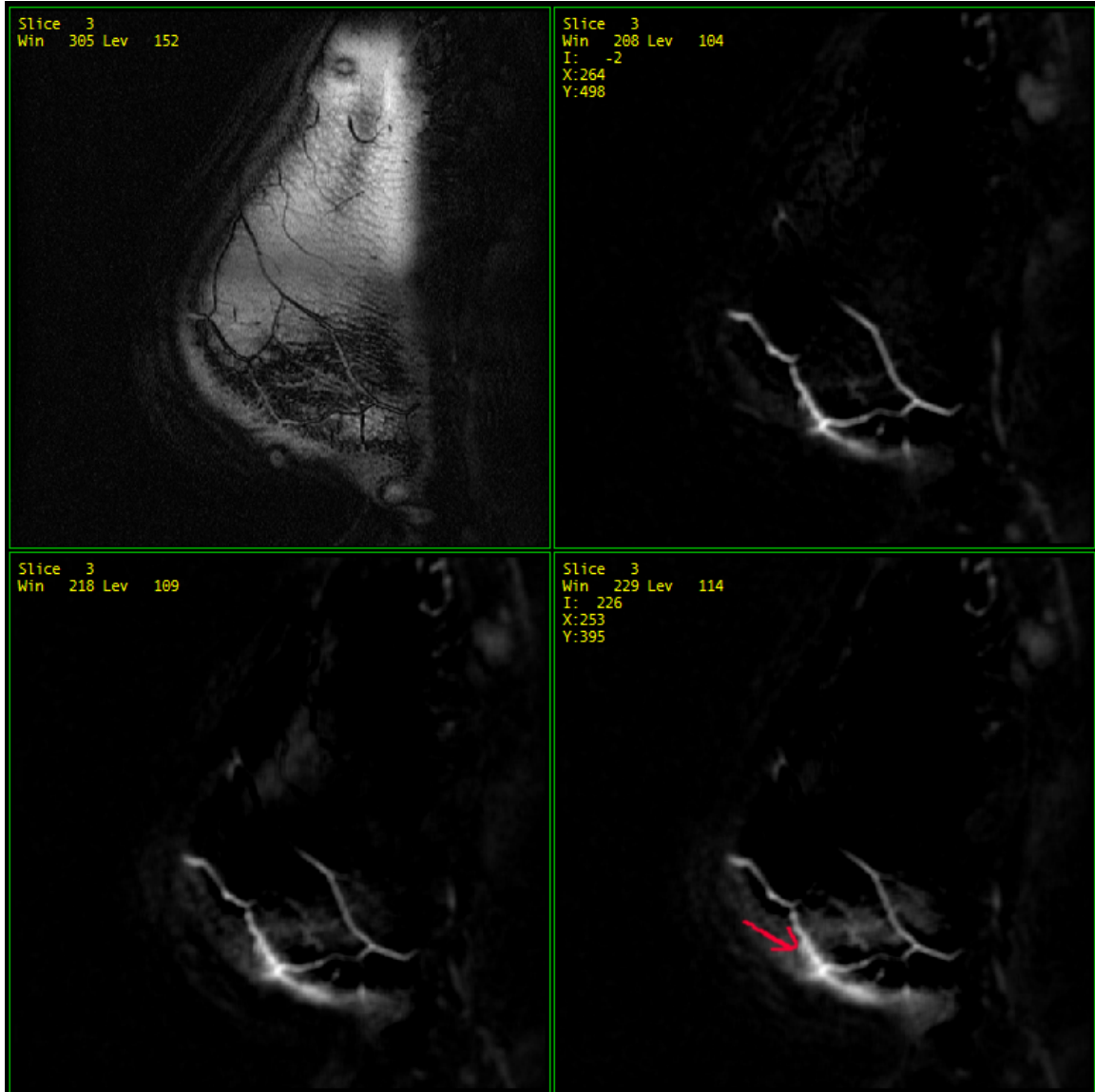


Figure 6.3: This figure shows that blood vessels (red arrow) enhance very strongly, but it is not a malignancy indication. (Case: 1806, Slice: 3).



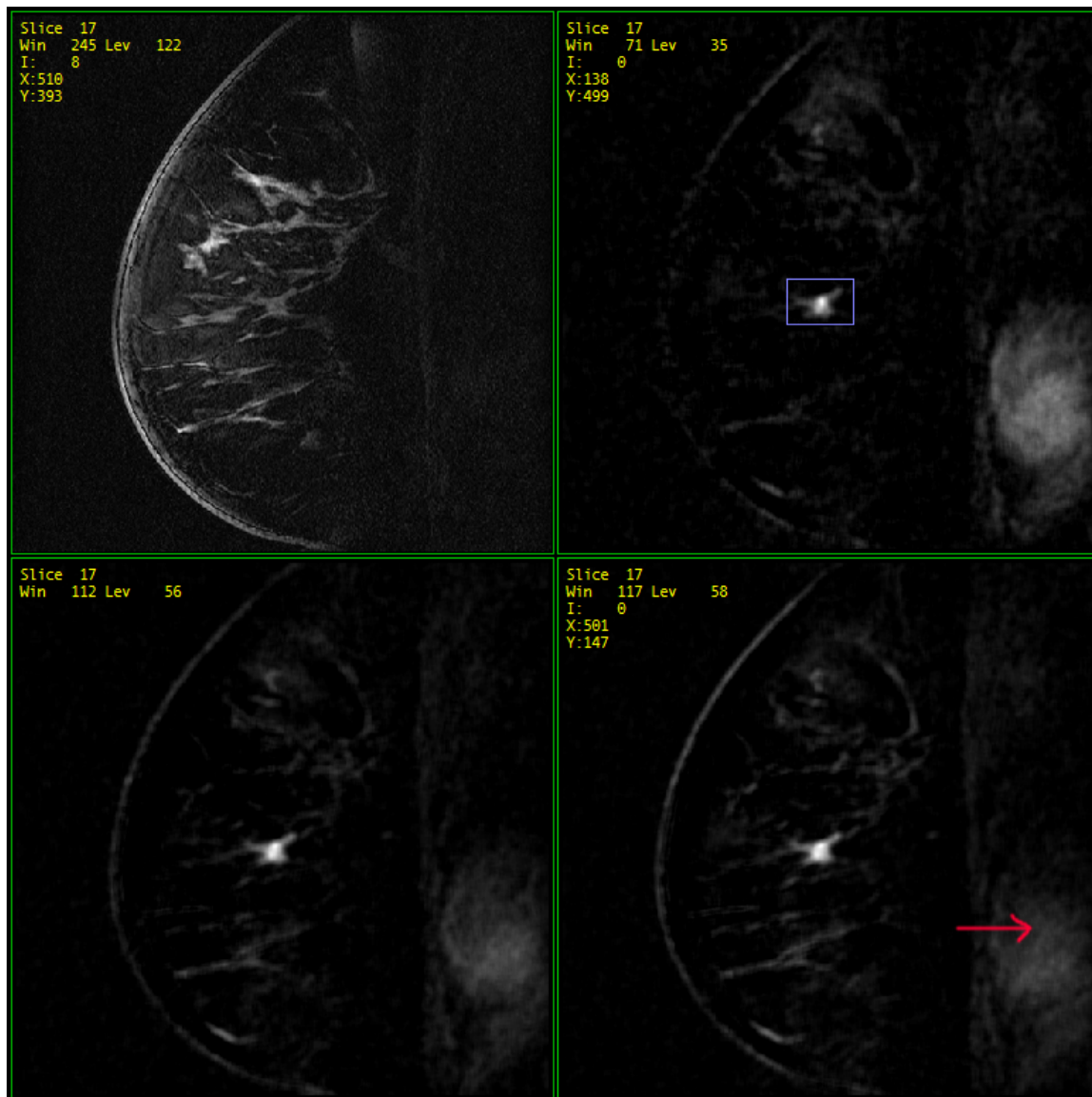


Figure 6.4: This figure shows a dataset where the heart (red arrow) is visible. (Case: 1610, Slice: 17).

Bayesian classifier (BC), using the biopsy results as the gold standard. All classifiers were trained using leave-one-out method.

Classification results from only the BNN system were used in the human reader experiments. For this aspect of the work, we used a three layer neural network with three neurons in the hidden layer and one neuron in the output layer. The back-propagation algorithm was used for training. During training, the backpropagation neural network (BNN) output was compared against the biopsy results (the targets). The mean-square difference with a minimum performance of 0.1 was used as the error measure. The Levenberg-Marquardt (L-M) optimization method was used during training.

### 6.2.2.1 Classifier Differences

Table 6.1 shows the differences in the BNN system performance for different number of hided neurons. There was no significant difference in the BNN system output when three hidden neurons vs. four neurons were used, we choose to use the three neurons because  $A_z$  was slightly higher.

Table 6.1: The differences in the BNN system performance for different number of hided neurons in terms of ( $A_z$ ) along with the  $p$  value associated with each difference.

Contrast	Difference	$p$
$N_h = 3$ vs. $N_h = 4$	0.006	–
$N_h = 4$ vs. $N_h = 20$	0.247	< 0.0001
$N_h = 10$ vs. $N_h = 20$	0.188	< 0.0001

Figure 5.14 on page 111 shows the ROC curves of each classifier. Table 5.6 on page 110 shows the area under the ROC curve ( $A_z$ ), the standard error,  $p$  value, and the %95 confidence interval of the area. The  $p$  value represents the probability that observed difference between two groups are due to random variation from a single distribution. Thus, in the testing stage of the BNN and the SVM systems (see Table 5.6 on page 110), the  $p$  value shows a statistically significant difference between their output for malignant vs. benign lesions ( $p < 0.0001$ ). This is also true for the BC system, but with ( $p = 0.0002$ ).

Table 6.2 shows the differences among the three classifiers in terms of ( $A_z$ ) along with the  $p$  value associated with each difference.

Table 6.2: The differences among the three classifiers in terms of ( $A_z$ ) along with the  $p$  value associated with each difference.

Contrast	Difference	$p$
BNN vs. SVM	0.114	0.0287
BNN vs. BC	0.264	< 0.0001
SVM vs. BC	0.149	0.0470

The  $p$  values in Table 6.2 indicate significant differences among the three classifiers ( $p < 0.05$ ), with the largest difference between the BNN system and the Bayesian classifier ( $A_z(\text{BNN}) - A_z(\text{BC}) = 0.264$ ), and the smallest difference between the BNN system and the SVM ( $A_z(\text{BNN}) - A_z(\text{SVM}) = 0.114$ ).

In theory, both support vector machines and Bayesian classifiers have the ability to perform as well as a neural network. The differences primarily occur in

what assumptions are made, and how you train the system.

An SVM only operates as a linear separator. You are able to overcome this deficiency by generating pseudo-features from your input data (*e.g.* using  $x_1 \times x_2$  as a feature, instead of only  $x_1$  and  $x_2$ ). An SVM can also bypass the explicit computation of the pseudo-features by using the kernel trick. However, this kernel is still performing a specific transformation of the input data (*e.g.* a Gaussian kernel is performing a Gaussian transformation on the data). Proper selection of pseudo-features (or kernel) is very important to the performance of an SVM. This is a very laborious process, as it is unknown in advance what the proper transformation is. Adding many (pseudo-)features is not the solution either, because then you run into the problems of sampling density and feature space size. SVMs are designed to be fast at training, and once they are trained, they generalize over the populations very well (*i.e.* they train fast, and they are difficult to over-train).

A Bayesian classifier also, in theory, can perform identically to any other system. It is modeled on the probabilities that a particular value of a feature belongs to a specific class. This is very general, and with appropriate probabilities, it allows very accurate classification. The difficulty is in defining the probabilities. With  $Q$  features, this is a  $Q$ -dimensional surface defining the probability a sample from particular class will have the specific values of all the features (there is one surface per class, in our case two). This provides great flexibility, as you can define whatever surface you want, but in practice it may be flexible. In our case, we made the simplification that the probability surface could be modeled as a  $Q$ -dimensional Gaussian. Most likely,

this is not an accurate model for our data, but it would be a time consuming process and require more data to try many models.

A primary advantage of using a neural network is that it does not require the user to do as much work on selecting features, or selecting an appropriate model for their data. The training stage of the network is designed to automatically determine what features are significant, how they are correlated, etc. Basically, it does not require as much *a priori* knowledge to get a working system. The trade-off is that more time must be spent in the training stage, it is very possible to over-train your network, and if your training data is not representative of the population you are likely to get poor performance.

The first limitation of a neural network is not that severe. Our SVM could be trained in less than a second, while the neural network might take 30 seconds or more. While this is a difference of approximately two orders of magnitude, the actual length of time is not that severe. The second limitation is handled by controlling the number of hidden nodes in the network. With a limited number of nodes, the network is forced to generalize. This means that performance may seem to go down with the training set, but on the full population, the performance may actually increase. The third limitation is a general issue of systems which do not use *a priori* knowledge, since all knowledge must be present in the training data. Proper selection of training data, by increasing number of samples and making sure the samples are properly randomized, should easily overcome this limitation. At first glance, the order of training samples appears to matter, since that the network state (weights) are updated. However,

since the training is continued for many cycles (100 epochs), it is unlikely to have significant effect.

Since there are not many publications on automatic classification for BMRI lesions, we will compare our results to that of automatic classification of mammographic masses. In terms of  $A_z$  value, the BNN system classification results ( $A_z = 0.970$ ) is better than the results reported by other groups like Kupinski *et al.* ( $A_z = 0.95$ ) [50] and Huo *et al.* ( $A_z = 0.94$ ) [45]. Table 2.1 on page 15 shows a comparison among different studies used artificial neural networks to classify mammographic masses.

The BNN performed the best among the three classifiers. This is due in part to our previous experience in designing neural network [4–6]. Using features based on shape, texture, and enhancement kinetics of the BMRI lesion improved the performance of the BNN classifier compared with the performance when the shape features were used alone ( $A_z = 0.913$ ) [8].

When we chose an operating point where the number of false negatives (FN) is zero, we were able to maintain specificity of 100%, but the sensitivity was dropped to 93% for the BNN system. Whereas, in both SVM and BC systems we were not able to reach FN=0.

### 6.2.2.2 Human Reader Experiments

In the human reader experiments, we discuss the results for both the ROC curve analysis, and the multi-reader multi-case (MRMC) analysis.

**ROC Curve Analysis** Figure 6.5 shows the ROC curves of the BNN system, BMRI expert without using the BNN assistance, and the BMRI expert with the BNN system assistance for classifying the 80 BMRI lesions. Table 6.3 on the next page shows the differences between the BMRI expert with and without the BNN assistance vs. the BNN system classification.

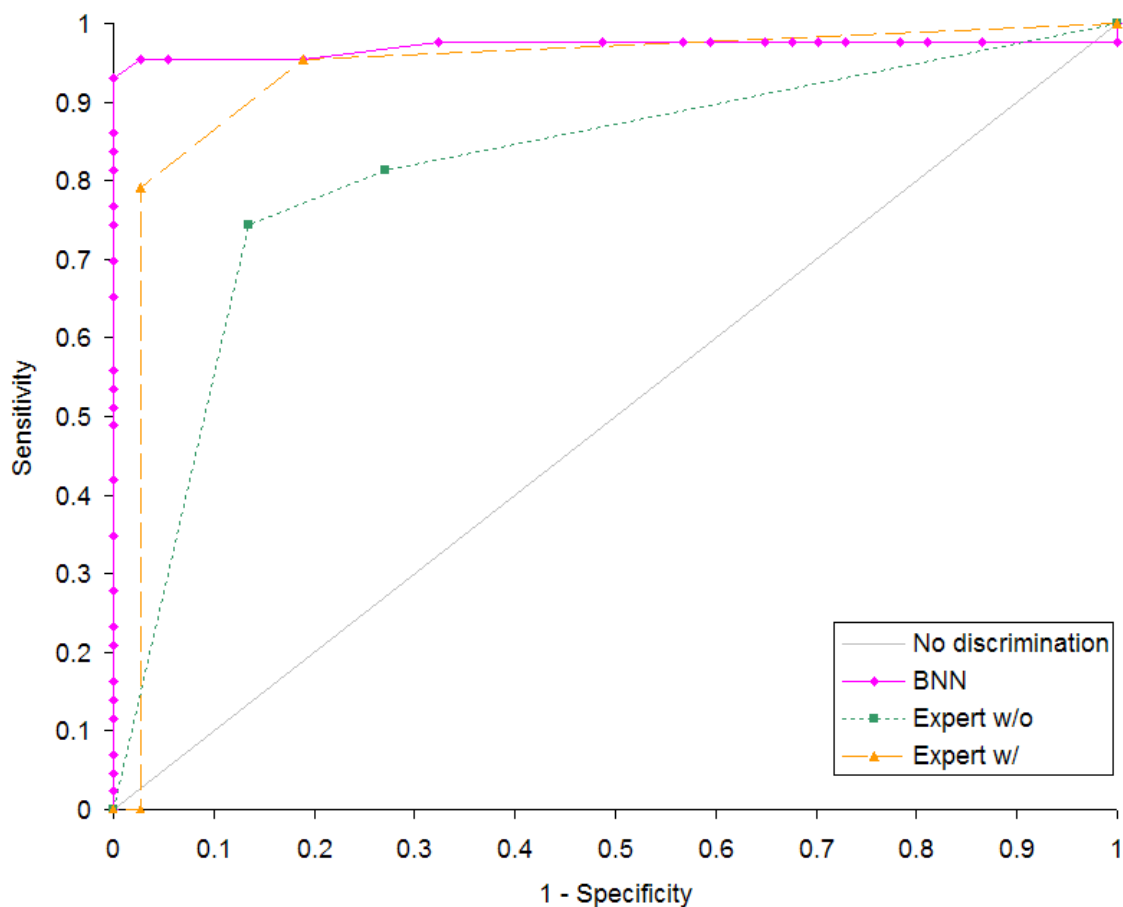


Figure 6.5: This figure shows the ROC curves of the BNN system, BMRI expert without using the BNN assistance, and the BMRI expert with using the BNN system assistance for classifying the 80 BMRI lesions.

Table 6.3: The difference between area under ROC curve ( $A_z$ ) between the BMRI expert with and without the BNN assistance vs. the BNN system classification.

Contrast	Difference	$p$
BNN vs. expert w/o	0.153	0.0026
BNN vs. expert w/	0.037	0.3298
expert w/o vs. expert w/	-0.115	0.0084

The  $p$  value in Table 6.3 indicate a significant difference between the BNN system and the expert without assistance in classifying BMRI lesions ( $p < 0.05$ ), and a significant difference for the expert without assistance vs. the expert with assistance ( $p < 0.05$ ). But there was no significant difference between the BNN and the expert with assistance ( $p \approx 0.33$ ).

Figure 6.6 on the next page shows the ROC curves of the BNN system, the average mammographer performance without using the BNN assistance, and the average mammographer performance with the BNN system assistance for classifying the 80 BMRI lesions. Table 6.4 shows the differences between the average  $A_z$  for the two mammographers with and without the BNN assistance vs. the BNN system classification.

Table 6.4: The difference between area under ROC curve ( $A_z$ ) between the average mammographer with and without the BNN assistance vs. the BNN system classification.

Contrast	Difference	$p$
BNN vs. mammographers w/o	0.182	0.0013
BNN vs. mammographers w/	0.064	0.1251
mammographers w/o vs. mammographers w/	-0.117	0.0015



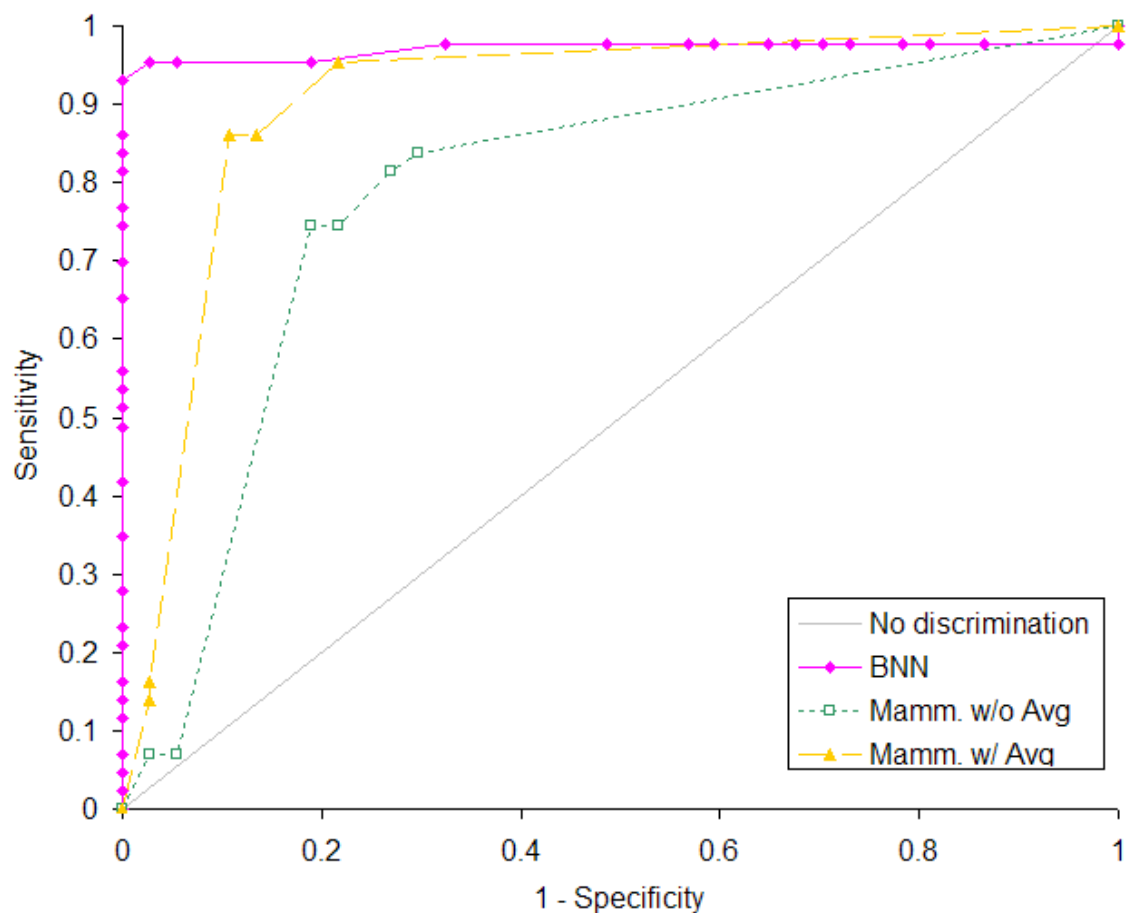


Figure 6.6: This figure shows the ROC curves of the BNN system, the average mammographer performance without using the BNN assistance, and the average mammographer performance with the BNN system assistance for classifying the 80 BMRI lesions.

The  $p$  values in Table 6.4 on page 130 indicates a significant difference between the BNN system and the mammographers without the assistance ( $p < 0.05$ ), and a significant difference for the mammographers without assistance vs. the mammographers with assistance ( $p < 0.05$ ). But there was no significant difference between the BNN and the mammographers with the assistance ( $p \approx 0.13$ ).

Figure 6.7 on the following page shows the ROC curves of the BNN system, the average body imaging fellow performance without using the BNN assistance, and the average body imaging fellow performance with the BNN system assistance for classifying the 80 BMRI lesions. Table 6.5 shows the differences between the average  $A_z$  for the two fellows with and without the BNN assistance vs. the BNN system classification.

Table 6.5: The difference between area under ROC curve ( $A_z$ ) between the average for the two fellows with and without the BNN assistance vs. the BNN system classification.

Contrast	Difference	$p$
BNN vs. fellows w/o	0.262	< 0.0001
BNN vs. fellows w/	0.101	0.0363
fellows w/o vs. fellows w/	-0.162	< 0.0001

The  $p$  values in Table 6.5 show a significant difference between the BNN system and the fellows without assistance ( $p < 0.05$ ), a significant difference for the fellows without assistance vs. the fellows with assistance ( $p < 0.05$ ), and a significant difference for the BNN and the fellows with the assistance ( $p < 0.05$ ).

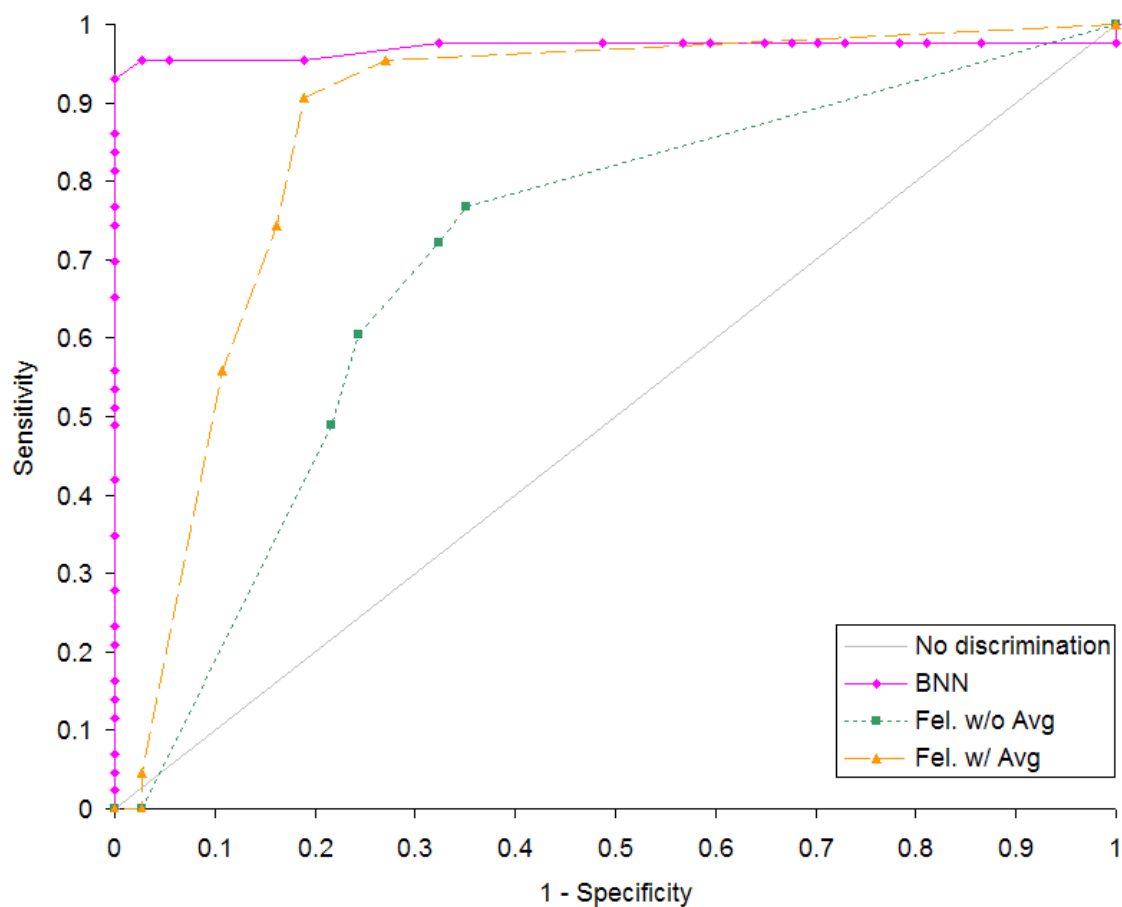


Figure 6.7: This figure shows the ROC curves of the BNN system, the average body imaging fellow performance without using the BNN assistance, and the average fellow performance with the BNN system assistance for classifying the 80 BMRI lesions.

This study showed a significant improvement in all readers when using the CAD assistance compared to no CAD assistance. For both the expert with CAD assistance and the mammographers with CAD assistance, there was no difference between the reader with CAD assistance and the BNN system at  $p = 0.05$  level. Even with CAD assistance, the body imaging fellows had significantly lower  $A_z$  than the other human reader groups and the BNN system.

In 1998, Giger *et al.* [31] performed a preliminary observer study conducted at the RSNA. They demonstrated that the average performance in terms of  $A_z$  value of 128 radiologists (who participated in the study) on 110 mammograms increased from an  $A_z = 0.89$  to an  $A_z = 0.94$  with computer assistance. We showed better improvement results for the three different categories of the human readers comparing with the study performed by Giger *et al.*, but the population of readers we used is not sufficient to generalize the outcomes of this study to the population of readers.

### 6.2.3 MRMC Analysis

In both types of MRMC analysis, there was a statistically significant improvement when readers were assisted by the CAD system. The second analysis showed a more significant improvement, but the first analysis is more defensible since the population of cases (patients) is well defined, but the populations of lesions is not. Since both are significant, we prefer to focus the discussion and conclusions on the more defensible position.

The MRMC results agree well with the individual ROC results, which is expected since each group significantly improved with CAD assistance, the readers as

a whole should also significantly improve.

### 6.3 Limitation of the Research

The limitations of this research can be grouped into two categories: Segmentation limitations and classification limitations.

The limitations on segmentation stem from using an expert to identify the lesions (semi-automatic segmentation) before computing the features. This makes the subsequent feature computation dependent on the expert and could be different with different experts. The difference is not expected to be large, but experts generally differ slightly on what they consider the extent of a lesion.

Another limitation is that all features were derived solely from the MR images. It is possible to use information such as personal and family medical history, age, presence of BRCA1/2 genes, *etc.* Radiologists frequently use these non-imaging features, and they could help improve the classification ability of the artificial classifiers.

The scope of the classification portion was also somewhat limited. We used 75 cases, but all acquired with the same scanner and under the same protocol. This prevents us from being definitive that these methods will work on all BMRI cases, but the quality of the results shows strong promise. We were also limited in the number of human subjects; to show strong statistical significance it is necessary to have at least 10 human readers, while 20 readers could be used in a very strong study to show how the general population of human readers would be affected by the CAD system.

Leave-one-out method does not demonstrate whether an algorithm is generalized for the population cases that are not seen by the training stage. Nonetheless, it

still uses a non-training sample for the testing stage. This method is useful when you have a limited number of samples, and it demonstrates the robustness of the training stage.

The classification results are restricted by the artificial classifiers limitations discussed in section 6.2.2.1 on page 124.

## 6.4 Summary

In this study, we have investigated different methods for automatic segmentation of BMRI lesions, and implemented a semi-automatic segmentation method. We have contributed validated algorithms for classifying BMRI lesions using artificial intelligence techniques. We also have developed a CAD system that uses these algorithms to assist readers with different levels of experience (experts, mammographers, and fellows) while they are reading breast MR cases. When they find lesions, the CAD system will improve their diagnosis, leading to fewer unnecessary biopsies, and more cancer found at an earlier, treatable stage. Finally, we have performed a thorough human reader experiment, showing that a CAD system can improve the performance of a human reader in breast MR classification.

We validated the algorithms using a data base of 75 BMRI images (80 lesions). Thirteen features were selected based on shape, texture, and enhancement kinetics of BMRI lesions. These features were fed to three artificial classifiers: a backpropagation neural network (BNN), support vector machine (SVM), and Bayesian classifier (BC). The three classifiers were trained using the leave-one-out method, and evaluated using ROC curve analysis. The BNN system significantly outperformed the

other two classifiers, thus it was used in the human reader experiments. Five radiologists participated in these experiments: one BMRI expert, two mammographers, and two body imaging fellows. The human readers' performance was compared with that of the BNN system, and evaluated using ROC curve and MRMC analysis with and without CAD assistance.

All readers showed statistically significant improvements in their ability to classify breast MRI lesions. Further, the BNN also significantly outperformed all readers without CAD assistance. With CAD assistance, there was no significant difference between the mammographers and the expert vs. the BNN. Though the fellows show a significant improvement with CAD assistance, their performance with CAD still did not match that of the mammographers and the expert.

At present, BMRI is best used as a diagnostic modality for high risk patients. Its high cost and large datasets make it difficult to be used as a screening method. As technology improves, cost generally decreases, and advanced CAD system can reduce the radiologists workload. Traditionally, BMRI has shown high sensitivity but limited specificity in finding cancer, Our CAD system has demonstrated high specificity for detected lesions. This enables BMRI to be a non-invasive alternative for handling suspicious lesions detected by mammography.

## CHAPTER 7 FUTURE WORK

In the future, this work can be extended in several directions. First, a fully automatic segmentation method can be further investigated. There are several possible ways of achieving automatic segmentation. One possibility is to use the enhancement kinetics, regional information, and a second neural network to classify each pixel into categories such as air, breast tissue, skin, blood vessels, and possibly enhancing masses. Connected components labeling should then be able to identify regions which need to be analyzed with the malignant/benign classifier described in this thesis. To overcome the problem of low specificity in MRI, the BNN could be augmented with a second output indicating the confidence that this lesion is important. Right now, a lesion that is not a cancerous mass (*e.g.* fibrosis), has to be classified as a benign mass, since there is no other category for it. This may confuse the network during training, because it has to reconcile two potentially very different classes. Adding the importance classification can simplify the segmentation stage, because extra regions from the segmentation stage could be properly classified as unimportant. Another possibility is to again use enhancement kinetics, only this time have them form a basis for thresholding. So instead of thresholding based on the difference between the pre-contrast and a specific post-contrast, thresholding could be performed looking at the entire enhancement profile.

More work can be done for tuning and validating the classifier. A larger collection of BMRI images can provide more confidence as to the systems performance



with a variety of lesion types. With a larger number of human readers, it would be possible to compare which categories of readers are helped the most by CAD assistance (experts, mammographers, fellows, residents). To become clinically useful, the system needs to be trained and validated on scanners from different manufacturers, using a wide set of protocols. This might also be a basis for creating optimized protocols for automatic segmentation and classification.

There are still more features which could be explored for the classification step. When diagnosing difficult cases, radiologists frequently use non-imaging information (*e.g.* personal and family medical history, age, presence of BRCA1/2 genes, *etc.*). These features can be easily incorporated into the system.

The CAD system could be adapted for use as a teaching tool. It can provide quick feedback for residents describing why this region is likely malignant or benign (high spiculation, large area, *etc.*). This type of training tool, with appropriate interaction, could boost the confidence of new readers as they interpret BMRI data.

## REFERENCES

- [1] M. Aizerman, E. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," in *Automation and Remote Control*, vol. 25, 1964, pp. 821–837.
- [2] American Cancer Society, "Cancer facts and figures 2004," 2004, [http://www.cancer.org/downloads/STT/CAFF\\_finalPWSecured.pdf](http://www.cancer.org/downloads/STT/CAFF_finalPWSecured.pdf).
- [3] L. Arbach, N. Al-Mudarres, and M. Ammar, "Computerized detection of breast masses from mammograms," in *Science Week*, vol. 37, Damascus, Syria, June 1997.
- [4] L. Arbach, D. L. Bennett, J. M. Reinhardt, and G. Fallouh, "Distinguishing between malignant and non-malignant breast masses from mammograms using AI techniques," *Basel Al-Assad Journal for Engineering Science*, no. 16, pp. 103–121, July 2002.
- [5] —, "Classification of mammographic masses: Comparison between back-propagation neural network BNN and human readers," in *SPIE in Medical Imaging*, vol. 5032, San Diego, CA, USA, Feb. 2003, pp. 810–818.
- [6] —, "Mammographic masses classification: Comparison between backpropagation neural network (BNN), K-nearest neighbors (KNN), and human readers," in *IEEE Canada, CCECE*, vol. 3, Montreal, Canada, May 2003, pp. 1441–1444.
- [7] L. Arbach and G. Fallouh, "Developing an expert system for distinguishing malignant and benign breast masses from mammograms using the fuzzy logic," *Basel Al-Assad Journal for Engineering Science*, Feb. 2002.
- [8] L. Arbach, A. Stolpen, and J. M. Reinhardt, "Classification of breast MRI lesions using a backpropagation neural network (BNN)," in *IEEE'04 International Symposium on Biomedical Imaging*, Washington, DC, 2004, pp. 253–256.
- [9] D. Aykac, E. A. Hoffman, G. McLennan, and J. M. Reinhardt, "Segmentation and analysis of the human airway tree from three-dimensional x-ray CT images," *IEEE Trans. On Medical Imaging*, vol. 22, no. 8, pp. 940–950, Aug. 2003.
- [10] D. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981.

- [11] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, "Learnability and the Vapnik-Chervonenkis dimension," *Journal of the ACM*, vol. 36, no. 4, pp. 929–865, 1989.
- [12] K. W. Bowyer, "Validation of medical image analysis techniques," in *Handbook of Medical Imaging*, M. Sonka and J. M. Fitzpatrick, Eds. SPIE Press, 2000.
- [13] L. Bruce and N. Shanmugam, "Using neural network with wavelet transforms for an automated mammographic mass classifier," in *SPIEMI*, 2000, pp. 985–987.
- [14] F. A. Cardillo, A. Starita, D. Caramella, and A. Cilo, "A neural tool for breast cancer detection and classification," in *EMBS*, Istanbul, Turkey, 25–28 Oct. 2001, pp. 2733–2736.
- [15] —, "A tool for breast MR images processing and classification," *IEEE Transactions on Medical Imaging*, pp. 181–184, 2002.
- [16] R. Carraway, T. Morin, and H. Moskowitz, "Generalized dynamic programming for multicriteria optimization," *European Journal of Operational Research*, vol. 44, pp. 95–104, 1990.
- [17] H. Chan and B. Sahiner, "Computerized analysis of mammographic microcalcifications in morphological and texture feature space," *Medical Physics*, vol. 25, no. 10, pp. 2007–2019, 1998.
- [18] C. Chen and G. Lee, "Image segmentation using multiresolution wavelet analysis and expectation-maximization (EM) algorithm for digital mammography," in *Imaging Systems and Technology*, vol. 8, no. 5, 1997, pp. 491–504.
- [19] H. Cheng, C. Chen, and R. Freimanis, "A neural network for breast cancer detection using fuzzy entropy approach," vol. 3, Washington, DC USA, 23–26 Oct. 1995, pp. 141–144.
- [20] I. Christoyianni, E. Dermatas, and G. Kokkinakis, "Neural classification of abnormal tissues in digital mammography using statistical features of the texture," in *IEEE ICECS*, York, England, July 1999, pp. 117–120.
- [21] —, "Fast detection of masses in computer-aided mammography," in *IEEE Signal Processing Magazine*, Jan 2000, pp. 54–64.

- [22] E. Claridge and J. Richer, "Characterization of mammographic lesions," in *2nd International Workshop on Digital Mammography*, York, England, July 1994, pp. 241–250.
- [23] H. E. Cline, D. R. Thedens, C. H. Meyer, D. G. Nishimura, T. K. Foo, and S. Ludke, "Combined connectivity and a gray-level morphological filter in magnetic resonance coronary angiography," *Magnetic Resonance in Medicine*, no. 43, pp. 892–895, 2000.
- [24] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines (and other kernel-based learning methods)*. Cambridge University Press, 2000.
- [25] B. Dhawan and P. Alam, "Artificial-neural-network-based classification of mammographic microcalcifications using image structure features," vol. 1905, pp. 820–831, July 1993.
- [26] D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method," *Investigative Radiology*, vol. 27, no. 9, pp. 723–731, Sept. 1992.
- [27] R. Duda and P. Hart, "Use of the Hough transform to detect lines and curves in pictures," *Communs Assn Comput*, vol. 15, pp. 11–15, 1975.
- [28] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. Wiley & Sons, 2000.
- [29] I. El-Naqa, Y. Yang, M. Wernick, and R. Nishikawa, "A support vector machine approach for detection of microcalcification," vol. 21, no. 12, pp. 1552–1563, Dec. 2002.
- [30] C. Enderwick and E. Micheli-Tzanakou, "Classification of mammographic tissue using shape and texture features," *IEEE/EMBS*, vol. 19, p. 810, 1997.
- [31] M. L. Giger and Z. Huo, "Artificial neural networks in breast cancer diagnosis: Merging of computer-extracted features from breast images," *IEEE Trans. on Medical Imaging*, vol. 17, no. 10, pp. 1768–1769, 1999.
- [32] M. Giger, "Current issues in for mammography," in *Digital Mammography*, vol. 2, 1996, pp. 53–59.
- [33] M. Giger, Z. Huo, and K. Horsch, "Computer-aided diagnosis of lesions on multimodality images of the breast," vol. 4322, 2001, pp. 656–660.

- [34] T. O. Gulsrud and J. Husoy, "Optimal filter-based detection of microcalcifications," *IEEE Trans. on Biomedical Engineering*, vol. 48, no. 11, pp. 1272–1281, Nov 2001.
- [35] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design*. Boston: PWS Publishing, 1996.
- [36] C. Hayes and T. Redpath, "Breast MRI protocols," in *Breast MRI in practice*, R. Warren and A. Coulthard, Eds. UK: Martin Dunitz, 2002.
- [37] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York, NY: IEEE Press, 1994.
- [38] S. H. Heywang, P. Viehweg, and A. Heining, "Contrast enhanced MRI of breast: Accuracy, value, solutions," *European Journal of Radiology*, vol. 24, no. 2, pp. 94–108, 1997.
- [39] S. H. Heywang and H. Wolf, "Misleading changes of the signal intensity on opposed-phase MRI after injection of contrast medium," *Journal of computer assistant tomography*, vol. 20, pp. 173–178, 1996.
- [40] S. L. Hillis and K. S. Berbaum, "Power estimation for the Dorfman-Berbaum-Metz method," *Academic Radiology*, vol. 11, pp. 1260–1273, 2004.
- [41] S. L. Hillis, N. A. Obuchowski, K. M. Scharz, and K. S. Berbaum, "A comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette methods for receiver operating characteristic (ROC) data," *Statistics in Medicine*, 2005, in press.
- [42] S. A. Hojjatoleslami and J. Kittler, "Region growing: A new approach," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1079–1084, July 1998.
- [43] J. P. Hornak, "The basics of MRI," 2002, <http://www.cis.rit.edu/htbooks/mri/inside.htm>.
- [44] C. A. Hulka, W. B. Edmister, and B. L. Smith, "Dynamic echo planner imaging of the breast: experience in diagnosing breast carcinoma and correlation with tumor angiogenesis," *Radiology*, vol. 197, pp. 33–38, 1997.
- [45] Z. Huo, M. L. Giger, C. Vyborny, D. Wolverton, R. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign mass lesions on digitized mammograms," *Academic Radiology*, vol. 5, pp. 155–168, 1998.

- [46] Imaginis, “Breast cancer: Statistics on incidence survival, and screening,” 2002, <http://www.imaginis.com/breasthealth/statistics.asp>.
- [47] —, “Breast cancer: Statistics on incidence survival, and screening,” 2004, <http://imaginis.com/breasthealth/mri.asp>.
- [48] N. Karssemcijer, “Adaptive noise equalization and recognition of microcalcification clusters in mammograms,” *Pattern Recognition and Artificial Intelligence*, vol. 7, pp. 1357–1376, 1993.
- [49] R. Kasturi and R. C. Jain, *Computer Vision: Principles and Applications*. Los Alamitos, CA: IEEE Computer Society Press, 1991.
- [50] C. Kocur, S. Rogers, K. Bauer, and J. Steppe, “Investigation of regularized neural networks for the computerized detection of mass lesions in digital mammography,” in *IEEE/EMBS*, vol. 19, 1997, pp. 1336–1339.
- [51] B. Kosko, *Neural Networks and Fuzzy Systems*. Prentice Hall Inc., 1992.
- [52] M. Kriege, C. Brekelmans, and C. Boetes, “Efficiency of MRI and mammography for breast cancer screening in women with familial or genetic predisposition,” *The New England Journal of Medicine*, vol. 351, pp. 427–437, 2004.
- [53] C. Kuhl, R. Schmutzler, and C. Leutner, “Breast MR imaging screening in 129 women proved or suspected to be carriers of breast cancer susceptibility gene: Preliminary results.” *Radiology*, vol. 215, pp. 267–279, 2000.
- [54] M. Kupinski and M. L. Giger, “Neural network feature selection for breast cancer diagnosis,” in *SPIEMI*, vol. 2492, 1995, pp. 905–918.
- [55] W. Land, T. Masters, and J. Lo, “Application of adaptive boosting to EP-derived multilayer feed-forward neural networks (MLFN) to improve benign/malignant breast cancer classification,” July 2001, vol. 4322, pp. 1717–1724.
- [56] F. Lefebvre and H. Benali, “Fractal analysis of clustered microcalcification in mammograms,” in *Acta. Stero*, vol. 11, 1992, pp. 611–616.
- [57] G. Lera and M. Pinzolas, “Neighborhood based Levenberg-Marquardt algorithm for neural network training,” *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1200–1203, 2002.

- [58] H. Li, R. Liu, and S. Lo, "Fractal modeling and segmentation for the enhancement of microcalcification in digital mammograms," in *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, Dec. 1997, pp. 785–798.
- [59] L. Li, R. Clark, and J. Thomas, "Computer-aided diagnosis of masses with full-field digital mammography," *Academic Radiology*, vol. 9, pp. 4–12, 2002.
- [60] L. Li, Y. Chu, A. F. Salem, and R. A. Clark, "Image segmentation and 3D visualization for MRI mammography," vol. 4684, 5 2002, pp. 1780–1789.
- [61] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. New York: IEEE press, 2000.
- [62] L. Liberman, "Breast cancer screening with MRI-what are the data for patients at high risk?" *The New England Journal of Medicine*, vol. 351, no. 5, pp. 497–500, 2004.
- [63] L. Lihua and R. Clark, "Computer-aided diagnosis of masses with full-field digital mammography," *Academic Radiology*, vol. 9, pp. 4–12, 2002.
- [64] S. Liu, C. F. Babbs, and E. J. Delp, "Multiresolution detection of spiculated lesions in digital mammograms," in *IEEE Conference on Image Processing*, Rochester, MN, June 2001, pp. 874–884.
- [65] Magnetic Resonance Science Center, "Breast MRI web site," 2002, <http://www.mrsc.ucsf.edu/breast/>.
- [66] J. E. Martin, "Breast cancer missed by mammography," *American Roentgen Ray Society*, vol. 132, pp. 737–739, May 1979.
- [67] J. L. McClelland and D. E. Rumelhart, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge: MIT Press, 1986, vol. 1: Foundations.
- [68] S. C. Mitchell, B. P. Lelieveldt, H. G. Bosch, J. H. Reiber, and M. Sonka, "Disease characterization of active appearance model coefficients," in *SPIEMI*, 2003, pp. 949–957.
- [69] N. R. Mudigonda and N. Rangayyan, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Transactions on Medical Imaging*, pp. 1215–1227, Dec 2001.

- [70] M. Walsh and M. Lee, "A review of false negative mammography in symptomatic population," *Clinical Radiology*, vol. 44, pp. 13–15, 1991.
- [71] National Breast Cancer Foundation, "Cancer incidence in the U.S." 2004, <http://www.nationalbreastcancer.org/>.
- [72] L. Nunes and M. D. Schnall, "Diagnostic performance characteristics of architectural features revealed by high spatial-resolution MR imaging of the breast," *AJR*, vol. 169, pp. 409–415, Aug. 1997.
- [73] L. Nunes, M. D. Schnall, and S. G. Orel, "Update of breast MR imaging architectural interpretation model," *Radiology*, pp. 484–494, May 2001.
- [74] A. R. Padhani, "Contrast agent dynamics in breast MRI," in *Breast MRI in practice*, R. Warren and A. Coulthard, Eds. UK: Martin Dunitz, 2002.
- [75] A. I. Penn, N. Umar, S. F. Thompson, M. D. Schnall, and F. Wang, "Preliminary performance analysis of breast MRI CAD system," in *SPIE on Medical Imaging*, vol. 4322, 7 2001, pp. 1944–1953.
- [76] R. Pfisterer and F. Aghdasi, "Tumor detection in digitized mammograms by image texture analysis," vol. 40, no. 2, pp. 209–216, Feb. 2001.
- [77] F. Podo, F. Sardanelli, and R. Canese, "The Italian multi-center project on evaluation of MRI and other imaging modalities in early detection of breast cancer in subjects at high risk." *Journal of Clinical Cancer Research*, vol. 21, pp. 115–124, 2002.
- [78] R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," vol. 4322, Jul 2001, pp. 1337–1346.
- [79] W. Polakowski and D. Cournoyer, "Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivation-based feature saliency," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 811–819, Dec. 1997.
- [80] M. Ponti and A. Verri, "Support vector machines for 3-D object recognition," *IEEEtrans. Pattern Analysis and Machine Intelligence*, vol. 20, pp. 637–646, June 1998.
- [81] W. Qian and I. Li, "Image feature extraction for mass detection in digital mammography," *Medical Physics*, vol. 26, no. 3, pp. 402–408, 1999.



- [82] M. Quenouille, "Approximate tests of correlation in time series," *J Roy Stat Soc Ser B*, vol. 11, pp. 68–84, 1949.
- [83] —, "Notes on bias in estimation," *Biometrika*, vol. 43, pp. 353–360, 1956.
- [84] T. W. Rauber, "TOOLDIAG: Pattern recognition toolbox," <http://www.inf.ufes.br/thomas/home/tooldiag.html>.
- [85] B. D. Ripley, *Pattern Recognition and Neural Networks*. New York: Cambridge University Press, 1996.
- [86] W. G. Scheiber, G. Brix, and M. Knopp, "Improved visualization of breast lesions with gadolinium-enhanced magnetization transfer MR imaging," *Magnetic Resonance in Medicine*, vol. 35, pp. 861–869, 1996.
- [87] B. L. Shih-Chung, H. Li, Y. Wang, and M. Freedman, "A multiple circular path convolution neural network system for detection of mammographic masses," *IEEE Jour. on Medical Imaging*, vol. 21, no. 2, pp. 150–158, Feb 2002.
- [88] M. Sniedovich, "Dynamic programming and principles of optimality," *Journal of Mathematical Analysis and Applications*, vol. 65, no. 3, pp. 586–606, 1978.
- [89] O. Stan and E. Kamen, "New block recursive MLP training algorithms using the Levenberg-Marquardt algorithm," in *Neural Networks*, vol. 3, 10-16 July 1999, pp. 1672–1677.
- [90] A. Starita, S. LaManna, D. Majidi, D. Caramella, and A. Cilotti, "Magnetic resonance in mammography: A quantitative evaluation tool of contrast-enhanced magnetic resonance of the breast," in *(IEEE) Conference of Advances in Medical Signal and Information Processing*, 4–6 Sep. 2000, pp. 200–205.
- [91] The UK MRI breast screening study advisory, "Protocol for a national multi-center study of magnetic resonance imaging (MRI) screening in women at genetic risk of breast cancer," *Breast*, vol. 9, pp. 78–82, 2000.
- [92] J. Tukey, "Bias and confidence in not quite large samples," *Annals of Math Statistics*, vol. 29, pp. 614–618, 1958.
- [93] T. Vapnik and A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability and its Applications*, vol. 16, no. 2, pp. 264–280, 1971.

- [94] B. Verma and J. Zakos, "A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques," *Information Technology in Biomedicine, IEEE*, vol. 5, pp. 46–54, March 2001.
- [95] C. Vyborny and M. Giger, "Computer vision and intelligence in mammography," *AJR*, vol. 162, pp. 699–708, 1994.
- [96] C. J. Vyborny, "Can computers help radiologists read mammograms," *Radiology*, vol. 191, pp. 315–317, 1994.
- [97] T. Wang and N. Karayiannis, "Detection of microcalcifications in digital mammograms using wavelets," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 498–509, Aug. 1998.
- [98] E. Warner, D. Plewes, and R. Shumak, "Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer," *Journal of Clinical Oncology*, vol. 19, pp. 3524–3531, 2001.
- [99] P. Weatherall, "MRI vs. histologic measurement of breast cancer following chemotherapy: Comparison with X-ray mammography and palpation," *Magnetic Resonance Imaging*, no. 13, pp. 868–875, 2001.
- [100] C. Wu, "Image feature analysis for classification of microcalcifications in digital mammography: Neural network and genetic algorithms," vol. 3034, pp. 499–507, April 1997.
- [101] Y. Wu, M. Giger, K. Doi, C. Metz, and R. Nishikawa, "Application of artificial neural networks in mammography for diagnosis of breast cancer," vol. 1778, pp. 19–27, 1992.
- [102] Z. R. Yang and R. G. Harrison, "Detecting false benign in breast cancer diagnosis," *Neural Networks, IEEE*, vol. 3, pp. 655–658, July 2000.
- [103] S. Yu, L. Guan, and S. Brown, "Automatic detection of clustered microcalcifications in digital mammograms based on wavelet and neural network classification," vol. 3338, June 1998, pp. 1540–1546.