Loughborough University

University Library

Author/Filing Title ..... YAP, M.H. .....

.....

Class Mark ..... T

Please note that fines are charged on ALL overdue items.

Enhanced Algorithms for Lesion Detection & Recognition in Ultrasound Breast Images

By

Moi Hoon Yap

A Doctoral Thesis

submitted in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

Department of Computer Science

Loughborough University

December 2008

© by Moi Hoon Yap 2008

Supervisors: Dr Eran Edirisinghe, Dr Helmut Bez

Director of research: Professor Alastair Gale

Loughborough University
Pilkington Library
Date 14/5/10
Class T
Acc No. 0403820097X

Abstract

Mammography is the gold standard for breast cancer detection. However, it has very high false positive rates and is based on ionizing radiation. This has lead to interest in using multi-modal approaches. One modality is diagnostic ultrasound, which is based on non-ionizing radiation and picks up many of the cancers that are generally missed by mammography. However, the presence of speckle noise in ultrasound images has a negative effect on image interpretation. Noise reduction, inconsistencies in capture and segmentation of lesions still remain challenging open research problems in ultrasound images.

The target of the proposed research is to enhance the state-of-art computer vision algorithms used in ultrasound imaging and to investigate the role of computer processed images in human diagnostic performance. The results of the research can be used to improve existing computer vision algorithms in US CAD systems and propose methods that can enhance the perceptual tasks in US training systems.

The thesis proposes an image pre-processing stage that uses hybrid filtering and multifractal analysis which reduces speckle noise in ultrasound images. Further it proposes two novel initial lesion detection techniques one of which is superior to other techniques, improved segmentation algorithms for lesion boundary identification, best feature descriptors that can be used to enhance the chances of better recognition and feature-classifier combinations that perform better than other systems in lesion type recognition. Their performances were better compared with that of the benchmark algorithms and suggestions are made for possible future enhancements. *Moi-Hoon Yap, Nov 2008*

Acknowledgements

I would like to express my appreciation to my supervisors, Dr Eran Edirisinghe and Dr Helmut Bez, for their excellent guidance and continuous support. Dr Edirisinghe provided me the opportunity to study overseas, acquire research skills and precious suggestions in writing papers and this thesis. Dr Bez helped me with the understanding of mathematical concepts and theories.

Many thanks to my director of research, Professor Alastair Gale, my supervisor in my Research Associate post. He has provided me ample opportunities to be professionally trained in medical imaging and research methodology. I would like to thank Professor Paul Chung for his advice during my first year. I would also like to thank my colleagues of the Digital Imaging Research group, the Applied Vision Research Centre and the Technical Staff members of the Department of Computer Science for their help, corporation and friendship, which proved to be invaluable.

Special thanks go to my parents and siblings, who have been supporting me for many years. I know, they will continue to support me in every possible way. I am profoundly thankful to my husband, for his love, faith, valuable advice and understanding during my study.

I would like to acknowledge the support of Dr Prapavesis et al. for allowing me to use their ultrasound image database in my research. Further I would like to pay my gratitude to the Department of Computer Science for its contribution towards funding my tuition studies.

TABLE OF CONTENTS

Abstract II

Acknowledgements III

Table of contents IV

List of Figures VIII

List of Tables XIII

Abbreviations XIV

Glossary XVI

PART 1: Introduction, Background and Literature Review

CHAPTER 1 INTRODUCTION 1

1.1 Research Motivation 2
1.2 Aims and Objectives 4
1.3 Proposed Approach 6
1.4 Contributions 8
1.5 Thesis Organisation 9

CHAPTER 2 BACKGROUND 13

2.1 Anatomy of Breasts 13
2.2 The UK Breast Screening Programme 14
2.3 Diagnostic Breast Ultrasound 16
2.4 Speckle Noise in US Imaging 20
2.5 Benign Tumours 21
2.5.1 Fibroadenoma 21
2.5.2 Cysts 22
2.6 Malignant Tumours 23
2.7 Review of Breast Cancer Detection Techniques 25
2.7.1 Screen Film Mammography (SFM) 26
2.7.2 Full-Field Digital Mammography (FFDM) 27
2.7.3 Ultrasound (US) 28
2.7.4 Magnetic Resonance Imaging (MRI) 29
2.7.5 Positron Emission Tomography (PET) 29
2.7.6 Terahertz Rays (T-rays) 30

2.8 Summary 31

CHAPTER 3 LITERATURE REVIEW 32

3.1 Introduction 32
3.2 Ultrasound 33
3.2.1 Initial Lesion Detection 37
3.2.2 Segmentation 38
3.2.3 Classification 41
3.2.3.1 Breast Imaging Reporting and Data System (BIRADS) 41
3.2.3.2 Computer Methods in Classification 42
3.2.4 Test Performance Measures 44
3.3 Survey in Ultrasound Breast Images Research Groups 47
3.4 Mammography 52
3.4.1 Mammography CAD System 52
3.4.2 Current Trend in Mammography Research 54
3.5 Summary & Conclusion 56

PART 2: Novel Algorithms for Breast Cancer Detection and Recognition (Contributory Chapters)

CHAPTER 4 INITIAL LESIONS DETECTION 59

4.1 Introduction 59
4.2 Motivation and Benchmark Algorithm 60
4.3 Measurement 61
4.4 Benchmark algorithm 62
4.5 Datasets 63
4.6 Pre-processing 64
4.6.1 Histogram equalisation 64
4.6.2 Hybrid filtering 67
4.7 Method-1 69
4.7.1 Watershed segmentation 70
4.7.2 Segment of interest 74
4.7.2.1 Statistical analysis 74
4.7.2.2 Fractal analysis 74
4.7.3 *k*-means clustering 76
4.7.4 Results and analysis 77
4.8 Method-2 83

4.8.1 Multifractal processing 84
4.8.2 Thresholding segmentation 88
4.8.3 Rule-based approach 89
4.8.4 Result and analysis 90
4.9 Comparison of the Proposed Methods against Benchmark Algorithm 93
4.10 Summary and Conclusion 96

CHAPTER 5 BOUNDARY DETECTION 98

5.1 Introduction 98
5.2 Method-1: Multifractal RGI approach 98
5.2.1 Research motivation and benchmark algorithms 99
5.2.2 The methodology 103
5.2.2.1 Region growing segmentation 103
5.2.3 Results and analysis 104
5.3 Method-2: Multifractal IsoGauss Approach 107
5.3.1 Research motivation and benchmark algorithms 107
5.3.2 The Methodology 108
5.3.2.1 Ultrasound images 109
5.3.2.2 Initial lesion detection 109
5.3.2.3 Initial contour detection 109
5.3.2.4 Active contour models 111
5.3.3 Results & Analysis 113
5.4 Summary & Conclusion 117

CHAPTER 6 CLASSIFICATION 119

6.1 Introduction 119
6.2 Appearance Based Classification 121
6.2.1 Principal Component Analysis (PCA) 122
6.2.2 Two Dimensional Linear Discriminant Analysis (2D-LDA) 125
6.3 Feature Based Classification 126
6.3.1 Feature Extraction Approaches 126
6.3.1.1 Shape descriptors 126
6.3.1.2 Texture descriptors 130
6.3.1.3 Edge descriptors 132
6.3.2 Classifiers 135
6.3.3 Feature Selection 138
6.4 Experimental Results & Analysis 139

6.4.1 Performance evaluation of the SVM classifier 141
6.4.2 Performance evaluation of other classifiers 144
6.4.3 Comparison of feature selection methods 150
6.5 Summary & Conclusion 158

CHAPTER 7 EFFECTS OF COMPUTER PROCESSED IMAGES ON HUMAN PERFORMANCE 160
7.1 Introduction 160
7.2 Research Motivation 161
7.3 Methodology 162
7.3.1 Experimental design and rationale 163
7.3.2 Test image dataset 164
7.3.3 Processed images 166
7.3.4 Subjects 168
7.3.5 Design of the questionnaire 168
7.3.6 Receiver Operating Characteristics (ROC) analysis 170
7.4 Results and Analysis 171
7.5 Summary & Conclusion 176

CHAPTER 8 SUMMARY AND FUTURE WORK 177
8.1 Summary 177
8.2 Future Work 180
8.3 Final Conclusions 182

References 184

Appendices
Appendix A Sample Questionnaires 197
Appendix B List of Publications 213
Appendix C Journal papers 215

List of Figures

Figure 1-1 An existing US CAD approach.....5
Figure 1-2 The aim of this research. (a) to automate the ROI selection, (b) to improve the segmentation and classification algorithms, (c) initial study in human performance by using the intermediate images.....5
Figure 1-3 Thesis Organisation.....12
Figure 2-1 Anatomy of breasts.....13
Figure 2-2 Frequency range of diagnostic ultrasound.....16
Figure 2-3 The radiologist holds the ultrasound transducer against the skin while imaging the breast.....19
Figure 2-4 Sonography layer anatomy.....19
Figure 2-5 Benign tumours – fibroadenoma.....22
Figure 2-6 Benign tumours – Cysts.....23
Figure 2-7 Malignant tumours.....24
Figure 2-8 The electromagnetic spectrum.....26
Figure 3-1 Diagram illustrates the flow of CAD processes.....32
Figure 3-2 Speckle noise causing difficulties in detecting the edges in ultrasound images.....34
Figure 3-3 Comparison between speckle reduction filters.....36
Figure 3-4 The trade off of the cut off value.....46
Figure 3-5 Examples of ROC curves.....47
Figure 4-1 Illustration of the use of the proposed solution.....59
Figure 4-2 Illustration of the operation of the intermediate stages of Drukker et al.'s algorithm (a) original image, (b) gray-scale inverted image, (c) median filtered image, (d) RGI filtered image, (e) thresholded RGI image, (f) final detection.....62
Figure 4-3. (a) original image (b) Gaussian blur (c) Nonlinear diffusion filtering. (d) hybrid filtering.....68
Figure 4-4 Overview of the methodology.....69
Figure 4-5 Illustration of over segmentation (a) original image (b) Over segmented Image.....70
Figure 4-6 Some speckle noise and artefacts are removed after applying Gaussian Blur, with (a) low scale of blur, σ=2; and (b) high scale of blur, σ=8. Note that image (b) consists of less noise, and is smoother than image (a).....71
Figure 4-7. Gaussian blur reduces the number of segments. Image in (a) shows the result of watershed segmentation on an image, obtained with a low scale of blur, σ=2, overlapped with the original image. Image in (b) shows the result of watershed segmentation on an image, obtained with a high scale of blur, σ=8, overlapped with

the original image.....71
Figure 4-8 Application of watershed segmentation to different types of lesions. Observe that there are differences between segmentation patterns for different types of lesions. For 'normal' lesions, the distribution of the segments are generally even, as illustrated in (a) and (b). For others types of lesions, the suspected regions consist of less segments. (c) and (d) are the images with cysts, where no segmentation has occurred within the ROI (areas pointed by the arrows). For images with fibroadenoma ((e) and (f)) and malignant ((g) and (h)) lesions, the number of segments in the suspected region is less than (a) in an image with a normal lesion (area pointed by the arrows).....72
Figure 4-9 (a) Markers. (b) Results of using Marker-controlled watershed segmentation.....73
Figure 4-10 (a) original image. After; (b) histogram equalization, (c) nonlinear diffusion filtering, (d) watershed segmentation, (e) mapping of watershed segmented image with the original image, (f) labelling of segments, (g) location of the segment of interest, (h) automatic labelling of the region of interest using a rectangle.....78
Figure 4-11 (a) the plot of minimum local mean, showing that segment number 50 is the point of interest. (b) the plot of minimum D_m, confirming that segment number 50 is the point of interest.....79
Figure 4-12 Illustration of the operation of the intermediate stages of the proposed method (a) Original image (b) Image after pre-processing (histogram equalization) (c) Image after hybrid filtering (d) Watershed segmentation mapped on to the original image (e) Segments with the same cluster with the initial lesion (f) Labelling of region of interest (ROI) by using the proposed method.....81
Figure 4-13 (a), (c), (e), (g),(i) Manual labelling of ROI. (b), (d), (f), (h),(j) The results of automated labelling by using our method.....82
Figure 4-14 (a), (c) Manual labelling of ROI. (b), (d) The results of automated labelling of the proposed method.....83
Figure 4-15 Overview of Second Proposed Method.....84
Figure 4-16 Graph plotted for multifractals dimension with different value of q.....86
Figure 4-17 Result of multifractal analysis with (a) $q<0$ or $q>1$, (b) $0<q<1$, (c) $q=1$, (d) threshold of (a), (e) threshold of (b), (f) threshold of (c).....87
Figure 4-18 Comparison of segmentation approaches (a) original image, when using only (b) Gaussian filtering + Multifractal (c) Non-linear + Multifractal filtering (d) Hybrid + Multi-fractal filtering.....90

Figure 4-19 Illustration of the operation of the intermediate stages of the proposed algorithm (a) The original image. Image after, (b) pre-processing (histogram equalization) (c) hybrid filtering (d) multifractal processing (e) thresholding segmentation (f) and labelling of region of interest (ROI).....91
Figure 4-20 The results of automated lesion ROI labelling when using the proposed method, (a) malignant tumour (b) simple cyst (c) fibroadenoma (d) complex cyst....92
Figure 4-21 Examples of unsuccessful lesion identification for two cases of fibroadenoma.....92
Figure 4-22 Graphical presentation of the ROI detection accuracy for each type of abnormality of Drukker's algorithm, method-1 (with the use of local mean and fractal dimension) and method-2.....94
Figure 4-23 Visual performance comparison of method-2 against that of method-1 and the benchmark algorithm. (a),(f) are the original images. (b),(g) are the results of Drukker et al's algorithm. (c),(h) are the results of method-1 (using local mean). (d),(i) are the results of method-1 (using the fractal dimension). (e),(j) are the results of the proposed method.....95
Figure 5-1 RGI based algorithm of Kupinski et al. (a) original image (b) inverted image preprocessed by median filtering. (c) detected boundary. (d) the detected boundary mapped on the original image.....101
Figure 5-2 Boundary detection algorithm of Joo et al. (a) original image. After, (b) median filtering, (c) unsharp masking, (d) contrast enhancement, (e) binary thresholding, (f) median filtering and (g) the detected boundary (h) the boundary mapped on to the original image.....102
Figure 5-3 Block diagram depicting stages of Approach-1.....103
Figure 5-4 The RGI value of each partition. Note that the partitioning is based on the threshold value used. The partition with the largest RGI value is returned as the final lesion partition.....104
Figure 5-5 Results of the proposed method. (a) The original image. The processed image after, (b) histogram equalization, (c) hybrid filtering, (d) multifractal dimension/analysis, (e) the boundary detected by using the region growing method (Note: the boundary with highest RGI is selected.), (f) the detected boundary mapped on to the original image.....105
Figure 5-6 Detected boundary when multifractal analysis is excluded.....105
Figure 5-7 Comparison between the three boundary detection methods. (a),(e),(i),(m),(q) are the original images. Results produced by; (b),(f),(j),(n),(r) Kupinski et al's method. (c),(g),(k),(o),(s) Joe et al's method. And (d),(h),(l),(p),(t) the proposed method.....106

Figure 5-8 Overview of methodology.....109
Figure 5-9 Output at various intermediate stages of stage 2: (a) Inverted image, (b) after applying Isotropic Gaussian on the inverted image, (c) after thresholding segmentation and (d) the final lesion boundary.....111
Figure 5-10 (a),(d) Original image (b),(e) Initial boundary (c),(f) Final boundary obtained by the gradient vector flow based approach.....113
Figure 5-11 Output at various stages: (a) Original image. After; (b) histogram equalization, hybrid filtering and multifractal enhancement, (c) thresholding segmentation, (d) Labelling of region of interest (ROI), (e) Isotropic Gaussian on inverted US images, (f) thresholding segmentation (g), initial boundary and (h) the final detected boundary.....114
Figure 5-12 Visual performance comparison with benchmark algorithms. (a),(d),(g) The original images. (b),(e),(h) The result of Drukker et al's algorithm. (c),(f),(i) The result of the proposed Multifractal IsoGauss approach 115
Figure 6-1 Resized test images (a-d) and its correspondence Eigen images(e-h)...121
Figure 6-2(i) Successful classifications in PCA classifier.....123
Figure 6-2(ii) mis-classifications in PCA Classifier.....124
Figure 6-3 Lesions and Convex hull. (a),(b) malignant tumours, (c),(d) benign tumours.....127
Figure 6-4 Illustration of the bounding rectangle with it's width and length.....128
Figure 6-5 Illustration of the major axis and minor axis.....129
Figure 6-6 Illustration of the shape of a: (a) benign tumour, (b) malignant tumour, with the respective roundness and form factor values.....129
Figure 6-7 Illustration of maximum radius and minimum radius.....130
Figure 6-8 Texture within contours of the detected lesions. (a) Cysts, (b) Fibroadenoma, (c) malignant.....130
Figure 6-9 (a) Single point. (b) multiple points.....132
Figure 6-10 Fourier descriptors for a (a) malignant tumour and a (b) benign tumour, and their corresponding reconstructions of edges using different numbers of descriptors/coefficients.....134
Figure 6-11 Analysis of single different features in classification. The blue and red areas represent malignant and benign classes, respectively. the x-axes represents the value of each parameter represented in feature extraction.....142
Figure 6-12 Fitted ROC curve shows the performance of ten classifiers with 19 features.....144
Figure 6-13 Fitted ROC curve illustrates the performance of different classifiers with features selected by CfsSubsetEval.....153

Figure 7-1. (a,b), (c,d), (e,f), (g,h), (i,j) are five pairs of original ultrasound images selected for the survey. (a, c, e, g, i) are selected randomly, while (b, d, f, h, j) are purposely selected to be of same level of diagnostic difficulty.....165
Figure 7-2. (a) Original image (b), (c), (d), (e), (f) Processed images by using different algorithms: (b) isogauss processing, (c) histogram equalization (d) Multifractal filtering (e) Non-linear filtering, (f) median filtering.....167
Figure 7-3. Comparison of the detection accuracy between subjects from different professional backgrounds.....172
Figure 7-4. Comparison of ROC curves (non-Radiologists).....174
Figure 7-5. Comparison of ROC Curves (Radiologists).....174

List of Tables

Table 2-1 Velocity figures of different material.....18
Table 3-1 Test characteristics.....45
Table 3-2 Summary of state-of-the-art mammographic CAD systems.....53
Table 4-1 The probability of an occurrence of grey level 65
Table 4-2 The cumulative distribution function (*cdf*).....66
Table 4-3 Initial lesion detection accuracy when using local mean and D_m, as metrics 80
Table 4-4 Summary of the ROI detection accuracy for each type of abnormality.....93
Table 5-1 Novelty of the multifractal IsoGauss approach.....116
Table 6-1 Summary statistics in using single feature in using SVM based classification.....143
Table 6-2 Summary statistics in using different groups of features in SVM based classification.....144
Table 6-3 Summary statistics for different classifiers implemented within WEKA when using all of the 19 features.....145
Table 6-4 Summary statistics for different classifiers when using a combination of all shape features.....147
Table 6-5 Summary statistics for different classifiers when using a combination of all texture features.....148
Table 6-6 Summary statistics for different classifiers when using a combination of all edge features.....149
Table 6-7(i) Feature sets selected by different feature selection algorithms implemented within WEKA.....150
Table 6-7(ii) Continue from Table 6-7(i) Feature sets selected by different feature selection algorithms implemented within WEKA.....151
Table 6-8 A comparison of statistics when features 2,4,9,11,12 (CfsSubsetEval selection) are used by different classifiers.....152
Table 6-9 Results for BayesNet with selected feature sets.....154
Table 6-10 Results for NaiveBayes with selected feature sets.....154
Table 6-11 Results for SVM with selected feature sets.....155
Table 6-12 Results for MLP with selected feature sets.....155
Table 6-13 Results for RBFNetwork with selected feature sets.....156
Table 6-14 Results for AdaBoost with selected feature sets.....156
Table 6-15 Results for Random Forest with selected feature sets.....157
Table 7-1. Detection accuracies.....171

Abbreviations

A_z	–	Area under Receiver Operating Characteristics curve
AdaBoost	–	Adaptive Boosting
ANN	–	Artificial Neural Network
ARD	–	Average Radial Gradient
BayesNet	–	Bayesian Network
BIRADS	–	Breast Imaging Reporting and Data System
B-scan	–	Brightness mode scan
BSE	–	Breast self-examination
BUS	–	Breast Ultrasound
CAD	–	Computer-Aided Diagnosis
CC	–	Cranio caudal
D	–	Multifractal dimension
D_m	–	Modified fractal dimension
FFDM	–	Full field digital mammography
FN	–	False negative
FP	–	False positive
FPF	–	False positive fraction
GVF	–	Gradient Vector Flow
IsoGauss	–	Isotropic Gaussian
ITS	–	Intelligent Tutoring System
LFS	–	Linear Forward Selection
LDA	–	Linear Discriminant Analysis
MCE	–	Minimum cross entropy
MLO	–	Medial lateral oblique
MLP	–	Multi Layer Perceptron

MRF	–	Markov Random Field
MRI	–	Magnetic Resonance Imaging
NaiveBayes	–	Naïve Bayesian Network
NHS	–	National Health Services
NHSBSP	–	National Health Services Breast Screening Programme
NPV	–	Negative predictive value
OGD	–	Oriented Gradient Detector
PACS	–	Picture Archiving and Communication System
PCA	–	Principal Component Analysis
PDE	–	Partial Differential Equation
PET	–	Positron Emission Tomography
PPV	–	Positive predication value
RBFNetwork	–	Radial Basis Function Network
RGI	–	Radial Gradient Index
ROC	–	Receiver Operating Characteristics
ROI	–	Region of interest
SFM	–	Screen film mammography
SMO	–	Sequential Minimum Optimisation
SVM	–	Support Vector Machine
T-rays	–	Terahertz rays
TN	–	True negative
TP	–	True positive
TPF	–	True positive fraction
US	–	Ultrasound
WEKA	–	The Waikato Environment for Knowledge Analysis

Breast Cancer Glossary of Medical Terms

(Courtesy of Imaginis [1] and Online Medical Dictionary [2])

Abscess: A closed pocket of tissue containing pus (a creamy, thick, pale yellow or yellow-green fluid that comes from dead tissue); most commonly caused by a bacterial infection.

Acini: Another term for the lobules of the breast. Lobules are milk-producing glands.

Adenosis: A more or less generalised glandular disease.

Benign: Not cancerous, not malignant. The main types of benign breast problems are fibroadenomas and fibrocystic change.

Breast density: Describes breast tissues that has many glands close together. Density shows up as a white area on a mammogram film. Though fairly common (especially in younger women), dense breasts may make microcalcifications and many other masses difficult to detect.

Calcifications: Small calcium deposits within the breast, singly or in clusters, that are usually found by mammography.

Carcinoma: A cancerous tumour that begins in the lining layer (epithelial cells) of organs. At least 80% of all cancers are carcinomas, and almost all breast cancers are carcinomas.

Duct: a hollow passage for gland secretions. In the breast, a passage through which milk passes from the lobule to the nipple.

Fat necrosis: The death of fat cells, usually following injury. Fat necrosis is a non-cancerous condition, but it can cause a breast lump, pulling of the skin, or skin changes that can be confused with breast cancer.

Fibroadenoma: A type of non-cancerous breast tumour composed of fibrous tissue and glandular tissue.

Fibroblasts: Connective tissue cells

Fibrocystic change: A term that describes certain non-cancerous changes in the breast, also called fibrocystic disease. Symptoms of this condition include cysts (accumulated packets of fluid), fibrosis (formation of scar-like connective tissue), lumpiness, area of thickening, tenderness, or breast pain.

Glands: Organs that produce and release substances used locally or elsewhere in the body.

Hyperplasia: The abnormal multiplication or increase in the number of normal cells in normal arrangement in a tissue.

Leucocyte: Generic term for a white blood cell.

Lumpectomy: Surgery to remove the breast tumour and a small margin of surrounding normal tissue.

Lumina: plural of lumen, the cavity or channel within a tube or tubular organ.

Malignant: Cancerous.

Mastectomy: Surgery to remove all or part of the breast and sometimes other tissue.

Nodularity: Increased density of breast tissue, most often due to hormonal changes, which cause the breast to feel lumpy in texture.

Pathology: The branch of medicine concerned with disease, especially its structure and its functional effects on the body.

Prognosis: A forecast as to the probable outcome of a disease, the prospect as to recovery from a disease as indicated by the nature and symptoms of the case.

Stromal cells: Connective tissue cells of an organ found in the loose connective tissue.

Chapter 1: Introduction

Breast Cancer is one of the most common and serious public health problems. Fortunately worldwide research efforts in a number of professional disciplines have led to significant progress against breast cancer. Better diagnosis and treatment methods that lower the chance of death from the disease and improve quality of life have been introduced. However through continuing research, knowledge about breast cancer keeps growing. Scientists are learning more about the causes of breast cancer and are exploring new ways to prevent, detect, diagnose and treat this disease.

Early detection of cancer plays a vital role in reducing mortality rates. Therefore many countries have established screening programmes where citizen groups who are at higher risks of developing cancer are routinely monitored. Breast screening in the UK is a national programme involving over 100 separate screening centres coupled with additional mobile screening vans which together annually screen approximately 1.8 million women [3] in the age group of 50-70. However virtually all screening is currently carried out using analogue films, with a rather small number of centres trialling computed radiology and full field digital mammographic screening. It is planned that by 2010 every screening centre will have at least one digital mammography unit. Therefore in the near future digital images are likely to play a major role in the screening, detection and treatment of cancer.

The underlying premise for breast cancer screening is that it allows for the detection of breast cancers before they become palpable. Breast cancer is a progressive disease. Thus small tumours are more likely to be at an early stage of the disease and therefore

have a better prognosis capability and are more successfully treated [4].

There are a number of research groups worldwide who are investigating breast sonograms. The main focus of these research groups is to create automated, ultrasound, Computer Aided Diagnosis systems with high sensitivity, specificity and consistency. Furthering these research efforts, in this thesis the aim is to develop a new method of ultrasound image processing for extracting relevant tissue structure information that will help differentiate between normal and malignant tissues. The ultimate goal is to provide fast and reliable tools for the early detection of malignant tissues in ultrasound images. The portability and low-cost features of the screening tools will allow for the deployment of this device on large scale.

This chapter provides an introduction to the problem domain and summarizes the solutions presented within this thesis. For clarity of presentation it is organized as follows: Section 1.1 identifies, defines and scopes the problem addressed in this thesis; Section 1.2 states the aims and objectives of the proposed research; Section 1.3 summarises the proposed approach; Section 1.4 lists the contributions made by the thesis to the state-of-the-art; Finally, section 1.5 provides an insight to the structure and presentation of the thesis.

1.1 Research Motivation

Mammography is the Gold Standard for breast imaging and early detection of breast cancer. However, this sensitive screening and diagnosis technique has high false-positive (FP) rates [5]. Some of the limitations are: inability to change image

contrast or brightness, problems in detection of subtle soft tissue lesions (dense glandular tissues), and the difficulty of archiving. Alternatively, it has been shown that ultrasound studies are good at picking up many of the cancers missed by mammograms, especially those which occur in women who have dense breasts [6, 7]. In addition it is non-invasive, portable, versatile, does not use ionizing radiations and is relatively low-cost.

Unfortunately due to the relatively poor quality of ultrasound images, their segmentation is a difficult problem [8]. It has been shown that reasons for ultrasound image degradations include, primarily, speckle noise, which can be modelled as a multiplicative degradation field, the blurring of spatial information perpendicular to the propagation direction of ultrasonic waves and the non-linear attenuation of ultrasound [9].

Unfortunately, segmentation algorithms based only on global information (such as thresholding techniques, global clustering approaches and intensity-dependant edge detection) is not suitable for ultrasound image segmentation and gives poor results [10]. This emphasizes a need to develop segmentation techniques that can avoid the intensity inhomogeneity problem, i.e., approaches that take into account the non-uniformity of tissue classes.

It is worthwhile noting that the automatic detection of Region-of-Interest (ROI) within a breast cancer screening scenario is not meant to replace the radiologist, but is to provide a tool to reduce the region labelling time of the radiologist. It has been shown that radiologists with different training backgrounds and experiences often get

rather different results in the reading of sonograms. Thus new techniques that assist physicians in improving the consistency of interpretation should be explored [11]. Currently there are many on-going research projects [12, 13] that are focusing towards creating Ultrasound Computer Aided Diagnosis (US CAD) systems with high sensitivity, specificity, and consistency. Unfortunately these systems are based on the assumption that the ROI will be pre-selected by the radiologist. Therefore there is a need to develop fully automatic ROI identification techniques that will assist in the decision making process of a radiologist.

Further in recent years there has been a workforce related issue within the UK breast screening programme in that there has been a shortage of consultant radiologists in the specialist domain [3]. Fortunately this has been offset by a growth in specially trained radiographers as part of the changing landscape for the radiographic profession within the UK. Currently there are some 109 screening centres regularly performing over 1.5 million screening examinations per year and a growing number of individuals (circa 650) nationally who report on these screening cases [3]. Hence, a need exists for the development and introduction of effective training tools that will help improve the performance quality of this relatively inexperienced workforce to acceptable national standards.

1.2 Aims and Objectives

Two main research efforts in early detection of breast cancer include the development of software tools to assist radiologists in identifying the abnormalities and the development of training tools to enhance their skills. Medical image analysis systems,

widely known as Computer-Aided Diagnosis (CAD) systems, play an important role in the development of such software tools. A typical US CAD system can be presented as in figure 1-1. In Ultrasound CAD systems the input consists of a rectangular region of interests, manually selected, and the output consists of a statistical analysis to aid the radiologist.

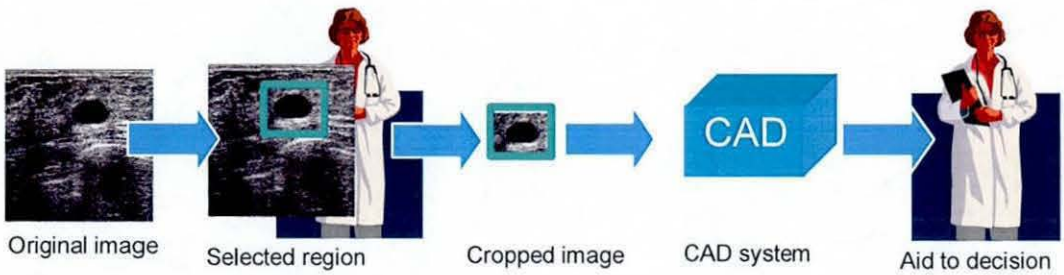


Figure 1-1 An existing US CAD approach.

The aims of the research presented in this thesis are: (i) to further enhance state-of-the-art computer vision algorithms and (ii) to improve human vision and decision making accuracy, in the development of training tools, by introducing the computer processed images in the training systems. Figure 1-2 illustrates the above aims.

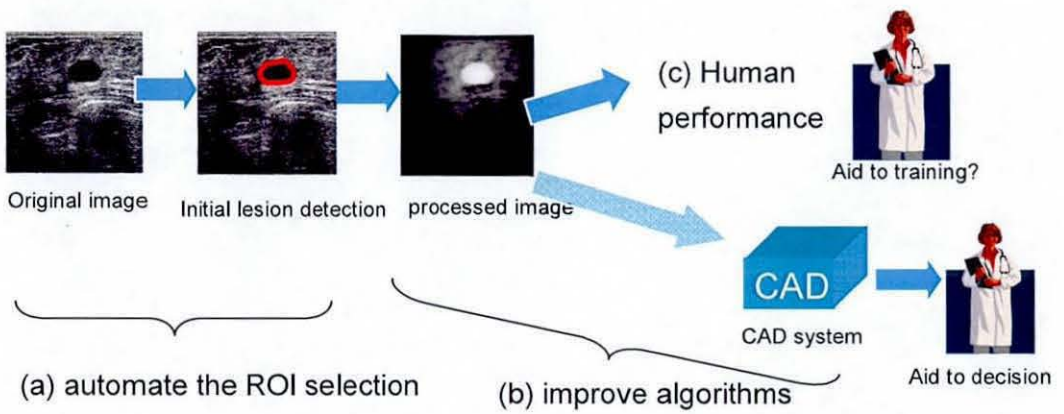


Figure 1-2 The aim of this research. (a) to automate the ROI selection, (b) to improve the segmentation and classification algorithms, (c) initial study in human performance by using the intermediate images.

A key objective of this research is to enhance the underlying algorithms in various stages of a CAD system, namely: fully automated initial lesion detection, boundary detection, and classification. In the case of breast cancer detection, an objective will be to develop an image-processing tool for extracting relevant tissue structure information that will differentiate between benign and malignant tissues. The ultimate goal is to provide fast and reliable techniques for the early detection of malignant tissue using ultrasound images. The portability and low-cost features of ultrasound imaging will allow the deployment of this device on large scale. This could be an interesting case study for sending medical US images from remote sites to general hospitals/referral centres. Further, in the development of training tools, the objective is to assist the radiologists in perceptual and cognitive tasks by adopting computer processed images into the training system.

1.3 Proposed Approach

The research presented in this thesis has been carried out in four different phases:

1. Understand the anatomy and pathology of breasts and review the state-of-the-art in breast cancer detection technology.

2. Review medical image analysis techniques used in reducing the speckle noise, shadowing and other artefacts in ultrasound images. Propose novel image filtering and segmentation methods to improve the initial lesion (ROI) detection and boundary detection accuracy.

3. Apply and adopt suitable classification methods to classify the detected and segmented lesions into different diagnosis categories, e.g., benign and malignant.

4. Conduct subjective studies in groups of experienced radiologists and non-radiologists / trainee radiologists on the effects of processed images in human perception.

In order to reduce the speckle noise in ultrasound images, effective filtering algorithms are needed. As a solution, the thesis proposes the use of a hybrid filtering approach, which combines nonlinear and linear filters. The result shows that this approach is more efficient than the state-of-the-art approaches, such as, nonlinear diffusion filtering, median filtering, and Gaussian filtering.

Accurate segmentation plays an important role in automatic ROI identification and ultimate lesion boundary detection. The thesis reviews state-of-the-art segmentation methods, such as, *k*-means, watershed, fractal, region growing and thresholding segmentation. A multifractal based approach is proposed in this thesis that contributes more effectively towards accurate segmentation of lesions.

The final computer based processing stage involves classification of detected lesions. This involves the use of feature extraction and their subsequent classification. To this extent the thesis investigates the use of standard feature extraction techniques and classification techniques such as appearance-based classifiers and texture-based classifiers. The review of literature has revealed that the application of standard

feature detectors and classifiers in ultrasound breast imaging has not been fully exploited. Therefore this thesis contributes to the knowledge of the state-of-art.

Finally to study the effects of processed images in the decision making process of radiological investigations, subjective tests were carried out on groups of expert and non-expert radiologists. Conclusions are drawn upon perceptual and cognitive tasks.

1.4 Contributions

The fundamental contributions of this thesis to the state-of-the-art are:

..1.1 A hybrid filtering approach that combines linear and non-linear filtering [P3, P4].

..1.2 Application of multifractal theory in initial lesion detection and segmentation [P1, P6].

..1.3 Two novel approaches to initial lesion detection:

(i) A combination of hybrid filtering, watershed segmentation and *k*-means clustering [P5].

(ii) A combination of hybrid filtering, multifractal processing, thresholding segmentation and a rule-based approach [P1, P7].

..1.4 A novel approach to the accurate segmentation of lesions in ultrasound breast images, to obtain the best possible boundary [P7].

..1.5 An Investigation into the use of appearance based and feature based approaches to lesion recognition in ultrasound images. In particular the use of various types of features (shape, texture and edge based), feature selection and classification algorithms to determine optimum computer vision pipelines for ultrasound image processing [P10, P11].

..1.6 Investigations in to the use of computer processed images in human perception and cognition [P2, P8, P9].

These contributions have resulted in seven conference papers, one poster presentation in radiological meeting, and two journal papers publications in proceedings of reputable international conferences and journals, respectively (see Appendix B).

1.5 Thesis Organisation

This thesis is organised into three parts (see figure 1-3). The first part includes non-contributory chapters providing fundamental and background knowledge of the subject area and the state-of-the-art in technological development. The second part of the thesis includes four contributory chapters. The third part concludes the thesis with an insight to future directions of research and development.

Part 1: Introduction, Background and Literature Review

(non-contributory chapters)

Chapter 1 provides an overview of the thesis. It defines the problem domain, states the research motivation and specifies the thesis aims and objectives. It further summarises the proposed approaches and highlights the contributions made by the thesis to the state-of-the-art. Finally it outlines the thesis organisation.

Chapter 2 presents fundamental knowledge about the anatomy, pathology of the breast, diagnosis of the cancer, the history of breast cancer, and how to reduce breast cancer rates. It further reviews different breast cancer detection techniques. The advantages and disadvantages of each approach are compared and discussed.

Chapter 3 provides a critical technical review of state-of-the-art image analysis approaches used in ultrasound breast imaging, including different filtering, segmentation and classification techniques. It further highlights constraints and limitation of the existing algorithms/algorithms.

Part 2: Novel algorithms for breast cancer detection and recognition

(contributory chapters)

Chapter 4 proposes two novel initial lesion detection techniques. The first approach uses hybrid filtering, watershed segmentation followed by *k*-means clustering. In hybrid filtering, a unique combination of nonlinear filter and linear filter is used. The second approach adopts multifractal processing and thresholding segmentation.

Results of both approaches are compared with a benchmark, state-of-the-art technique by Drukker et al [14].

Chapter 5 presents the novel solution proposed for effective lesion boundary detection in ultrasound breast images. The results are compared with the state-of-the-art approaches of Kupinski et al. [15] and Joo et al. [16]

Chapter 6 presents a detailed investigation on the use of appearance and feature based classification approaches to lesion recognition. In particular the use of appearance based approaches such as PCA, LDA and the use of feature based approaches such as Support Vector Machines, Neural Networks, Bayesian Neural Networks and other classifiers are investigated. The best feature-classifier combinations that are able to perform the recognition tasks most accurately are studied. Finally conclusions are made based on the analysis of experimental results.

Chapter 7 investigates the effects of computer processed images in human performance, both in perceptual and cognitive tasks. Two experiments were conducted on two distinct groups of subjects, namely, radiologists and non-radiologists. Results are critically analysed and conclusions made.

Finally, Chapter 8 concludes the thesis with a summary of contributions made by the thesis, limitations of the present context of research and an insight into future directions of research.

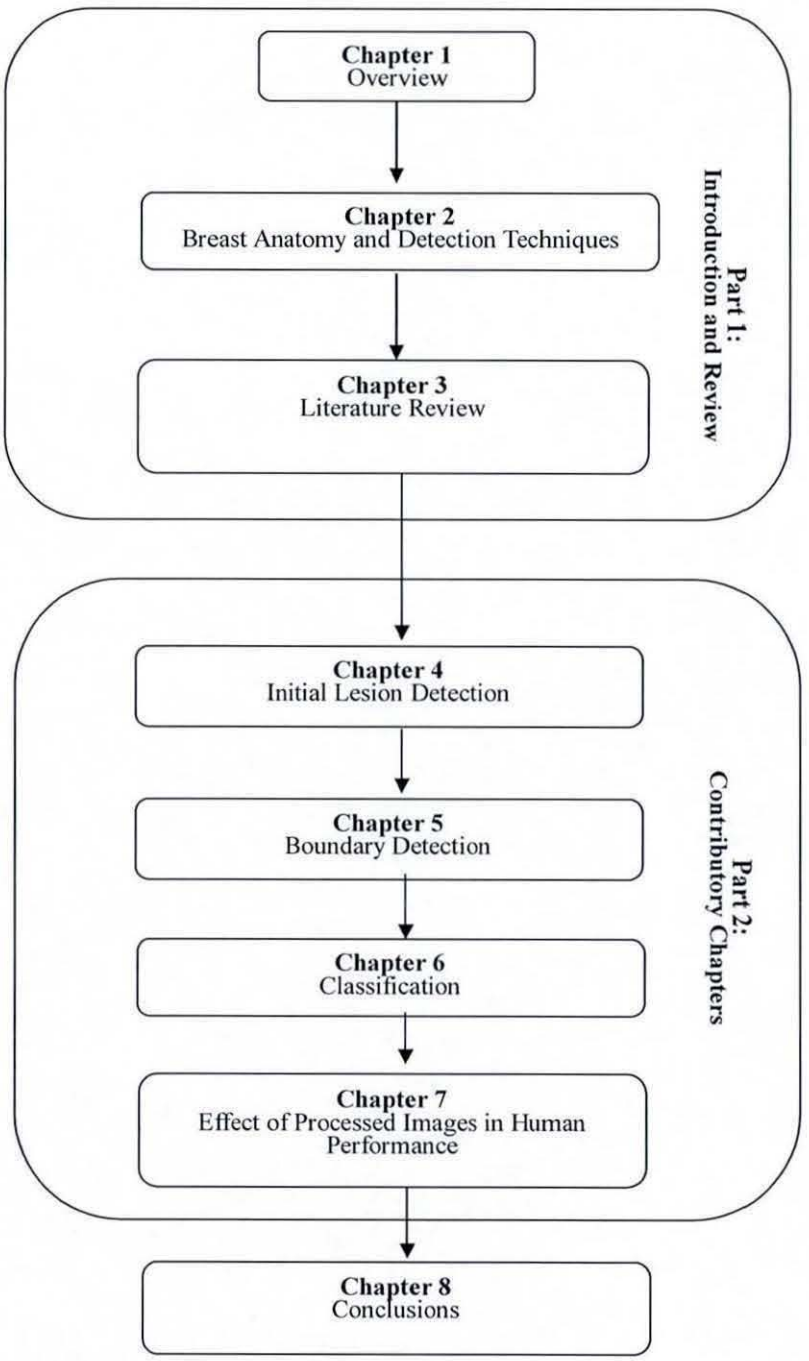


Figure 1-3 Thesis Organisation

Chapter 2: Background

This chapter introduces the readers to the anatomy of the breasts, breast screening and popular imaging techniques used. It provides vital information of the context in which research presented in this thesis has been built upon.

2.1 Anatomy of Breasts

The breasts are a complex structure as illustrated in figure 2-1. Their sensitivity to hormonal influences predisposes them to a number of pathological conditions.

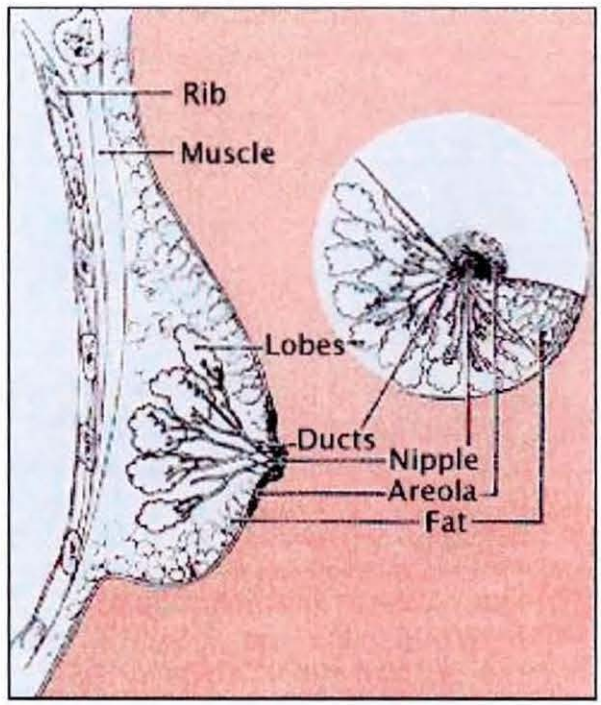


Figure 2-1 Anatomy of breasts. (Courtesy of [17])

The breasts are composed of 15 to 20 lobes in most women during their reproductive years [7], which have a relatively larger risk of developing pathological disorders. For

readers who are interested in studying details of the anatomy of breasts, please refer to [7].

Most of the human breast diseases are palpable, painful, nodules in the masses which are recognizable via human inspection often aided by effective medical imaging techniques. Fortunately there are many features/characteristics that can help a Sonographer, i.e. a medical technician training in performing ultrasound exams [18], identify the abnormalities of a sonographic image of the breast to provide an informed diagnosis.

Due to the relatively high risk of developing various pathological abnormalities in breasts, in some countries, routine breast screening programmes have been introduced for women of specific age groups.

2.2 The UK Breast Screening Programme

In the United Kingdom, Screening for breast abnormalities (that can lead to cancer) is carried out nationally for all woman aged between 50 and 70 years at regular three year intervals [3]. Fundamentally, screening involves imaging both medio-lateral oblique (MLO) and cranio-caudal (CC) views of each breast. These are then examined and are typically dual read by breast screening radiologists and/or specially trained radiographers (known in the UK as 'advanced practitioners'). It is noted that both professional groups are referred to as 'Mammographers' in this thesis, unless otherwise specifically stated. In breast screening, after a careful manual examination, if the

presence of an abnormality is suspected, imaging procedures (e.g. ultrasound) and/or further clinical investigations are subsequently undertaken. The role of Mammographers is therefore pivotal in first identifying suspicious mammographic appearances, making their level of professional skill crucial to the overall success of the national breast screening programme.

The most recently available data from the NHS Breast Screening Programme (NHSBSP) for the period 2005/2006 indicated that a total of 2,382,122 women were invited for screening and approximately 1,891,408 underwent screening. Of this, some 87,469 were recalled for further assessment of which 14,841 women with cancerous conditions were identified. The aim of the NHSBSP is to reduce the mortality caused by breast cancer through their early detection. The latest data show that this national screening programme has been instrumental in saving approximately 1,400 lives annually [3].

Currently in the UK, breast screening is performed using mammographic films although several trials of digital mammography that has greater potential in machine aided diagnosis have been undertaken. It is highly likely that nationwide full field digital mammography screening will be introduced in the near future. Several research projects are underway in the UK examining the use of digital screening and Computer Aided Diagnosis (CAD), building on well established, worldwide research efforts. Despite the lack of use of digital mammography in breast screening, it is noted that within the remainder of the UK health system, radiological imaging is predominantly digital, with PACS (Picture Archiving and Communication System) being very prevalent. The large high resolution digital images required for the

examination of breast images require associated high resolution workstations for their examination, coupled with robust hardware and software to enable these images to be viewed, when required, often at places remote to the originating screening centre. This has been identified as the bottleneck to the full scale digitization of the NHSBSP.

2.3 Diagnostic Breast Ultrasound

The objective of a diagnostic Breast Ultrasound (BUS) is to make a more specific non-invasive diagnosis in patients who have clinical or mammographic abnormalities than could be achieved with mammography and clinical findings alone [7]. Ultrasound has a greater ability than mammography to differentiate among types of normal tissues and to characterize complex cysts and solid nodules [7].

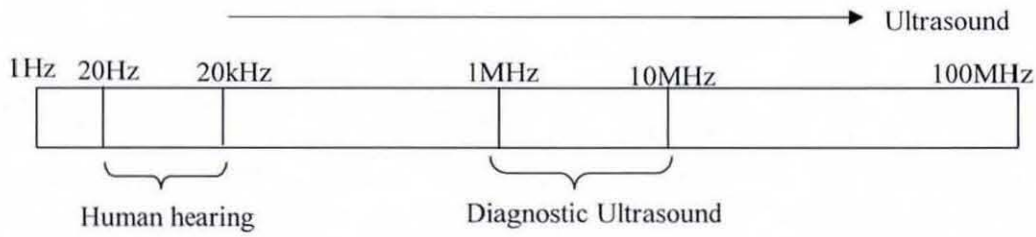


Figure 2-2. Frequency range of diagnostic ultrasound

A diagnostic Ultrasound (US) lies in the frequency range of approximately 1 to 10 megahertz (MHz) [6], as illustrated in Figure 2-2. The basic idea of diagnostic US is to transmit waves through the human body, which are partially reflected at the interfaces between tissue [19]. These reflections help form the US images that can

later be inspected for abnormalities. US imaging or Sonography is an important modality in the evaluation and treatment of breast masses. It is most often tailored to the patient, so as to answer a specific questions raised in the findings from mammography or physical examinations [20].

In US imaging, a so-called 'Brightness-mode scan' (B-scan) provides a two-dimensional image which is generated by the diffraction or back-scattering of an ultrasonic pulse from a cross section of the object that is scanned [21]. The term, 'attenuation', refers to total propagation loss due to absorption, scattering, and reflection. Absorption is a process to convert ultrasound beam energy to heat. Scattering at a tissue interface is called backscatter, while scattering from an irregular tissue interface is called diffusion. Both backscatter and diffusion result in increased ultrasound attenuation [6]. Meanwhile, reflection is proportional to the difference in acoustic impedance. In mechanical waves, the velocity is decreased when it is more compressible and denser [19].

The ultrasound wave propagation speed is generally called 'ultrasound velocity'. The velocity figures of different material can be summarized as in table 2-1 [6].

Table 2-1. Velocity figures of different material

Material	**Velocity (m/s)**
Air	330
Fat	1450
Water	1480
Human Soft tissues	1540
Blood	1570
Muscle	1585
Bone	4080

The most widely used modalities to construct a B-scan are linear, sector and arc shaped transducer movements. In US image database CD used in our experiments [22], the images have been obtained via a linear scan, which means the transducer has been moved in a straight line while scanning.

It is noted that the clinical use of ultrasound imaging commenced in 1962 and has since become very popular particularly in breast cancer screening. Given that the imaging volume capable of a US B-Scan is less than 20cm it is well applicable in breast imaging.

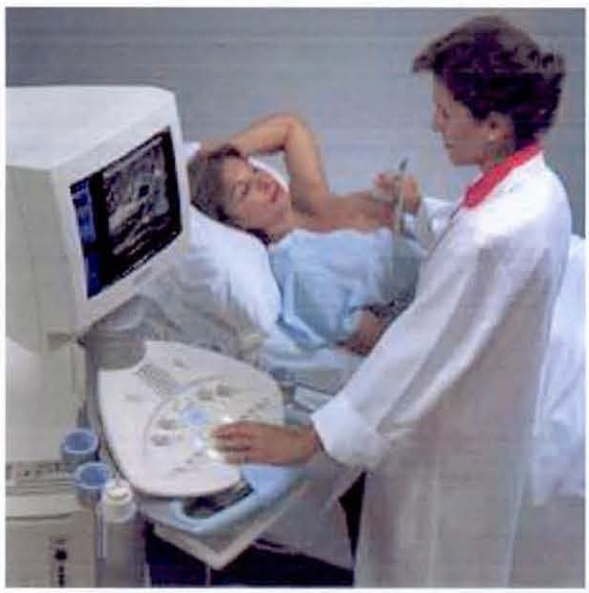


Figure 2-3. The radiologist holds the ultrasound transducer against the skin while imaging the breast. Image courtesy of [23].

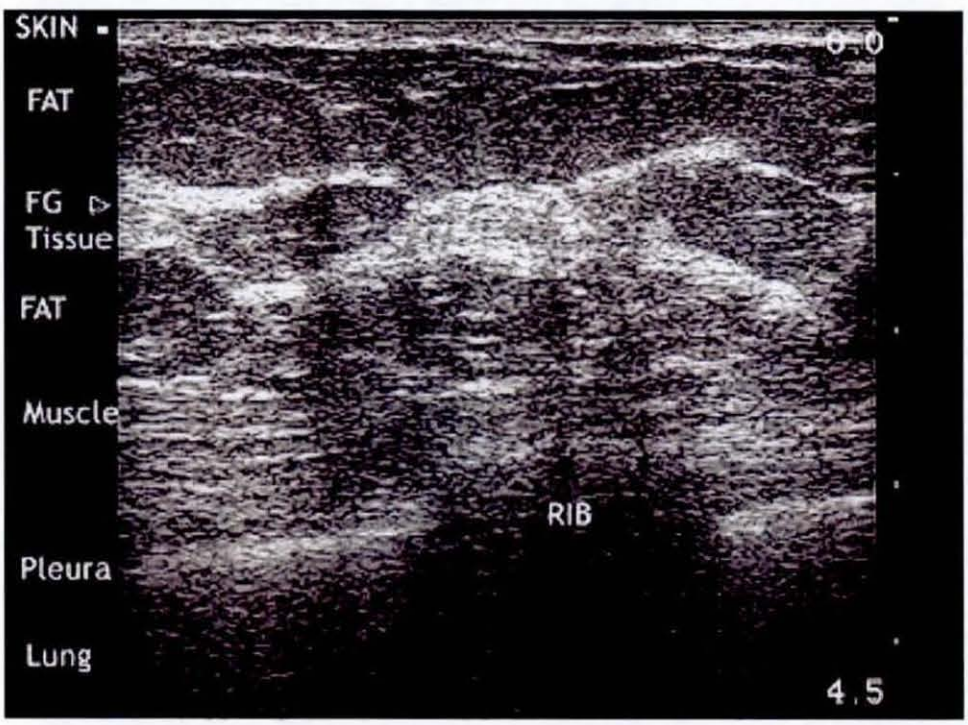


Figure 2-4. Sonography layer anatomy.
(Courtesy of Breast Ultrasound CD [22])

Figure 2-3 shows the radiologist holds the ultrasound transducer against the skin while imaging the breast. The ultrasound image appearance is illustrated in figure 2-4. Typically the skin of the breasts shows a thickness of up to 2mm (and up to 3mm in the periphery of the lower quadrants). During ultrasonography with a stand-off pad, the skin is displayed as a double hyper-echoic layer with a hypo-echoic intervening layer lying superficially [22]. Ultrasound can distinguish among different types of normal water-density tissues by echogenicity as well as thickness and compressibility [7]. However, from the figure 2-4, the limitation of ultrasound imaging is the existence of speckle noise.

2.4 Speckle Noise in US Imaging

US images in general are complex due to data composition, which can be described in terms of speckle information. The term 'speckle' is used for the granular pattern that appears in B-scan images and can be considered as a kind of multiplicative noise [24]. Upon visual inspection, speckle noise consists of a relatively high grey level intensity, qualitatively ranging between hyper-echoic (bright) and hypo-echoic (dark) domains [25].

Imaging speckle is a phenomenon that occurs when a coherent source and a non-coherent detector are used to interrogate a medium, which is rough on the scale of the wavelength [26], in other words, it is caused by the constructive and destructive interference of back-scattered signals due to unresolved tissue inhomogeneity. Due to this, speckle noise is a major limitation of ultrasound images, which affects human

interpretation of the images as well as the accuracy of the Computer-assisted diagnostic techniques/tools [27].

2.5 Benign Tumours

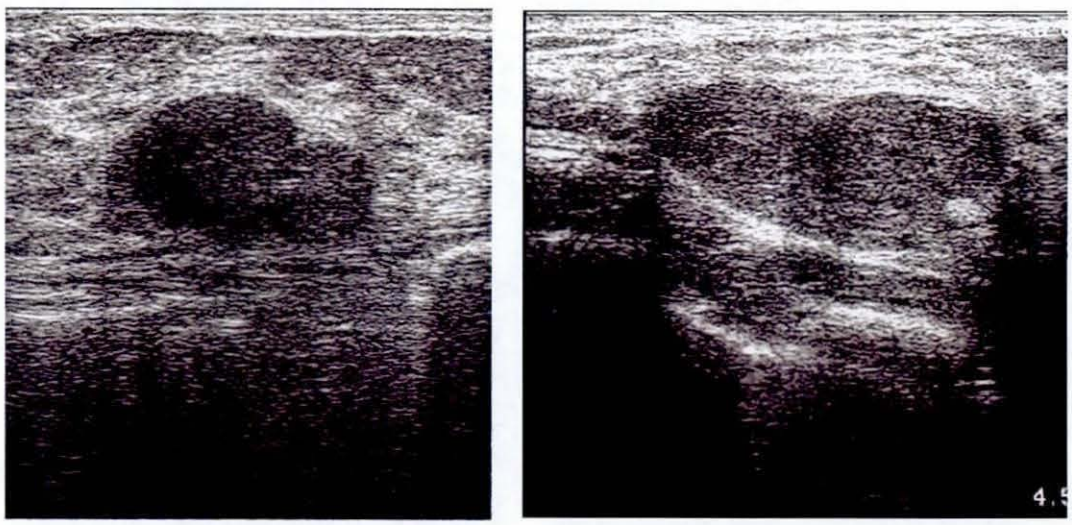
The histological spectrum of benign diseases includes lesions that produce microcalcifications such as cysts, adenosis, and irregular mammographic densities mimicking cancer; e.g. radial scars and fibrocystic change. None of these have malignant potential unless associated with epithelial hyperplasia [28]. Radiological correlates are changes in density with or without nodularity, ductal prominence, calcification or microcalcifications and mass lesions. These appearances result from changes in the lobular, ductal and stromal components of the breast.

Benign masses of the soft tissue density include cysts, and fibroadenomas [22]. In general, benign masses tend to be low density, and vessels may be seen through the mass. Malignant masses however are often denser than the adjacent parenchyma and may appear too dense for their size. A number of different types of benign tumours exist as described below:

2.5.1 Fibroadenoma

Fibroadenomas are the most common benign tumours in women between 20-30 years of age [28]. Figure 2-5 shows the Fibroadenomas tumours. From visual appearance, the masses present as firm, smooth, oval shaped, well-marginated (sometimes

lobulated) [28], freely movable entities. They grow as spherical nodules that are usually sharply circumscribed and movable from the surrounding breast substance. Histologically, a Fibroadenoma tumour is composed of various amounts of connective tissues and glandular elements and often has epithelium clefts [28].



(a) (b)

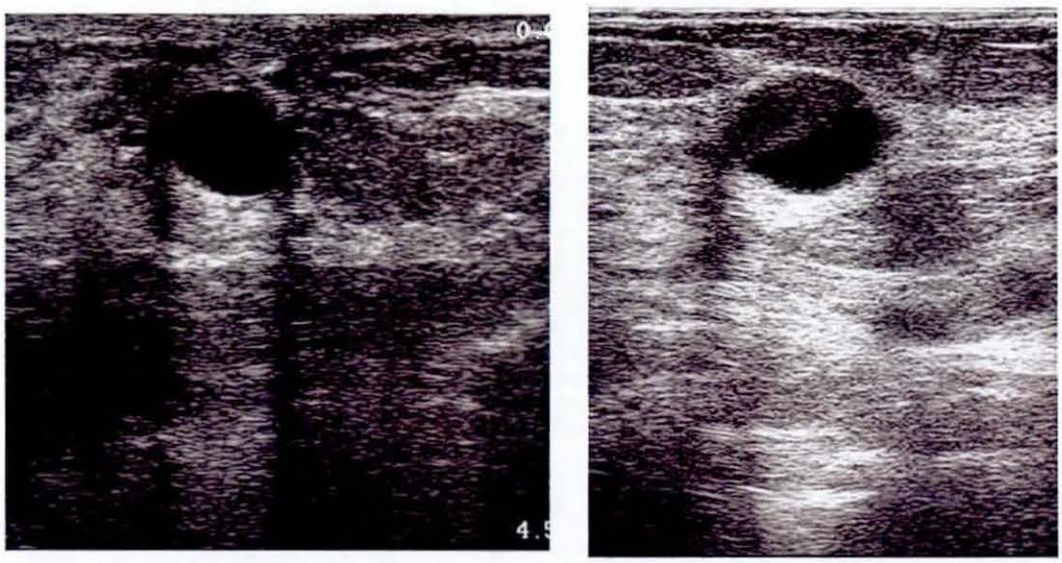
Figure 2-5. Benign tumours – Fibroadenoma

Fibroadenomas frequently occur in the upper outer quadrant of the breast. They vary in size from under 1cm to 15cm diameter. On a sectional view they are greyish white, and contain silt like spaces, often well circumscribed and shelled out.

2.5.2 Cysts

A cyst is round, anechoic, well-marginated, ovoid, or lobulated. Besides, from figure 2-6, it shown that cyst has well-defined posterior walls with acoustic enhancement. Breast cyst developed when the lumina of the ducts and acini become dilated and are

lined by atrophic epithelium. Simple cysts are often moveable within the breast, although they can also appear deep within the breast tissue [28]. Cysts may be divided into two categories, micro cysts, which are less than 3mm in diameter and 'gross cysts' which are 3mm or more in diameter.



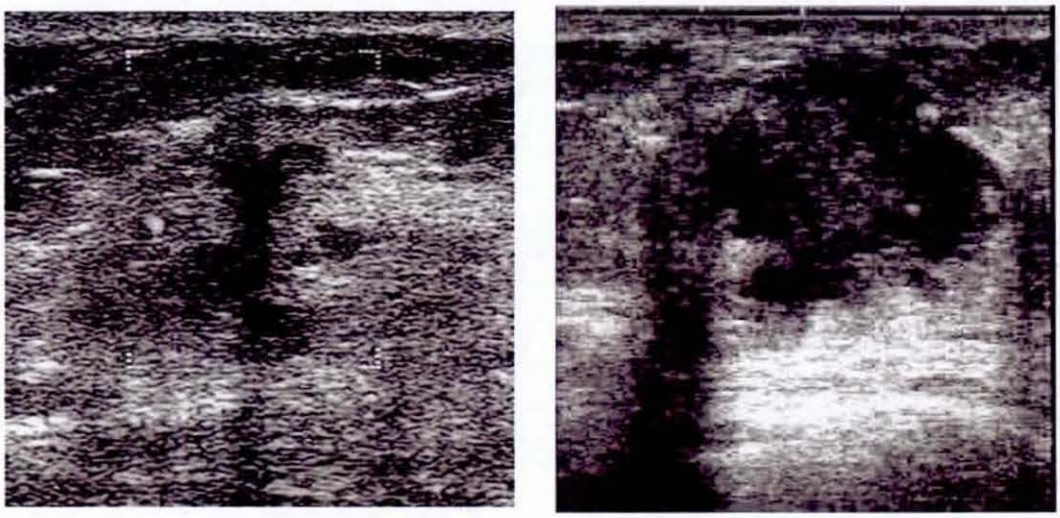
(a) (b)

Figure 2-6. Benign tumours – Cysts

2.6 Malignant Tumours

Characteristics of malignant lesions include hypo-echoic nature and irregular margination (especially interior margins, heterogeneous internal echoes, and acoustic shadowing), as illustrated in Figure 2-7. Malignant proliferation of breast epithelium results in breast carcinoma [22]. Carcinoma is the most common malignancy seen in the female breast. Breast carcinoma can be categorized into two main groups: (i) ductal carcinoma, where the malignancy originates from ductal epithelial cells and (ii)

lobular carcinoma, where the malignancy originates from more distal cells of the lobule [7].



(a) (b)

Figure 2-7. Malignant tumours

Each category can be subdivided into 'in-situ' (non-invasive) and infiltrating (invasive) types depending on whether the malignant cells have penetrated the underlying basement membrane. Infiltrating carcinomas have a much worse prognosis than in situ carcinomas. Each individual breast carcinoma is composed of a heterogeneous population of cell types including fibroblasts, leucocytes, and blood vessels [7]. With improved mammographic techniques, in situ ductal carcinomas are being detected with increased accuracy. On the other hand invasive carcinoma can be detected at very early stages, when the mass is only a few millimetres in size.

2.7 Review of Breast Cancer Detection Techniques

Breast cancer can be detected at a very early stage of development with modern technologies. It has been shown that early detection can significantly reduce the morbidity and mortality rates. The detection and diagnosis of breast cancer involves emotive issues and a test is required that is sensitive and specific for both the symptomatic and screened population [29] .

Breast self-examination (BSE) [30] or manual palpation is a very important part of every adult woman's personal health care. It is recommended that BSE should be performed once each month beginning at age 20 and should continue each month throughout a woman's lifetime. In addition to BSE, it is recommended that adult women should receive regular physician-performed clinical breast exams.

However there are some limitations to BSE and clinical breast exams. The major limitation is that the abnormal cells can only be palpated or felt when it grows to a certain size [30]. Therefore these techniques are not sensitive enough in breast cancer detection especially when the tumours are small in size.

A viable alternative to manual examination is the use of modern technology, such as imaging devices and signal processing algorithms, in developing tools/systems that are capable of improving the accuracy of human vision and judgment. In the following sections the popular techniques are introduced to the readers.

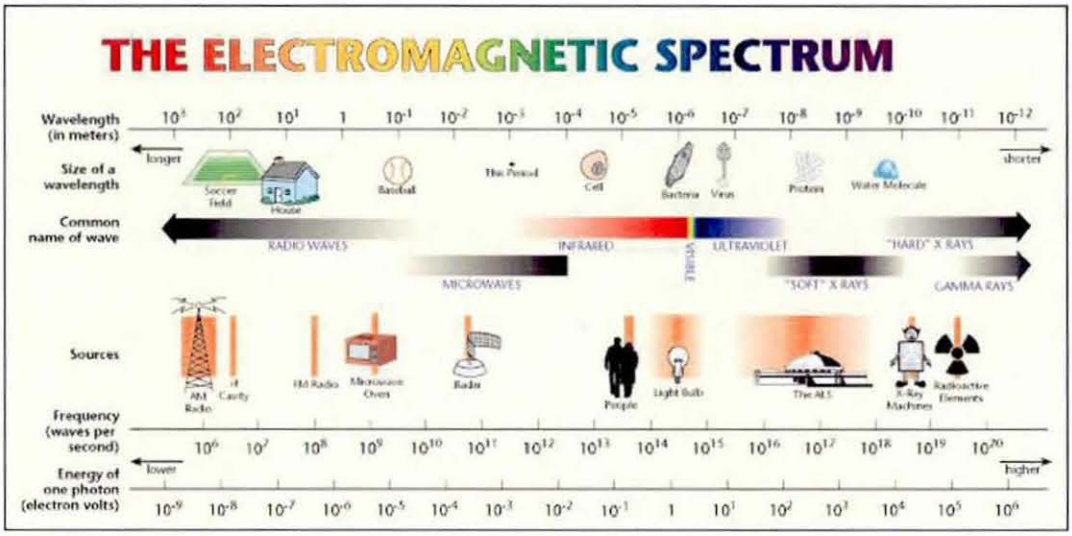


Figure 2-8. The electromagnetic spectrum
(Courtesy of [31])

2.7.1 Screen Film Mammography (SFM)

In general a mammogram is an x-ray of the breasts, consisting of two pictures per breast that are usually obtained on an annual basis [5]. They can be used to detect small cancers, known as a lump or a mass. Besides, calcifications and micro calcifications can be detected, which are white spots or granules similar to grains of salt.

SFM in particular is the Gold Standard for imaging of the breast and detection of early breast cancer. In this approach, a 'photographic film' serves as the image receptor, display as well as the storage medium [5]. The major advantage of SFM is that likely breast cancers can be found before palpation. However, compression of the breasts and radiation are necessary to create clear images. This causes considerable pain to those who have to undergo SFM examination.

Although SFM is sensitive in screening and diagnosis of breast cancer, it results in a high false-positive (FP) rate. Some of the limitations of this technique are inability to change image contrast or brightness, problems in the detection of subtle soft tissue lesions (dense glandular tissues), and difficulty in archiving.

2.7.2 Full-Field Digital Mammography (FFDM)

A Digital mammography unit looks like a conventional mammography unit as both use x-rays to image the breast. However, in a digital system, a phosphor screen is used instead of a photographic film as the image receptor. This screen converts these x-ray photons into light and the light is converted into a digitized signal that is displayed on a computer monitor.

The main advantage of such a digital imaging system is the possible separation of image acquisition, processing and display, allowing optimization of each of these steps. In addition, advanced applications such as computer-assisted detection/diagnosis (CAD) can be easily applied to the digital mammograms to assist in image interpretation [32].

Although there are benefits from image manipulation, transfer and storage, data from studies have not demonstrated a meaningful improvement in screening accuracy [33]. Some of the disadvantages of FFDM are high-cost, spatial resolution of detector used for high contrast objects being less than SFM, and no difference detection accuracy when comparing DM using both soft copy and hard copy formats [34].

2.7.3 Ultrasound (US)

Ultrasound is not perceptible to the human ear as it consists of sound waves with a frequency higher than 20 kHz. US uses some harmless and painless sound waves to produce a visual picture of the breast.

It has been shown that ultrasound studies are good at picking up many of the cancers missed by mammograms, especially those which occur in women who have dense breasts [35]. Ultrasound is non-invasive, portable and versatile. It does not use ionizing radiations and is relatively low-cost. Ultrasound applications for breast imaging include aspirating cysts, distinguishing solids from cystic lesions and guiding biopsies.

The main disadvantage of ultrasound is the poor quality of images due to multiplicative speckle noise and artefacts that cause information loss. However the performance of a high quality breast ultrasound examination not only depends on the scanner, but also requires a trained and experienced examiner with knowledge of the normal echo anatomy of the breast and the changes caused by the pathology. Therefore having trained staff is a vital part of the success of the use of US imaging in breast examination.

Modern US devices are digital in their design and can use real-time, image enhancement and computer-aided diagnosis systems (US CAD) very easily. For example on-device enhancement of images is possible via contrast changes and colour manipulation. Novel techniques for US colour pre processing are presently under

development and promises new insights into the use of US in breast imaging.

2.7.4 Magnetic Resonance Imaging (MRI)

Research into breast MRI began in the 1980s [33]. MRI uses magnetic fields and the magnetic properties of the body to image various atoms and molecules in the body. Breast tumours have their own blood supply when they reach a certain size. MRI does not detect calcifications but images blood flow to determine the shape and size of tumours.

MRI is a sensitive method for the detection of breast cancer. Sensitivity is one of the strengths of MRI [36]. Some other advantages are the non-invasive nature and the ability to generate multivariate images. It is also useful in the detection of recurrences after surgery has been performed for mastectomy or lumpectomy cases.

The disadvantages of MRI are that it is expensive and time consuming. Besides, patients using cardiac pacemakers and other metallic devices or implants cannot be examined via the use of MRI. The lack of specificity will decrease its acceptability to clinicians and patients [29]. MRI is an impractical tool for routine screening, but plays a major role as an adjunct to mammography and US imaging for specific problems.

2.7.5 Positron Emission Tomography (PET)

PET is one of the imaging modalities that can be used for taking pictures of the breast to see if cancer is present in the breast tissues. The basis of tumour imaging with PET

is a specific uptake mechanism of positron emitting radiopharmaceuticals [37]. PET appeared to be effective in the study of patients with ambiguous mammography.

PET provides more accurate information in discriminating between viable tumour, scar or necrosis. PET will be primarily utilized for four indications in breast cancer: patients who have breasts that are difficult to evaluate, patients who have had previous surgery, patients who have had radiation therapy, and patients who have had breast implants [38].

Unlike most other imaging tests that are based on changes tumours cause in the body's structure, PET scanning depends on changes in tissue metabolism. PET is being used to detect the metastatic disease (cancer spread) and has been successful in this role. It is not currently used for primary breast cancer detection because it does not reliably detect tumours smaller than 1 cm, but research to improve the accuracy of this test is currently in progress.

2.7.6 Terahertz Rays (T-rays)

Terahertz radiation lies in between microwave and infrared radiation in the electromagnetic spectrum and it attracts attention for various imaging purposes [39]. Terahertz imaging is becoming a more recognized imaging technique, which is still at the early stages of development.

T-rays carry much less energy than X-rays, but T-ray can penetrate through matter without the damage associated with the ionizing X-rays [40]. T-ray imaging does not

pose a safety risk and allows for excellent imaging detail. T-rays provide good contrast between diseased and health cells, because they are sensitive to the water content and chemical markers of cancer and other diseases.

The major complication in the use of T-rays is that the signal is not 'pure'. Their generation is noisy. Their detection, each point on the object, is represented by the burst of energy extended in time and space. Hence the image received does not consist of simple pixels. The use of Terahertz imaging has been limited due to lack of cheap methods for generating and detecting terahertz rays.

2.8 Summary

This chapter introduced the readers to the anatomy of the breasts, breast screening and popular imaging techniques used in breast screening. It provided vital information of the context in which the research presented in this thesis will be built upon. The advantages of using ultrasound imaging in breast screening were highlighted. The importance of image enhancement in ultrasound imaging and the possibility of subsequent computer assisted diagnostics were discussed.

In Chapter 3 the reader is presented with a report of the state-of-the-art in automatic/semi-automatic breast cancer detection. The idea is to critically review them so as to be able to design novel systems that will perform better in terms of accuracy and functionality as compared to those of the state-of-the-art.

Chapter 3: Literature Review

3.1 Introduction

The typical pipeline of a conventional Ultrasound Computer-Aided Diagnosis (CAD) system (or the typical processing stages for 2D medical breast images) is illustrated in figure 3-1.

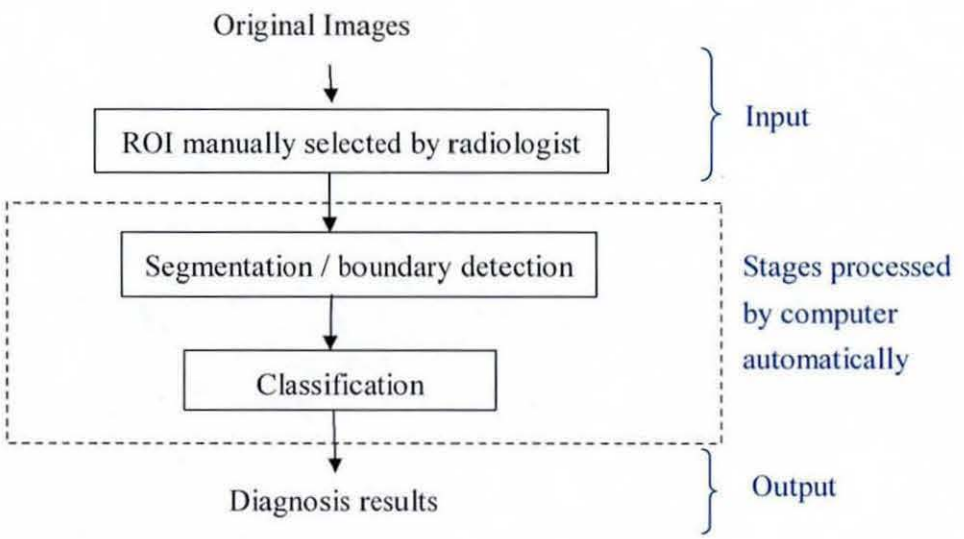


Figure 3-1 Diagram illustrates the flow of CAD processes.

In this chapter, the detailed reviews on literature are organized and presented according to the processing stages depicted in figure 3-1. In Section 3.2, research associated with common pre-processing and filtering issues in ultrasound images is discussed. In Section 3.2.1, a literature review and analysis of automated and manual selection of ROIs is presented. Section 3.2.2 reviews the segmentation methods that have recently been applied in a number of state-of-the-art US CAD systems. The classification methods used in existing research work are discussed in section 3.2.3.

Finally, a review of research in mammography is conducted and presented in section 3.3.

3.2 Ultrasound

A crucial step in ultrasound image processing is 'despeckling' or speckle noise reduction. To improve clinical diagnosis, speckle reduction is generally used for visualization enhancement and auto-segmentation improvement [41].

To visually illustrate the difficulties in processing ultrasound images, three efficient edge detectors, namely, Canny, Susan and Oriented Gradient Detector (OGD) are applied on the popular greyscale test image of Lena and on a number of ultrasound breast images (see figure 3-2). It is shown that the edge detectors worked well for the Lena image, but not for the ultrasound test images used. US images produce too many edges and fail to show the edges of the likely regions of interest. Based on this fact it can be concluded that edge detectors are not suitable in labelling the regions of likely tumours in an US image. A thorough review in filtering algorithms is needed to obtain the best filtering algorithm for ultrasound images.

Median Filter is a well-known filtering algorithm in computer vision. It is popularly used in filtering speckle noise in images. A number of researchers, to name a few, Joo et al [16], Kupinski et al [15], Drukker et al [14] implemented median filtering in the pre-processing stage of the proposed CAD systems. However further analysis of the effects of using median filtering revealed that it affected the accuracy of the boundary

detection as it removes important edge information, while removing speckle noise.

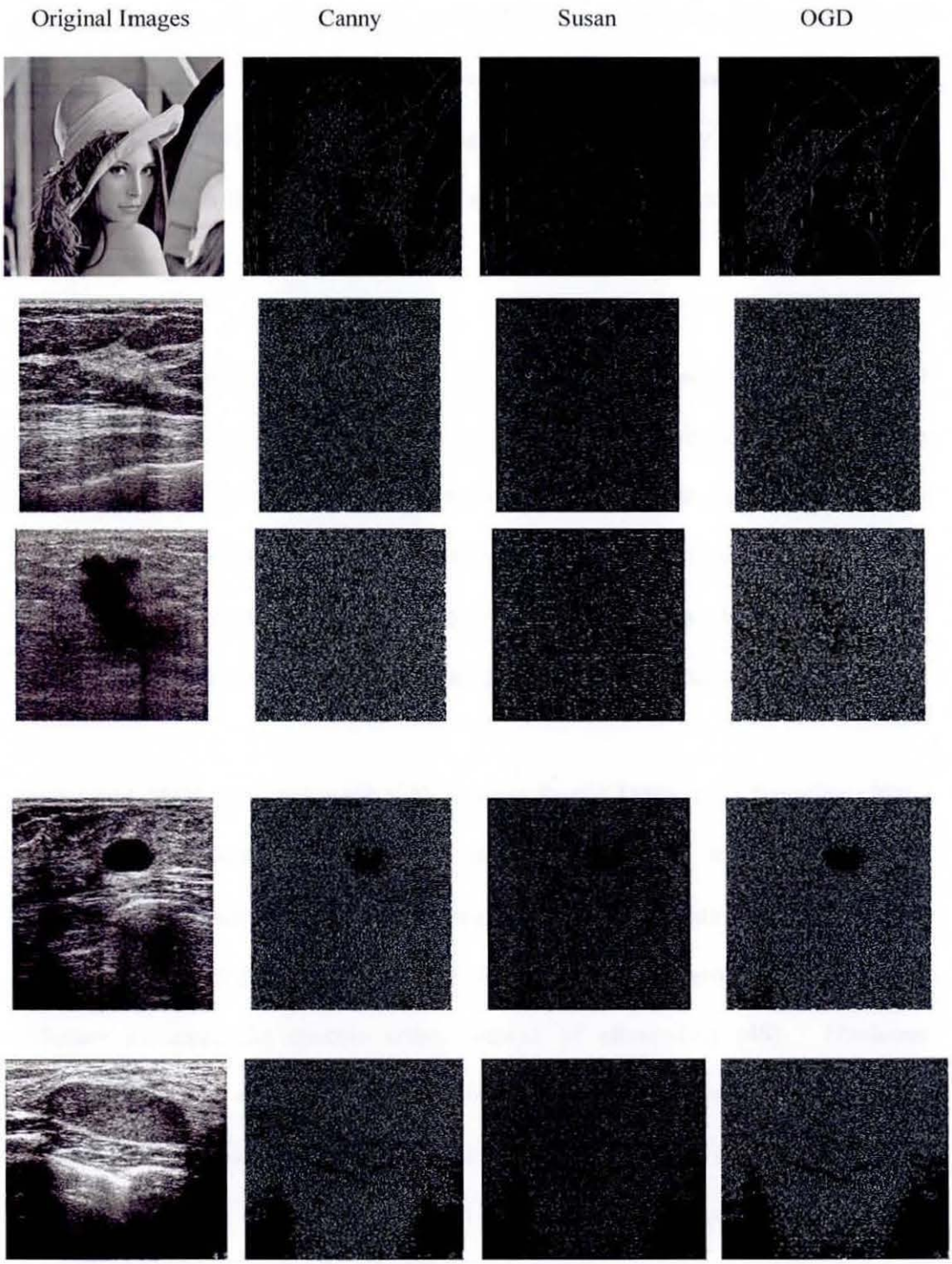


Figure 3-2 Speckle noise causing difficulties in detecting the edges in ultrasound images

Chen at al [42] implemented Gaussian blur in their work. Gaussian blur is a linear filtering approach which is implemented to reduce the common oversegmentation problem in ultrasound images. This is a very effective way in removing the speckle noise, but blurs and dislocates the edges when moving from finer to coarser scales [43]. Therefore, even though linear filtering is unique and is easy to handle, it cannot positively handle any additional information on structures, which are worth being preserved or enhanced [44].

Chen D.R et al [45] pre-processed the ultrasound images using a combination of morphological operations and histogram equalization. Morphology is an approach to image processing based on shape information. Morphological approaches, such as erosion and dilation, are based on curve evolution and are therefore difficult to handle and require prohibitively small time steps for effective operation. Further, they suffer from the problem of coping with singularities and topological changes [46].

Perona and Malik [47] proposed a Non-linear Partial Differential Equation (PDE) approach for smoothing images on a continuous domain. It was shown that Anisotropic Diffusion performs well for images corrupted by additive noise. However in cases where images contain speckle noise, it can be shown that anisotropic diffusion enhances the speckle noise, instead of eliminating [48]. Nonlinear Diffusion filtering, introduced by Weickert [44] has deservedly attracted much attention in the field of image processing for its ability to reduce noise while preserving (or even enhancing) important features of the image, such as edges or discontinuities as opposed to linear diffusion.

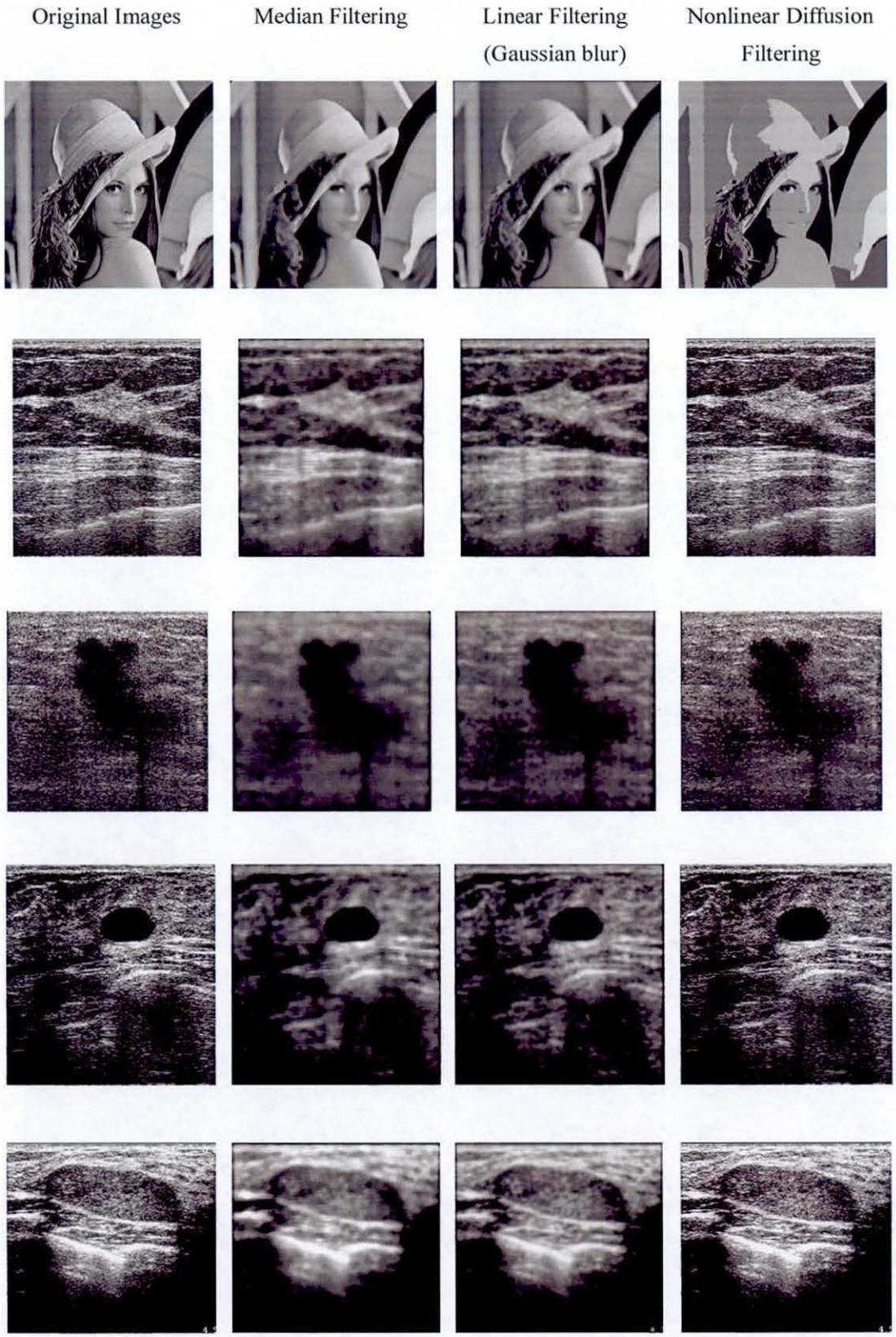


Figure 3-3 Comparison between speckle reduction filters.

Summarising the above review it can be said that the existing algorithms have their strong points as well as shortcomings. Thus there is scope for further improving the pre-processing algorithms used in US imaging. In Chapter 4 we propose improvements to the existing algorithms and compare their performance with that of the key pre-processing algorithms discussed above.

3.2.1 Initial lesion detection

A region of interest (ROI) is defined as a region that contains the suspected lesion/tumour. In computer based analysis these regions are often considered as rectangular areas enclosing the lesion/tumour. It is worthwhile noting that the automatic detection of Region-of-Interest (ROI) within a breast cancer screening scenario is not meant to replace the radiologist, but is to provide a tool to save the region labelling time of the radiologist. It has been shown that radiologists with different training backgrounds and experiences often get rather different results in the reading of sonograms [11]. Thus new techniques that assist physicians in improving the consistency of interpretation should be explored.

A significance proportion of on-going research projects are focused towards creating Ultrasound Computer Aided Diagnosis (US CAD) with high sensitivity, high specificity, and high consistency, to name a few, Gurney 1994 [49], Boone 1993 [50], Chen et al 2003 [51], and Sehgal et al 2004 [52]. Unfortunately these systems are based on the assumption that the ROI will be pre-selected by the radiologist. Chen et al managed to develop an accurate CAD system [51], but according to them, the manual operations by the physician to locate the ROI before the computer analysis are

still needed. In [53], the lesion boundaries were delineated manually. In [8, 10, 45, 54], the ROI is manually picked up by experts.

The review of existing literature revealed that most CAD systems rely on an initial, manual identification of the ROI by an experienced radiologist. Only Drukker et al (2002) [14] has initially worked on the full images to subsequently obtain the labelling of lesion using an automated approach. In this work they used a combination of median filtering, Radial Gradient Index (RGI) filtering, thresholding, in order to detect the initial lesions automatically. Segmentation of lesion candidates is done by selecting the maximum average radial gradient (ARG). At an overlap of 0.4 with a radiologist lesion outline, 75% of the lesions were correctly detected.

3.2.2 Segmentation

Methods for segmentation vary depending on the type of application, image modality, and other factors. There is currently no single segmentation method that can produce acceptable results for every medical image. Further, segmentation algorithms have had fairly limited application in ultrasound imaging.

This section describes several common segmentation approaches that have been proposed in the recent literature. The segmentation methods can be divided into thresholding approaches, region growing approaches, clustering approaches, Markov random field models, multiresolution texture segmentation, watershed segmentation, and the use of deformable models.

Thresholding approach is the simplest segmentation algorithm. A thresholding procedure attempts to determine an intensity value, called the threshold, which separates the pixels into desirable classes. Joo et al implemented one valued binary thresholding and edge detection to obtain the boundary of ultrasound images [16]. Thresholding is often used as an initial step in a sequence of image processing operations. For instance, an automated thresholding is applied by [51] to obtain the initial boundary of the tumour, which is subsequently enhanced by an active contour model.

Region growing is a technique for extracting a region of an image that is connected based on some predefined criteria, which can be based on the intensity information and/or edges in the image [55]. Drukker et al [14] implemented region growing by using the Average Radial Gradient index in obtaining the best boundary of the lesion. Kupinski et al [56] implemented region growing with Maximum Radial Gradient Index to obtain the best boundary.

Clustering algorithms are an unsupervised method. In order to compensate for the lack of training data, clustering methods iterate between segmenting and characterizing the properties of the each class. Some of the breast ultrasound imaging approaches that uses clustering approaches for segmentation includes, Boukerroui et. al's [54] and Ashton et. al's [8] work. The common clustering algorithms used in medical image analysis are *k*-means clustering and adaptive clustering.

Markov random field (MRF) modelling is a statistical model that can be used in segmentation. In practice a MRF is incorporated into clustering segmentation

algorithms under a Bayesian prior model. Pappas [57] modified the popular *k*-means clustering algorithm, by incorporating spatial smoothness constraints. Markov random fields are used to model the region segmentation process and assumes that the pixel intensities are given by slowly varying class mean corrupted by globally uniform additive white Gaussian noise.

Muzzolini et al. [58] proposed a multiresolution texture segmentation approach. They generalized the conventional simulated annealing methods to a multiresolution framework and minimized an energy function, which is dependent of the resolution and the size of the texture blocks of the image.

Watershed Segmentation is simple and intuitive, can be parallelized, and always produces a complete division of the image [59]. Huang et al. [60] integrated the advantages of neural network classification and morphological watershed segmentation to extract the precise contours of breast tumours from ultrasound images.

Deformable models, snakes, or active contour models are curves defined within an image domain that can move under the influence of internal forces within the curve and external forces computed from the image data [61]. Yezzi et al.[62] applied the deformable model to segment cysts in ultrasound breast images. In Chen et al.'s [51] research, after an initial contour of the segment has been detected, an active contour model is used to improve the initial segmentation results. Furthermore, the real-time acquisition capability of ultrasound images makes it better suited for motion estimation tasks [63, 64], where dynamic nature of active contours are often exploited.

Future research in medical image segmentation will strive towards improving the accuracy, precision, and computational speed of the state-of-the-art methods.

3.2.3 Classification

The final stage of a CAD system is classification. In classification, it is very important to achieve high sensitivity and specificity. It is known that benign and malignant tumours differ in shape, contour, margin and echogenicity of the lesions [35], which provide means of differentiation between the two types of lesions. To further elaborate on ultrasound breast image classification, section 3.2.3.1 explains the existing related standards of human interpretation and associated classification, followed by section 3.2.3.2, which discusses the state-of-the-art in computer aided classification algorithms.

3.2.3.1 Standards of human interpretation

The Breast Imaging Reporting and Data System (BIRADS) [65] is an international standard used for the reporting of mammographic findings, clarification of its interpretations and facilitates communication between clinicians. It was published by American College of Radiology (ACR) in 1993. A careful study of the Positive Predictive Value (PPV) of mammographic features described in the mammography BIRADS lexicon, shows that it is useful in differentiating the diagnosis of the breast lesions [65].

From the expert interpretation of humans, Heinig et al. [65] characterized lesions using BIRADS ultrasound descriptors of:

i) Mass margin: circumscribed, obscured, microlobulated, ill-defined/indistinct, or speculated.

ii) Shape: oval, round, lobular, or irregular

iii) Orientation: parallel or not parallel to skin

iv) Matrix echogenicity and homogeneity: anechoic, hypo-echoic or hyper-echoic; homogeneous or heterogeneous

v) Attenuation: indifferent, shadowing or enhancement

vi) Others: any associated findings (architectural distortion) or axillary lymphadenopathy

3.2.3.2 Computer based methods in classification

A neural network is a model that is inspired by the functions of biological neurons. They are extensively used in data classification. In 1999, Chen et al. [66] proposed the use of a neural network classifier with autocorrelation features to classify tumours in ultrasound images. In year 2000, Chen et al. [67] implemented another neural network model, using 24 autocorrelation texture features in classification. Chen et al. 2002 [68]

classified the benign and malignant tumours with a multilayered perceptron (MLP) neural network which trained using the error back-propagation algorithm with momentum. The classification was based on three feasible features, the variance contrast, autocorrelation contrast, and distribution distortion of wavelet coefficients. In 2003, Chen et al. 2003 [53] extracted seven morphological features, and classified the tumours by using a multilayer feed-forward neural network. Joo et al. 2004 [16] used ANN models to distinguish the benign and malignant tumours based on five morphological features representing the shape, the edge characteristics, and the darkness of the nodule. Drukker et al. (2002) [14] implemented a Bayesian neural network with round robin analysis.

Fractal analysis based classification methods have mainly focused on the classification of digital mammographic images. Garra et al. [69] used fractal analysis and statistical texture analysis methods in ultrasound images and concluded that the most useful features in classification are the co-occurrence matrix of the ultrasound images. Subsequently, Chen et al. [45] developed a CAD system in which fractal analysis was used to classify tumours.

Principal component analysis (PCA) and Linear Discriminant Analysis (LDA) played important roles in classification. Horsch et al. [70] implemented LDA in their research, while Hadjiiski et al. [71] classify the malignant and benign masses based on adaptive resonance theory mixed with a supervised linear discriminant classifier, namely the hybrid ART2LDA approach.

Most of the latter work in ultrasound image classification involved the use of Support Vector Machines (SVM), for example, Chang et al. 2005 [12], Huang et al. 2005 [72] and Huang et al. 2008 [13]. Support Vector Machines (SVMs) are a set of related supervised learning methods used for classification and regression [73]. Feature extraction is a pre-stage in SVM. Morphological features such as form factor, roundness, aspect ratio, solidity, convexity, and extent features were utilized in Chang et al.'s work [12]. Huang et. al 2005 [72] evaluated a series of pathologically proven breast tumours using SVM in the differential diagnosis of solid breast tumours. The inter pixel textual features are used in the SVM classifier. Huang et al. 2008 [13] applied the combination of morphological features in Chang et al.'s work [12] and Chen et al.'s work [53] to enhance the accuracy of classification.

From the study, it is important to realize that the use of texture descriptors alone, or shape descriptors alone, does not provide effective classification. Hence, research into the combined use of texture descriptors, shape descriptors, fractal dimensions and Fourier descriptors is important. Further the use of feature selection and feature based classifiers can be used as means of providing improved classification accuracy.

3.2.4 Test Performance Measures

In medical diagnosis, the principles of diagnosis modelling and computation are identical to that used in engineering, scientific evidence, finance etc. The test characteristics can be illustrated as in Table 3-1.

Table 3-1 Test characteristics

Disease / Test result	**Disease present**	**Disease absent**	**Total**
Positive	True positive (TP)	False positive (FP)	TP+FP
Negative	False negative (FN)	True negative (TN)	FN+TN
	TP+FN	FP+TN	

There are two types of errors in medical diagnosis: *false-negative (FN)* and *false positive (FP)* findings. *True positive fraction (TPF)* or *sensitivity* is the percentage of diseased patients that have a positive test. *Sensitivity* also known as *recall*. *True negative fraction (TNF)* or *specificity* is the percentage of non-diseased patients that have a negative test. Positive predictive value, also known as *precision*, is the percentage of positive tests on patients that actually have the disease. Negative predictive value is the percentage of negative tests is on patients who are free from the disease.

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

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

It is noted that the *sensitivity* and *specificity* depend on the cut off value between normal and abnormal. It can be shown that there is always a trade off in setting the cut-off point. (See figure 3-4).

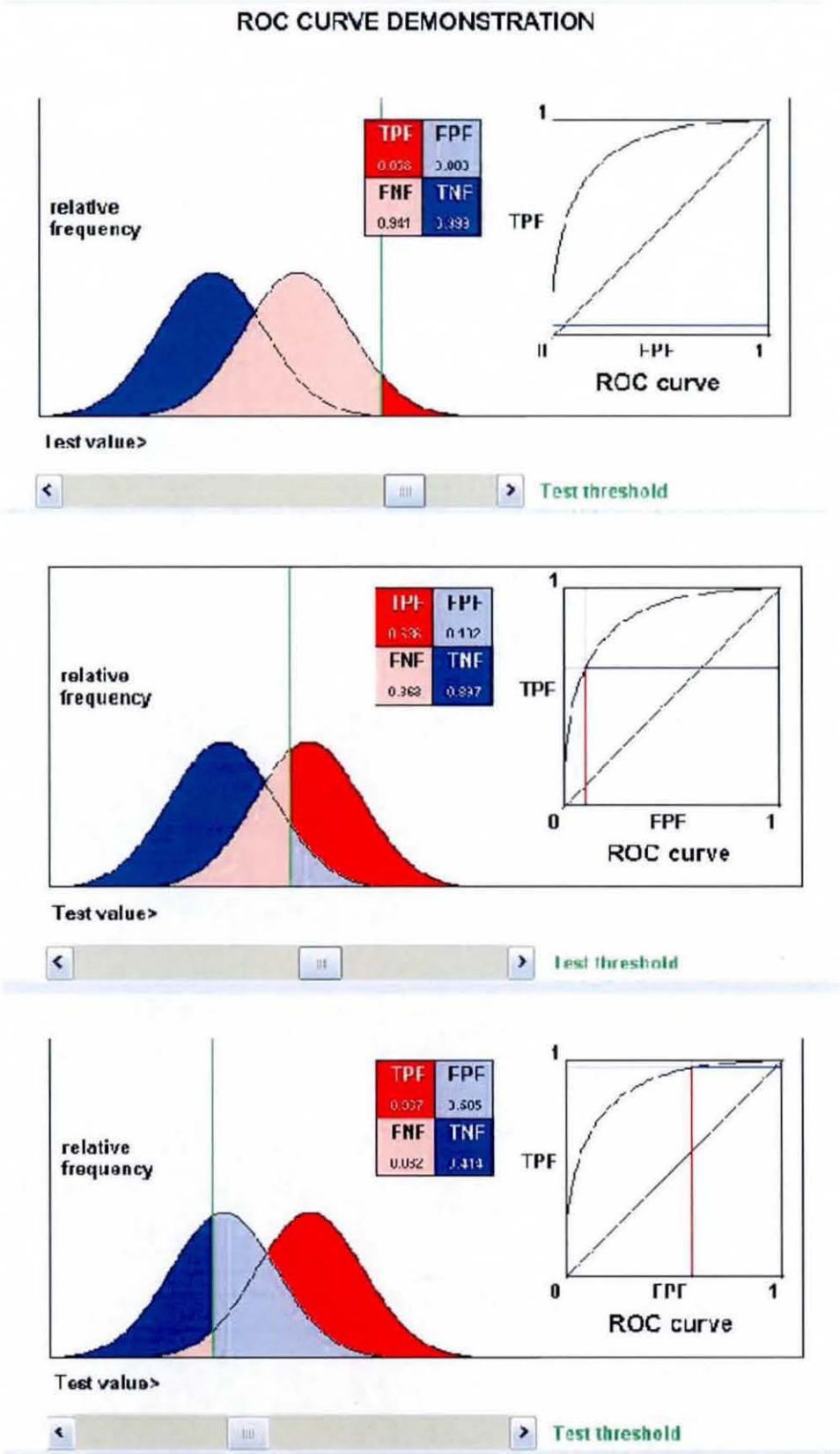


Figure 3-4 The trade off of the cut off value (courtesy image to [74])

The performance of a CAD system or the results of a classification experiment are often evaluated/measured using Receiver Operating Characteristic (ROC) curves. ROC curves are a way of estimating the best cut-off point for a test measured on a continuous scale. For comparing the relative performance of two tests, ROC curves plot the sensitivity versus 1-specificity. The single measure, area under curve (A_z) is normally used for comparison. Figure 3-5 shows an example of the ROC curves.

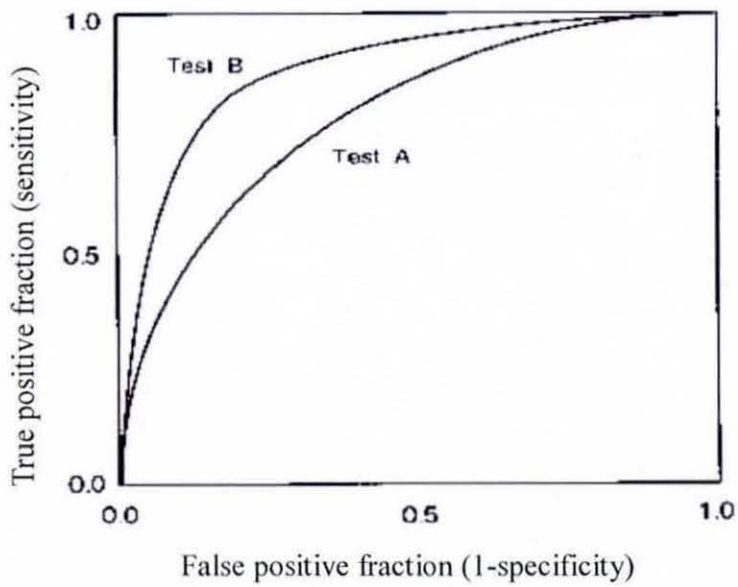


Figure 3-5 Examples of ROC curves.

3.3 A Survey of state-of-the-art CAD systems

This section presents a number of Computer Aided Diagnostic systems, i.e. systems that provide a total solution to the detection and recognition/classification problem of US images.

Stavros et al. [35] proposed a set of classification rules for human judgement. Their research had reported 99.5% (424/426) negative predictive value and 98.4% (123/125) sensitivity when following their approach. It was based on 20 specific sonographic features of breast masses, including morphologic descriptors of shape, margin, texture of a mass and acoustic properties (sound transmission, mass echogenicity), etc.

Chen et al. [66] proposed an alternative CAD system, which was tested on a US database containing 140 pathologically proved tumours, i.e., 88 benign and 52 carcinomas. Initially the ROI is manually extracted by the physician. The texture information of the ROI is subsequently extracted, and a neural network classifier with autocorrelation features is used to classify the tumour. The accuracy of the proposed system was 95.0%, with sensitivity 98%, specificity 93%, positive predictive value 89%, and negative predictive value 99%.

Chen et al. [67] studied the texture analysis of breast tumours on sonograms. A total of 1020 images from 255 patients were used in the experiment. The ROI images were initially identified by the physician. Subsequently, a neural network model, using 24 autocorrelation texture features, was used in the classification of tumours. The area under the ROC curve (A_z) for the model was reported to be 0.9840±0.0072.

In 2002 Chen et al. [68] proposed a further novel CAD system. Initially the ROI is manually selected by the physician. This is followed by a segmentation algorithm based on wavelet transform. Three features, variance contrast, autocorrelation contrast, and distribution distortion of wavelet coefficients were extracted from the ROI images. These features are subsequently used in a multilayered perceptron (MLP) neural

network, which is trained by a error back-propagation algorithm with momentum. The authors reported that A_z for the proposed system was 0.9396±0.0183, with sensitivity is 98.77%, specificity is 81.37%, positive prediction value is 72.73%, and negative predictive value is 99.24%.

Horsch et al. [70] presented a CAD method for the detection of breast lesions in ultrasound images that is based on the automatic segmentation of lesions and the automatic extraction of four features related to lesion shape, margin, texture, and posterior acoustic behaviour. The database used consisted of 400 cases (94 malignant lesions, 124 complex cysts, and 182 benign solid lesions). The use of linear discriminate analysis as a classification method of breast lesions was investigated. An average A_z value of 0.87 was reported in the task of distinguishing malignant from benign lesions.

Drukker et al. (2002) [14] used a radial gradient index (RGI) filtering technique to automatically detect lesions of a breast ultrasound image, and segmented the lesion candidates by maximizing an average radial gradient (ARD) index for regions grown from the detected points. Testing on a large database of 757 images it was reported that at an overlap of 0.4 with a lesion outline identified by an expert radiologist, 75% of the lesions were correctly detected. Bayesian neural network was used in classification stage and the quality was assessed by a round robin analysis, and yielded an A_z value of 0.84, with 94% sensitivity at 0.48 false-positive per image.

Chen et al. [53] proposed a CAD system by setting independent features (i.e., features extraction) and using a ANN. The database contained 2 sets of images, first set with

160 images, and second set with 111 images. Seven morphological features were used, and a multilayer feed-forward neural network was used as the classifier. The A_z reported for the ROC curve was 0.950±0.005, overall.

Joo et al. [16] proposed a CAD algorithm by using multiple ultrasound features and a ANN. The test database consisted 584 histological confirmed cases. The ROIs of images were manually selected by expert radiologists. Median filtering was initially applied to the ROI images and a simple segmentation consisting of contrast enhancement and thresholding segmentation, were subsequently applied. The reported accuracy of the system was 91.4%, sensitivity: 92.3% (131/142), and specificity: 90.7% (136/150). Further experimental results were reported during which the cut-off level of the thresholding segmentation was adjusted. It was shown that this resulted in an increase in accuracy (99.3% (141/142) and 100% (142/142)) but resulted in a decrease of sensitivity (53.3% (136/150) and 7.3% (11/150) respectively). In the above work, even though a simple segmentation approach was adopted, high level of classification accuracy was achieved. The errors in classification were thought to be caused mainly due to the inaccuracies of edge detection rather than due to the ANN.

Chen et al. [45] proposed a CAD system based on fractal features. The database contained 110 malignant tumours and 140 benign tumours. The ROI was selected by the authors. The images are pre-processed by using morphological operation and histogram equalization. The fractal dimension of the ROI images was calculated by using the fractal Brownian motion, and the tumours were classified by *k*-means classification methods. The reported ROC area index A_z was 0.9218.

Chang et al. [12] proposed a CAD system with automated ultrasound segmentation and morphology based diagnosis of solid breast tumours. The database consisted of 90 malignant tumours and 120 benign tumours. The ROIs were pre-selected by a radiologist. Anisotropic diffusion filtering is used to pre-process the images, followed by the level set method of segmentation introduced by the authors. Support Vector Machine (SVM) based on six morphological features is used to classify the tumours. They reported results were: accuracy 90.95%, sensitivity 88.89%, specificity 92.5%, positive predictive value 89.89%, and negative predictive value 91.74%.

Huang et al. 2005 [72] evaluated two pathologically proven ultrasonic databases, DB1, which contained 140 images, and DB2, which contained 250 images. The ROIs of the images were selected by a physician. Textural features were used in SVM based classification of tumours. The ROC indexes reported for for DB1 and DB2 are 0.9695±0.0150 and 0.9552±0.0161 respectively.

In the most recent work in classification of lesions, i.e., Huang et al. 2008 [13], 118 breast lesions were evaluated out of which 34 were malignant and 84 were benign. The ROI were manually selected by a physician. 19 morphological features from the extracted contour were obtained and PCA was used to find the independent features. A SVM classifier was finally used in the classification. The reported A_z value when using all morphological features and then when using only the lower-dimensional principal vector were, 0.91 and 0.90, respectively.

3.4 Mammography

Mammography is a specific type of imaging that uses a low-dose x-ray system as the source [75]. It plays an important role in early detection of breast cancers as it can show changes in the breast up to two years before it is palpable. As a result there is an increasing interest by governments, radiologists and medical service providers in using mammography in breast screening programmes [76].

3.4.1 Mammographic CAD Systems

Due to the specific nature of mammographic images, a number of special image processing techniques have been developed in the past, specifically applicable to such images [77]. They include: Manual intensity windowing (MIW), histogram-based intensity windowing (HIW), mixture-model intensity windowing (MMIW), contrast-limited adaptive histogram equalization (CLAHE), unsharp masking, peripheral equalization, Trex processing, MUSICA [77] (a Multiscale wavelet-based contrast enhancement technique developed by Agfa).

CAD techniques in medical imaging are developed for the automated differentiation between benign and malignant lesions and go beyond computer-aided detection by providing a measure of the likelihood of cancer for a detected lesion, given image and/or patient characteristics [78].

Table 3-2 Summary of state-of-the-art mammographic CAD systems

Method Used	Classification features	Algorithm	Performance	Comment
Artificial Neural Networks (ANN) [78]	Descriptors of individual character. Distribution of the cluster.	Automated detection, segmentation and classification steps (wavelet filters and ANN).	Sensitivity: 100%. Specificity: 85%. A_z = 0.98±0.01.	Insensitive to segmentation and detection errors (FP signals).
Generalized Dynamic Fuzzy Neural Networks (GDFNN). [79]	Texture parameters derived from 1st order gradient distribution and gray-level co-occurrence matrices, computed from the ROI.	GDFNN algorithm. DB: DDSM. 180 benign masses. 163 malignant masses.	TP fraction: 95%. FP fraction: 52.8%. A_z = 0.868±0.020. Accuracy: 70%.	Novelty: alleviates the problem of acquiring a designer to examine all the IO relationship of a training db.
Knowledge-based Approach [80]	A learning process to establish a knowledge base that is then used to determine whether a previously identified suspicious region is likely to depict a true mass.	Quantitatively characterizing the set of known masses, based on all "known" masses to determine the state of the suspicious region.	A_z = 0.83. 51% of previous identified FP regions were eliminated. 90% sensitivity maintained.	Significant reduction in FP detections. Maintains reasonable sensitivity. A potentially good approach.

Over the last two decades, CAD systems have been developed to aid radiologists in detecting mammographic abnormalities suspicious of being breast cancers. A typical mammographic CAD system uses computer software to assist the film reader in identifying abnormalities on a mammogram; this is done by placing prompts over areas of concern.

There are two commercially available mammographic CAD systems [32] at present:

1. R2: "Image checker", R2 Technology, Los Altos, California, USA
2. CADx: "Second Look", CADx Medical Systems, Quebec, Canada.

3.4.2 Current Trends in Mammographic R&D and Applications

There are currently 127 Picture Archiving and Communications Systems (PACS) in England; which have been employed in over 25.5 million patient studies generating some 640 million stored images. Traditionally breast screening has been outside such PACS. However at present a number of novel approaches are being investigated to integrate breast screening into hospital radiology information systems (RIS) and PACS. Intelligent tutoring systems play an important role is this respect.

In the UK, all screening personnel involved in the NHS Breast Screening Programme undertake PERFORMS, an annual self assessment scheme [81]. Twice a year 560 individuals interpret carefully selected sets of recent difficult screening cases. In doing this they identify certain key mammographic features, as well as defining their location, and rates each case on whether they consider it to be normal, benign or malignant. Participants gain immediate feedback on their decisions as well as

subsequently being able to see how they have fared as compared anonymously to all other screeners. An individual can rapidly gain insight into the level of their mammographic interpretation skills. Importantly, their skill in correctly locating and identifying key mammographic features, represents their detection ability and the overall decision of the case. The correct identification of the key mammographic features, indicates their interpretation skill. Knowledge of such detection and interpretation abilities can then be used as an input to inform the subsequent structure of a training scheme for each person whilst still maintaining the anonymity of each individual's performance data.

Ideally any tutoring system requires a large cohort of suitable mammographic images which can be used in different ways to aid learning. Such a large database of high resolution cases can be difficult both from a storage point of view (if stored at the point of training delivery) and also from a logistical point of view – as different breast screening and university research departments develop their own collections of images for differing purposes [82]. However, the 'Grid' concept greatly facilitates such a large endeavour by allowing images both to exist, and be harvested, via different databases in different geographical locations without the need for the full dataset and user to coexist at the same physical location. [The Grid is defined as "flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources" [83]]. Grid technology enables that the medical community can explore collaborative approaches for managing image data and exchanging knowledge [84]. The EU-funded MammoGrid and UK-funded eDiamond projects are to achieve the above aims [85]. In breast cancer research, MammoGrid project has recently delivered its first proof-of-concept prototype

enabling clinicians to store (digital) mammographic images along with appropriately anonymised patient meta-data and to provide controlled access to mammograms stored both locally and remotely [84].

3.5 Summary & Conclusion

The aim of this chapter was to critically review existing approaches to Ultrasound CAD. Apart from analyzing approaches which provided an end-to-end solution for CAD, the chapter also reviews existing research on particular stages of the CAD systems, namely pre-processing for noise reduction, initial lesion detection etc.

The chapter initially provided details about the presence of speckle noise in ultrasound images, which has motivated a number of research projects in the past to investigate the de-speckling of US images before processing within CAD systems. Continuing with these efforts in this thesis we propose a hybrid filtering approach (see Chapter 4) for noise reduction in ultrasound images, which is experimentally proved to perform better than existing de-speckling algorithms.

The detailed review presented in this chapter concludes that in most US CAD systems the labelling of the initial ROI is done manually, by a radiologist. Only a single effort has been made to automate the initial lesion detection process. However this approach has known shortcomings. Therefore in Chapter-4 we propose a novel algorithm for fully automated initial lesion detection.

Further, the review particularly emphasised the problem of segmentation in ultrasound breast images. It was pointed out that due to the inaccuracies in segmentation, the overall classification accuracy, sensitivity, specificity, and A_z metrics of the CAD systems can suffer significantly. Although a number of segmentation algorithms have been proposed in literature, developing an efficient algorithm still remains an open research problem. Towards providing a solution to this problem, in Chapter-5 we propose a lesion segmentation approach based on multifractal analysis and Isotropic Gaussian processing.

Classification algorithms also play a key role in US CAD systems. A number of approaches for classification of lesions have been investigated in literature. However the determining the best classifier for a given CAD system not only depends on the general accuracy of the classifier selected, but also on the nature of the dataset (e.g. US, mammographic) and the feature set used in classification. To this effect in the classification of US images there is a significant research gap in identifying the best feature set – classifier combinations that can directly impact in improving the accuracy, sensitivity and specificity of CAD systems. Addressing this issue chapter-6 presents a detailed study of using different feature sets, feature selection approaches and classifiers in identifying the best feature set – classifier combinations for US CAD systems.

The literature review presented in this chapter also shows that despite proposing number of approaches for the CAD of US images, no study has investigated the effects of computer processed images when used as an aid in human processing of US images. Despite the presence of automated CAD systems, human intervention of these

systems at various key stages of the image processing pipeline (e.g. initial lesion detection) is often envisaged. Therefore in Chapter-7 we present a study that was carried out as a part of this research to investigate the effects of processed images on human perception of US images.

In general breast cancer diagnosis systems are moving towards digital and grid-enabled systems. Further multi-modal systems are widely considered to be more accurate and are therefore better trusted. Mammography CAD systems enhanced by the co-existence of US CAD systems can demonstrate better accuracy and robustness against specific conditions under which mammography alone may perform sub optimally. Hence further development of the state of the art methods adopted within US CAD systems is of significant importance.

Chapter 4: Initial Lesions Detection

4.1 Introduction

This chapter introduces and evaluates two novel methods for initial lesion detection. It further introduces and analyses Drukker et al.'s [14] algorithm for initial lesion detection, which is subsequently used as a benchmark to evaluate the performance of the algorithms proposed.

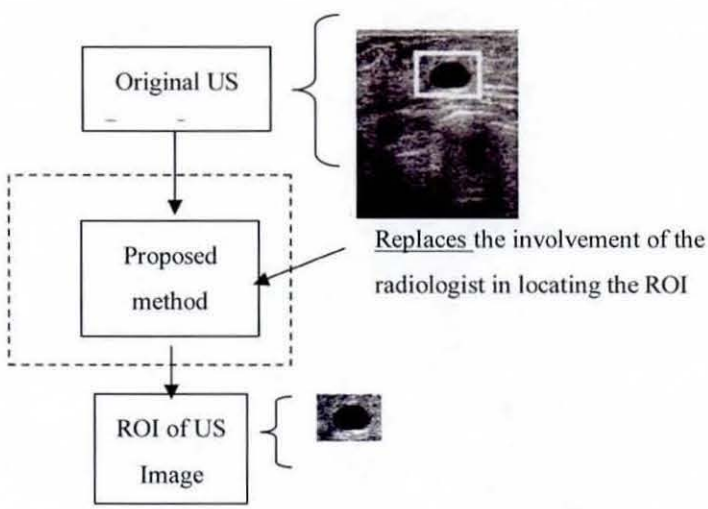


Figure 4-1 Illustration of the use of the proposed solution

4.2 Research Motivation

In most existing breast screening approaches, the so-called 'initial lesion' (i.e. the suspected region) is manually located by a trained radiologist in a pre-processing stage by marking its topmost, leftmost, bottommost and rightmost boundary limits with crosses. These crosses (and hence the initial lesion) are then manually encompassed within a rectangular Region of Interest (ROI) [42] by the radiologist. The ROI is subsequently presented to a CAD system for further analysis leading to the segmentation and classification of the tumour. Therefore both the selection of extreme points (i.e. crosses) and the rectangular region, require human intervention and are open to subjectivity and human error. As a result a well trained and experienced examiner with knowledge of the normal echo anatomy of the breast and the changes caused by the pathology is required for accurate breast cancer screening. The non-inclusion of the entire lesion within the ROI, missing ROI's containing lesions etc., can severely undermine the performance of a CAD system. Our present research focus is to provide the radiologist with an automated tool that can effectively assist in the selection of the ROI. It is worthwhile noting that the automatic detection of ROI is not meant to replace the radiologist, but is to provide a tool to reduce the region labelling time of the radiologist and to warn of possible ROI's that could be otherwise missed due to the poor quality of the ultrasound image. It has been further shown that radiologists with different training backgrounds and experiences often get different results in the reading of sonograms [11]. This provides further justification to the idea of researching into novel techniques that assist physicians in improving the consistency of interpretation of ultrasound images.

4.3 Measurement

The evaluation of the performance of the proposed algorithm requires a suitable test image database, an evaluation metric and a design goal. Due to the practical difficulties in obtaining databases with ultrasound images of normal and nearly normal breasts, current state-of-the art algorithms [14] of initial lesion detection have used US image databases that solely consist of malignant and benign tumours. In the proposed experiments we have used a US image database of 360 malignant and non-malignant images. This database additionally consists of location information of the extreme points (in the form of marked, 'crosses') of tumours, marked by a number of expert radiologists which is useful for evaluation purposes. The metric used to evaluate the performance of the algorithm on such databases is an 'overlap' figure [14] defined as the ratio of the intersection and the union of the two lesion areas that have been manually identified by the radiologists and the computer based algorithm. Specifically in Drukker et al.'s work achieving an overlap value of in excess of 0.4, has been used as the design goal. This can be represented as:

$$overlap = \frac{X \cap Y}{X \cup Y} \geq 0.4 \qquad (4\text{-}1)$$

Where X is the lesion area extracted by the computer based algorithm, and Y is the lesion labelled manually by the radiologist. In other words it has been assumed that if an overlap of more than or equal to 0.4 results, a computer based algorithm has been successful in accurately and automatically performing the otherwise manual task of initial lesion identification. Therefore a secondary metric of 'accuracy' can be defined as the percentage of experiments obtaining an overlap value of beyond 0.4. Within the context of the proposed research we use 'accuracy' as the objective metric to evaluate

and compare the results of the performance of the proposed algorithm with that of the state of art methods. Further, subjective results have been illustrated for visual comparison.

4.4 Benchmark Algorithm

The state-of-the-art technique of automatic ROI labelling of US breast images, used as benchmark to compare the performance of the proposed technique, can be summarised as follow:

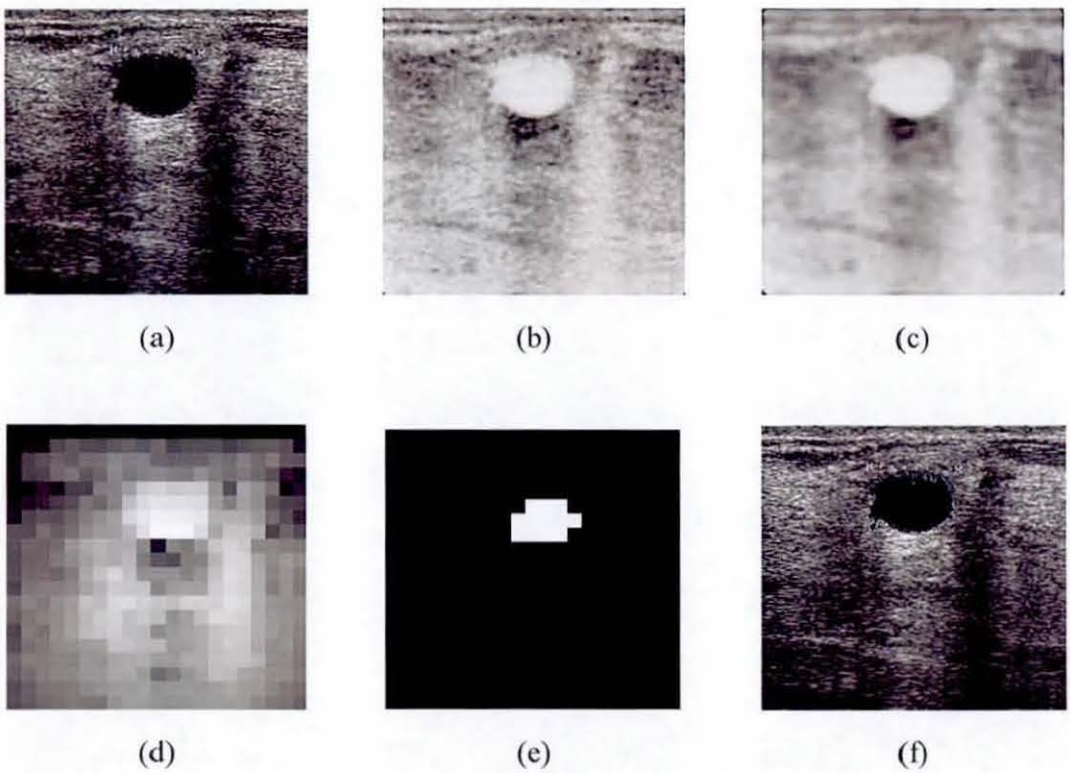


Figure 4-2 Illustration of the operation of the intermediate stages of Drukker et al.'s algorithm (a) original image, (b) gray-scale inverted image, (c) median filtered image, (d) RGI filtered image, (e) thresholded RGI image, (f) final detection.

In [14] Drukker et al. investigated the use of the radial gradient index (RGI) filtering technique to automatically detect lesions on breast ultrasound images (see Figure 4-2). It is common that in ultrasound images, lesions are almost invariably darker than the background. Thus in Drukker's work the gray-scale of the original ultrasound images are initially inverted. Subsequently images are pre-processed by a median filter to remove speckle noise and the resulting image is fed to a RGI filter. In RGI-filtering, the images are sub-sampled by a factor of 4. The threshold for the RGI-filtered image is varied iteratively from 0.74 to 0.62, until at least one lesion of interest is detected. The detected areas, smaller than $5mm^2$, are discarded. Lesion candidates are segmented from the background by maximizing an Average Radial Gradient (ARD) index [14] for regions grown from the detected points. According to Drukker et al., maximizing the ARD index is more accurate than maximizing the RGI index. At an overlap level of 0.4 with lesions outlined by a radiologist, 75% accuracy of lesion detection was reported.

4.5 Datasets

It should be noted that any automatic system that is designed to detect abnormal lesions in ultrasound images should be finally verified/compared with the judgment of a medical expert/radiologist. The test images used in this thesis are obtained from a professionally compiled Breast Ultrasound CD [22], which consists of explanations and verifications from several qualified expert radiologists. A total of 360 images from the Breast Ultrasound CD were selected for the experiments. Of the 360 images, 20 were malignant, 76 were simple cysts, 76 were complex cysts, 58 were fibroadenoma, 38 were carcinoma, 18 were occult lesions, 15 were adenosis, and 59

were combination of other diagnosis. Each image has been manually processed by an expert radiologist and the extreme points of the suspected lesions have been marked with 'crosses'.

4.6 Pre-processing

It was mentioned previously that the credibility of a high quality breast ultrasound examination depends on the scanner, the quality of the original US image and the experience of the examiner. The pre-processing stage deals with the issue of guaranteeing the homogeneity of the original ultrasound images, thus improving the chance of more accurate, subsequent, lesion ROI detection. In the proposed approaches, histogram equalization and hybrid filtering is used as pre-processing stages to achieve the above.

4.6.1 Histogram Equalisation

Histogram equalisation [86] is similar to contrast stretching in that it attempts to increase the dynamic range of the pixel values in an image. However unlike contrast stretching, there is no provision for interactivity, as applying histogram equalisation algorithm to an image, with a fixed number of bins, will always yield the same result. The ultrasound images are in 8-bit greyscale format. Let i be the gray level, and n_i be the number of occurrence of gray level i. The probability of an occurrence of a pixel of level i in the image, x, is by:

$$p_i(x) = \frac{n_i}{n} \qquad for\ i \in 0,....,L-1 \tag{4-2}$$

where L is the total number of grey levels in the image ($L=256$ in this case), n is the total number of pixels in the image. Let *cdf* be the cumulative distribution function corresponding to p, *cdf* is defined by:

$$cdf_i(x) = \sum_{j=0}^{i} p_j(x) \tag{4-3}$$

Let x be a 3×3 8 bit grey scale image has the following value:

$$x = \begin{bmatrix} 142 & 76 & 80 \\ 63 & 50 & 100 \\ 76 & 100 & 142 \end{bmatrix}$$

The probability of an occurrence of the grey levels in image x is shown by table 4-1. Please note that grey level with count 0 are excluded for brevity.

Table 4-1. The probability of an occurrence of grey level.

grey level, i	$p_i(x)$	grey level, i	$p_i(x)$	grey level, i	$p_i(x)$
50	$\frac{1}{9}$	63	$\frac{1}{9}$	76	$\frac{2}{9}$
80	$\frac{1}{9}$	100	$\frac{2}{9}$	142	$\frac{2}{9}$

The cumulative distribution function (*cdf*) is calculated, as in table 4-2.

Table 4-2. The cumulative distribution function (*cdf*)

grey level, i	$cdf_i(x)$	grey level, i	$cdf_i(x)$	grey level, i	$cdf_i(x)$
50	$\frac{1}{9}$	63	$\frac{2}{9}$	76	$\frac{4}{9}$
80	$\frac{5}{9}$	100	$\frac{7}{9}$	142	$\frac{9}{9}$

Let $cdf_{min}(x)$ be the *cdf* with minimum grey level, i.e., $\frac{1}{9}$ for this case. And $cdf_{max}(x)$ be the *cdf* with maximum grey level, i.e., 1. To get the corresponding output image y for grey level i, the general histogram equalisation formula is defined by:

$$y_i(x) = round\left(\frac{cdf_i(x) - cdf_{\min}(x)}{cdf_{\max}(x) - cdf_{\min}(x)} \times (L-1)\right) \tag{4-4}$$

For instance, $i=76$

$$y_{76}(x) = round\left(\frac{\frac{4}{9} - \frac{1}{9}}{1 - \frac{1}{9}} \times (256-1)\right) = 96$$

By repeating for all grey level, i, the output image:

$$y = \begin{bmatrix} 255 & 96 & 128 \\ 32 & 0 & 191 \\ 96 & 100 & 255 \end{bmatrix}$$

4.6.2 Hybrid Filtering

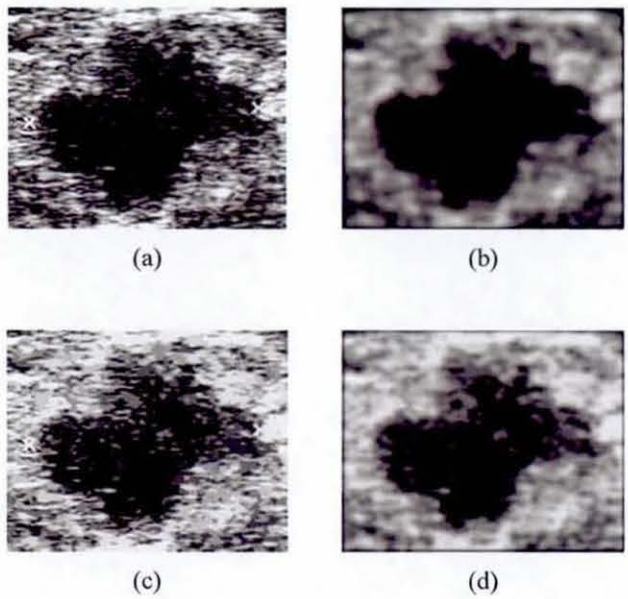
Filtering stage is used to remove noise, which is a major obstacle for accurate segmentation of US images. The fundamental requirements of noise filtering methods for medical images are: safeguarding important information of the object boundaries and detailed structures, ability to efficiently remove noise in the homogeneous regions and the ability to enhance morphological definitions by sharpening discontinuities [12]. It was discussed that generally the quality of ultrasound images is poor, as a result of multiplicative speckle noise and artefacts that cause information loss. There are many noise filtering methods, which can be broadly categorized into, linear diffusion filtering and non-linear diffusion filtering. Linear diffusion filtering, e.g. the Gaussian filter, can remove the noise efficiently, but at the same time, the semantically useful information can be eliminated. In addition to this, linear diffusion filtering dislocates edges when moving from finer to coarser scales [44] of image representation.

In [47], Perona and Malik proposed a Non-linear Partial Differential Equation (PDE) approach for smoothing images on a continuous domain. It was shown that *Anisotropic Diffusion* performs well for images corrupted by addictive noise. However in cases where images contain speckle noise, anisotropic diffusion actually enhances the speckle noise, instead of eliminating [48]. Despite this fact anisotropic diffusion techniques are used for filtering ultrasound images [12].

On a more positive note, *Nonlinear Diffusion* filtering [44] has attracted much attention in the field of image processing for its ability to reduce noise while preserving (or even enhancing) important features of the image, such as edges or

discontinuities as opposed to *linear diffusion filtering* (e.g., Gaussian filtering, linear scale-space representation) which not only removes noise but also blurs and dislocates edges [43].

In order to take advantages of both, Nonlinear and Linear Diffusion Filtering approaches we propose the use of a *Hybrid* filtering approach as a pre-processing stage. *Hybrid* filtering combines the key strength of *Nonlinear Diffusion* filtering, i.e. its ability to produce edge-sensitive speckle reduction, with linear filtering (*Gaussian blur*), i.e. its ability to smoothen the edges, and eliminate over segmentation.



(a) (b)

(c) (d)

Figure 4-3. (a) original image (b) Gaussian blur (c) Nonlinear diffusion filtering. (d) hybrid filtering.

Figure 4-3 compares the three filtering approaches when applied to lesion boundary detection of a complexly shaped lesion. It is clear that the hybrid filtered image has sharper, better defined edges that will increases its chances of more accurately being detected at the latter processing stages.

The above pre-processing stage has been used in both of the proposed approaches to initial lesion detection. In the following sections of this chapter present details of the subsequent processing stages of the two approaches and evaluate their performance in comparison to the benchmark algorithm (see Section 4.4).

4.7 Method-1

Figure 4-4 illustrates the block diagram of the first approach adopted. The design and operational details of the relevant stages are detailed below. Note that the ultrasound images are first pre-processed using Hybrid Filtering and subsequently enters the segmentation stages.

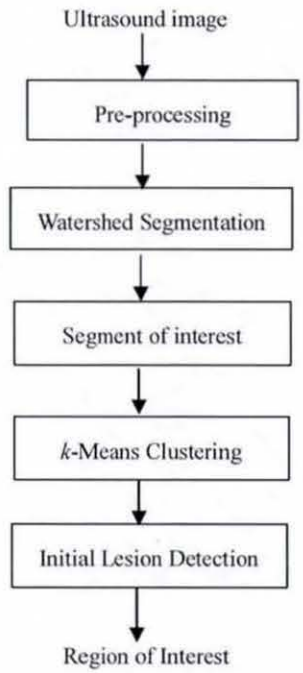


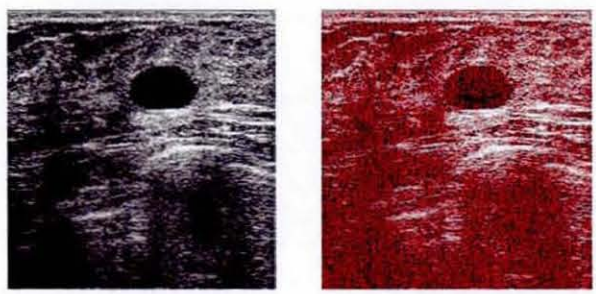
Figure 4-4 Overview of the methodology.

The individual stages of the approach proposed can be detailed as follows:

4.7.1 Watershed Segmentation

In general segmentation is a process used to divide an image into its constituent parts. The watershed transform is a well established tool for the segmentation of images. Instead of using the image directly, the transform uses a gradient image extracted from the original image for segmentation [87]. [Note: In geography, watershed is the ridge that divides areas drained by different river systems. A catchment basin is the geographical area draining into a river or reservoir. The watershed transform applies these ideas to gray-scale image processing in a way that can be used to solve a variety of image segmentation problems [88]]

Direct application of the watershed transform to greyscale images usually leads to oversegmentation due to noise and other local irregularities [88], as illustrated in Figure 4-5 (note: lines in red illustrate the segment boundaries).

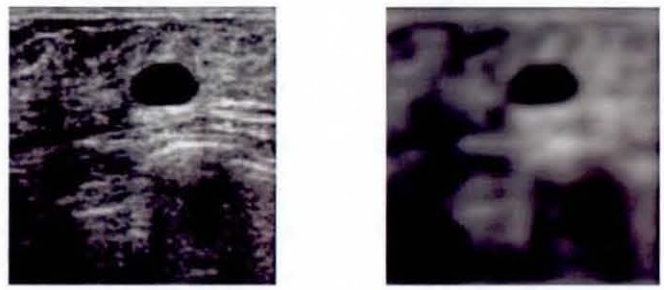


(a) (b)

Figure 4-5 Illustration of over segmentation (a) original image (b) Over segmented image

A solution is to apply a noise reduction algorithm before segmentation. Speckle noise

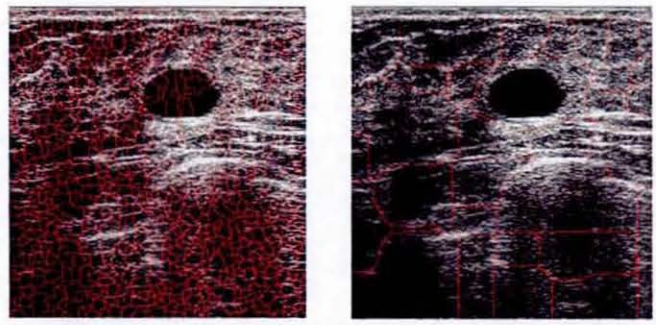
and artefacts can be removed by implementing Gaussian blurring [89] to ultrasound images (see Figure 4-6). In figure 4-6, it is illustrated that the image on the left, with higher scale of blur, (i.e. higher *sigma,* σ), yields a smoother image.



(a) (b)

Figure 4-6 Some speckle noise and artefacts are removed after applying Gaussian Blur, with (a) low scale of blur, σ=2; and (b) high scale of blur, σ=8. Note that image (b) consists of less noise, and is smoother than image (a).

Results illustrated in Figure 4-7 (b) shows that the number of segments will be reduced when watershed segmentation is applied on the blurred image. Note that in Figure 4-7 (a), the region of interest is segmented into smaller segments making it hard for the subsequent stages of processing.



(a) (b)

Figure 4-7. Gaussian blur reduces the number of segments. Image in (a) shows the result of watershed segmentation on an image, obtained with a low scale of blur, σ=2, overlapped with the original image. Image in (b) shows the result of watershed segmentation on an image, obtained with a high scale of blur, σ=8, overlapped with the original image.

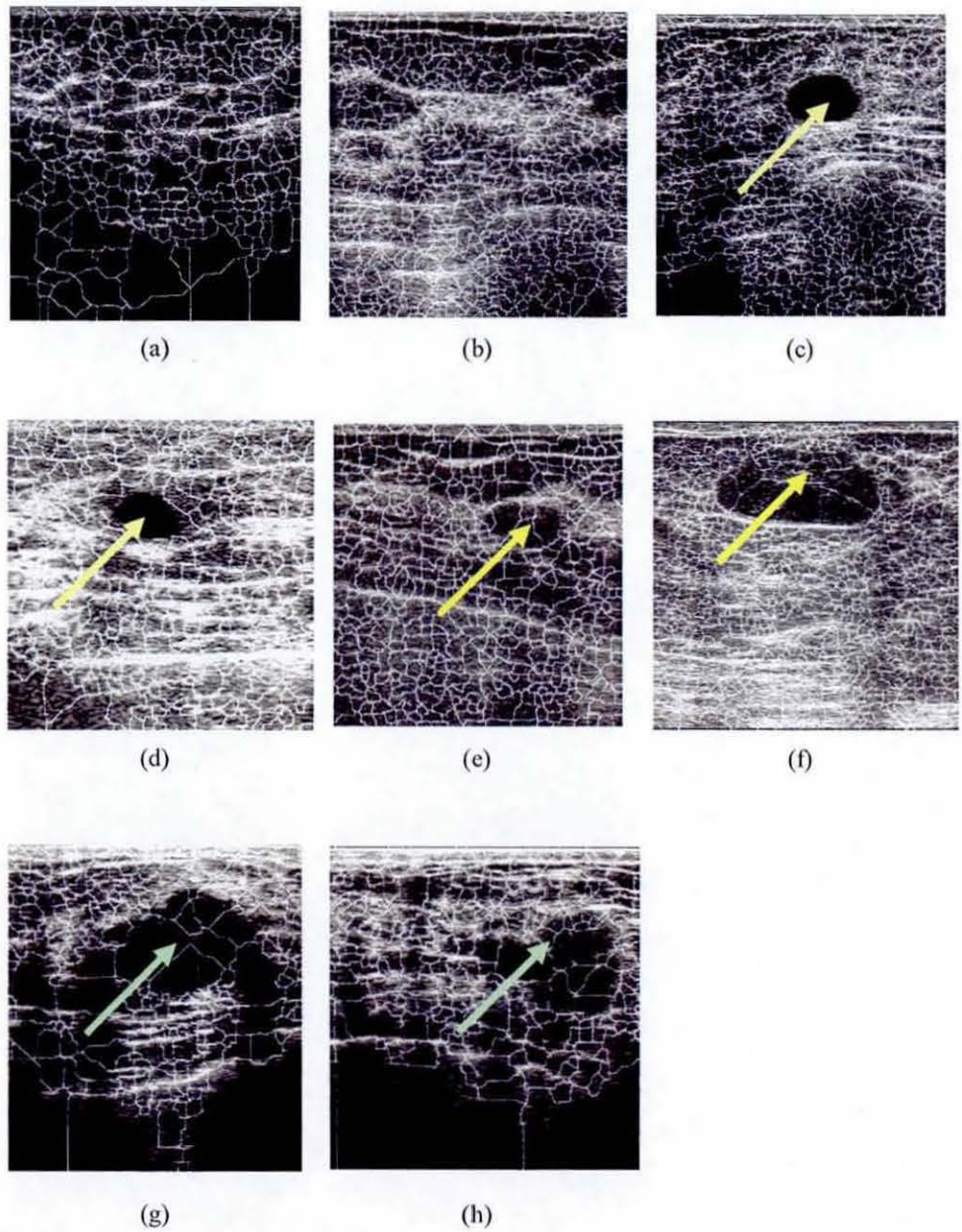
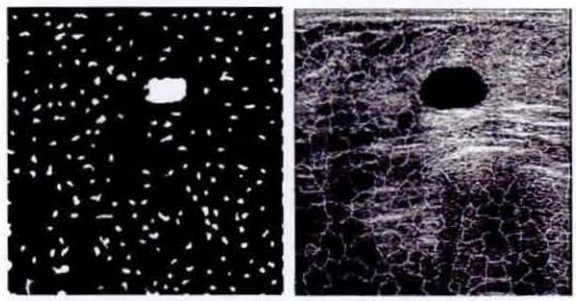


Figure 4-8 Application of watershed segmentation to different types of lesions. Observe that there are differences between segmentation patterns for different types of lesions. For 'normal' lesions, the distribution of the segments are generally even, as illustrated in (a) and (b). For others types of lesions, the suspected regions consist of less segments. (c) and (d) are the images with cysts, where no segmentation has occurred within the ROI (areas pointed by the arrows). For images with fibroadenoma ((e) and (f)) and malignant ((g) and (h)) lesions, the number of segments in the suspected region is less than (a) in an image with a normal lesion (area pointed by the arrows).

Experiments were carried out on four different types of lesion categories diagnosed by the expert radiologists, namely, normal, cyst, fibroadenoma, and malignant. It is expected that there are some differences in the number of segments for different diagnosis. The results are illustrated in Figure 4-8. It is seen that comparatively, watershed segmentation can give better results in segmenting cysts, when compared to the segmentations of other lesion types. This is expected as cysts consist of clear, more accurately defined boundaries of simpler, rounded shapes.

A more efficient approach used to control oversegmentation is based on the concept of markers, namely, Marker-Controlled Watershed Segmentation [88]. The result of using marker controlled watershed segmentation on ultrasound images is illustrated in Figure 4-9. It clearly shows that the approach is better than simple watershed segmentation. Hence in the proposed approach we use marker controlled watershed segmentation for the purpose of segmenting the ROI.



(a) (b)

Figure 4-9 (a) Markers. (b) Results of using Marker-controlled watershed segmentation.

4.7.2 Segment of Interest

The location of the Region of Interest (ROI) of an abnormal lesion is crucial. The location of the abnormal lesions requires the specification of both position and orientation (assuming 2D ultrasound images). Within the context of current research, two methods of analysis have been used to define the segment of interest, namely, statistical analysis and fractal analysis.

4.7.2.1 Statistical Analysis

In statistical analysis, the suspected lesions, both of benign and malignant type, are compared to the normal lesions. The local mean of each segment, centred at its centre of gravity (cog) and with radius of 9 (for example) is calculated. The segment with the minimum local mean (i.e. the darkest segment), is considered the segment of interest.

4.7.2.2 Fractal Analysis

Fractal analysis has been investigated in many application domains, such as, medical, satellite and texture image analysis [90]. In the proposed approach, a modified fractal dimension (D_m) calculation based on a dimension normalization technique is carried out to extract the information of each segment in ultrasound breast images. In the proposed modification, the fractal dimension, N, is first normalized by the surface area covered by the window size, $L\,unit \times L\,unit$. The normalized N_n can then be related to maximum gray level H by the Power Law [91]:

$$N_n = \frac{N}{L^2} = H^f \qquad (4\text{-}5)$$

where f is the partial fractal surface dimension, and from equation 4-5, logarithm of equation 4-5 yields:

$$f = \frac{\log(N_n)}{\log(H)} \qquad (4\text{-}6)$$

In the case of segments, since all the segments consist of uneven areas, the surface area, L^2, is substituted by the total area of the segment. The modified fractal surface dimension, D_m is obtained by adding the physical dimension (2 for 2-D ultrasound image) to the partial fractal surface dimension, f.

$$D_m = 2 + f \qquad (4\text{-}7)$$

Note that the modified fractal dimension will enable the differentiation between smooth and rough textured areas. In fractal analysis, the normal lesions have rougher surfaces as compared to the abnormal lesions. The rough surfaces have higher value of D_m, while smooth surfaces have low value of D_m. Hence, in detecting the segment of interest, the segment with minimum D_m is selected.

4.7.3 *k*-means clustering

The basic idea of clustering is to group similar objects together. More formally, clusters are connected regions of a multi-dimensional space containing a relatively high density of points, separated from other such regions by a region containing a relatively low density of points [92]. Clustering problems arise in many different applications, such as data mining and knowledge discovery, data compression, vector quantization, pattern recognition and pattern classification [93]. Among clustering formulations that are based on minimizing a formal objective function, perhaps the most widely used and studied is *k*-means clustering [93].

k-means clustering is an unsupervised machine learning algorithm. The objective of *k*-means clustering is to minimize the sum of distances to the nearest centre and the geometric *k*-centre problem, and to minimize the maximum distance from every point to its closest centre. *k*-means clustering aims at minimizing an objective function [94], in this case a squared error function *J*, where:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (4\text{-}8)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is the distance measure between the data point $x_i^{(j)}$ and cluster centre c_j.

The algorithm is explained in [94] with the following steps:

1. Make initial guesses for means(cluster centres) in k clusters
2. Check the distance of the data points to the cluster centres, assign each data point to the cluster that has the minimum distance.
3. When all the data points has been assigned, recalculate the means
4. Repeat steps 2 and 3 until there are no changes in any mean.

In order to capture the complete ROI, the similar segments are clustered as a group. Thus k-means clustering is implemented, and the segments with a low value of local mean (section 4.7.2.1) or a minimum of fractal value (section 4.7.2.2) are grouped under the same cluster as the Point-of-Interest. In our experiments, $k=5$ is used, which produced the best results, compared to other values tested. LTI-LIB [95], a computer vision library, is used as the clustering program.

4.7.4 Results and Analysis

Figure 4-10 illustrates the results obtained at various intermediate stages of the proposed approach to initial lesion detection. The graphs in figure 4-11 plots the local mean and the fractal dimension, D_m of each segmented region. The graphs indicate that both metrics identify that segment no. 50 is the segment of interest.

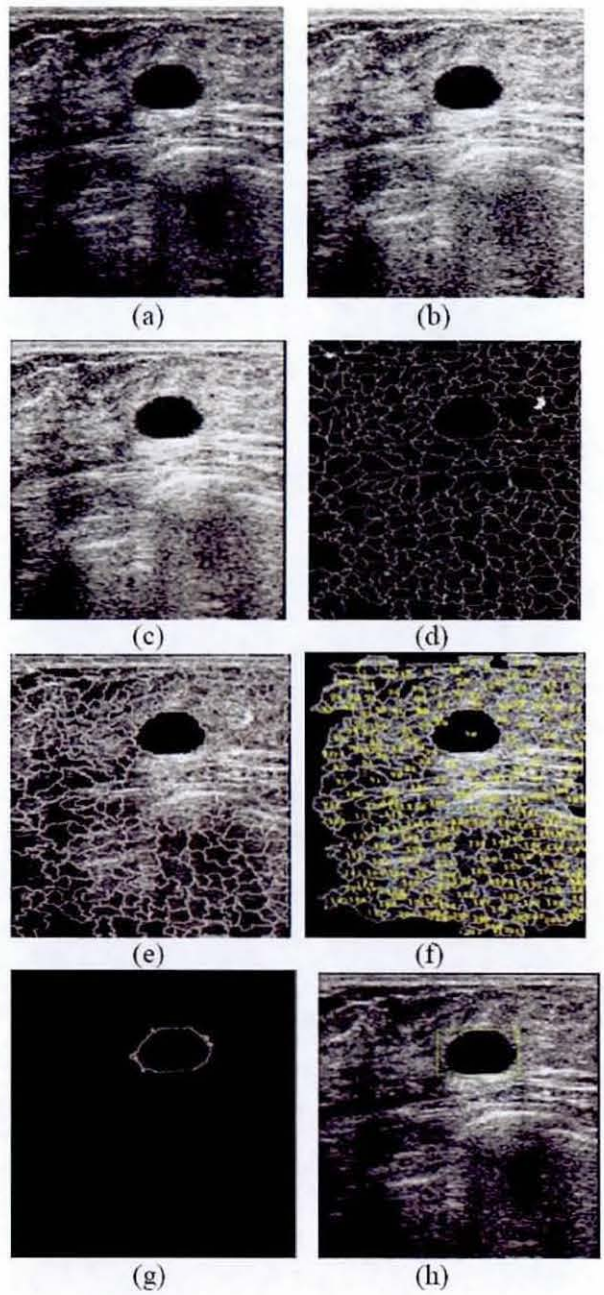
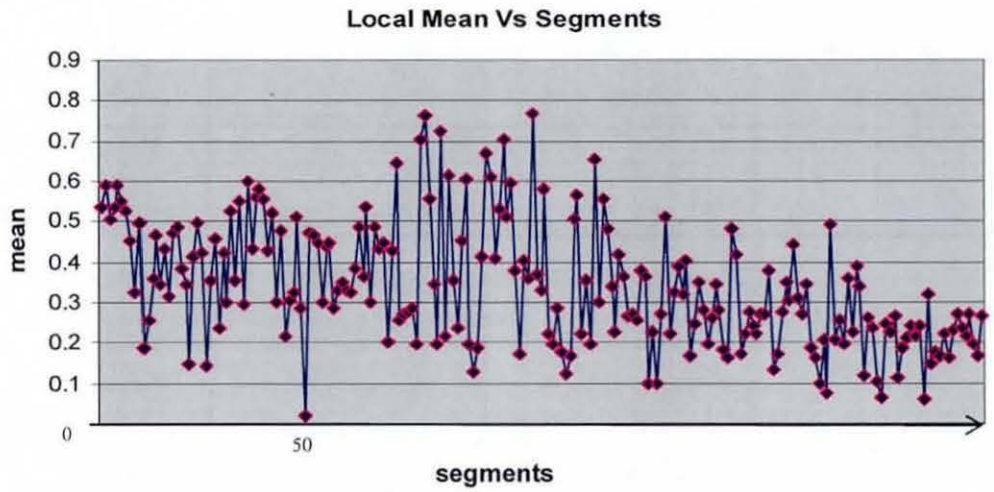
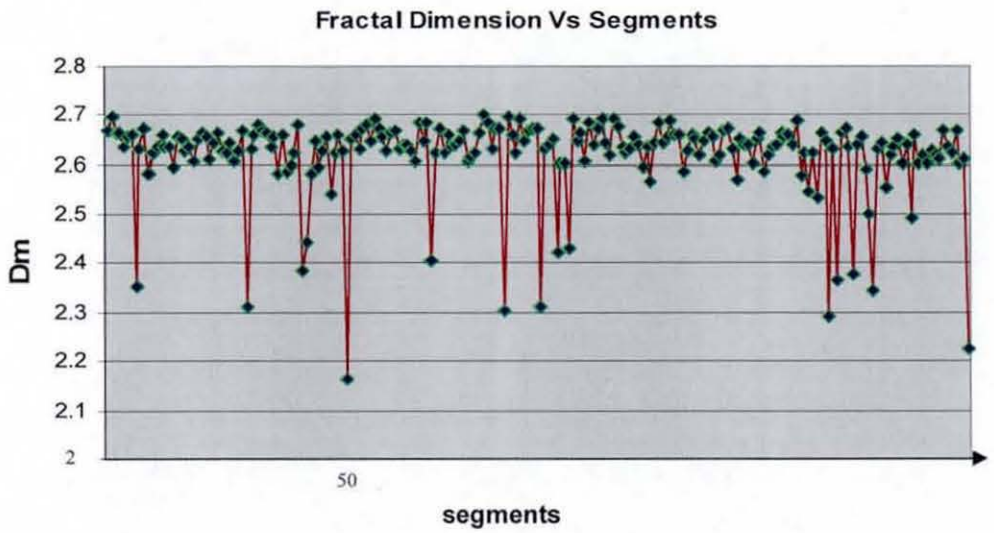


Figure 4-10 (a) original image. After; (b) histogram equalization, (c) nonlinear diffusion filtering, (d) watershed segmentation, (e) mapping of watershed segmented image with the original image, (f) labelling of segments, (g) location of the segment of interest, (h) automatic labelling of the region of interest using a rectangle.



(a)



(b)

Figure 4-11 (a) the plot of minimum local mean, showing that segment number 50 is the point of interest. (b) the plot of minimum D_m, confirming that segment number 50 is the point of interest.

Using the measure of 'Accuracy' defined in section 4.2.1, the overall results of the proposed initial lesion detection approach can be summarized as in Table 4-3.

Table 4-3 Initial lesion detection accuracy when using local mean and D_m, as metrics

Method	Positive	Negative	Accuracy
Local mean	263	117	69.21%
Fractal Analysis	206	174	54.21%

The results illustrate that Fractal analysis is significantly less accurate in initial lesion detection as compared to using the local mean. The reason is that the value of D_m is highly dependent on the characteristics of the neighbourhood pixels. The lesions with high intensity and homogeneity may produce very low values of D_m. However, referring to the graphs in figure 4-11, there is a clear cut separation between the average and low value points in fractal analysis as compared to that of statistical analysis. This implies that the segments in fractal analysis are more separable. Further detailed analysis of our results revealed that statistical analysis is accurate in the detection of cysts and malignant lesions, but less accurate in the case of detection of fibroadenoma. Most of the positive results in the experiment refer to the detection of cysts and malignant lesions, i.e., the lesions with low intensity. On the other hand, the negative cases in this experiment refer to fibroadenoma.

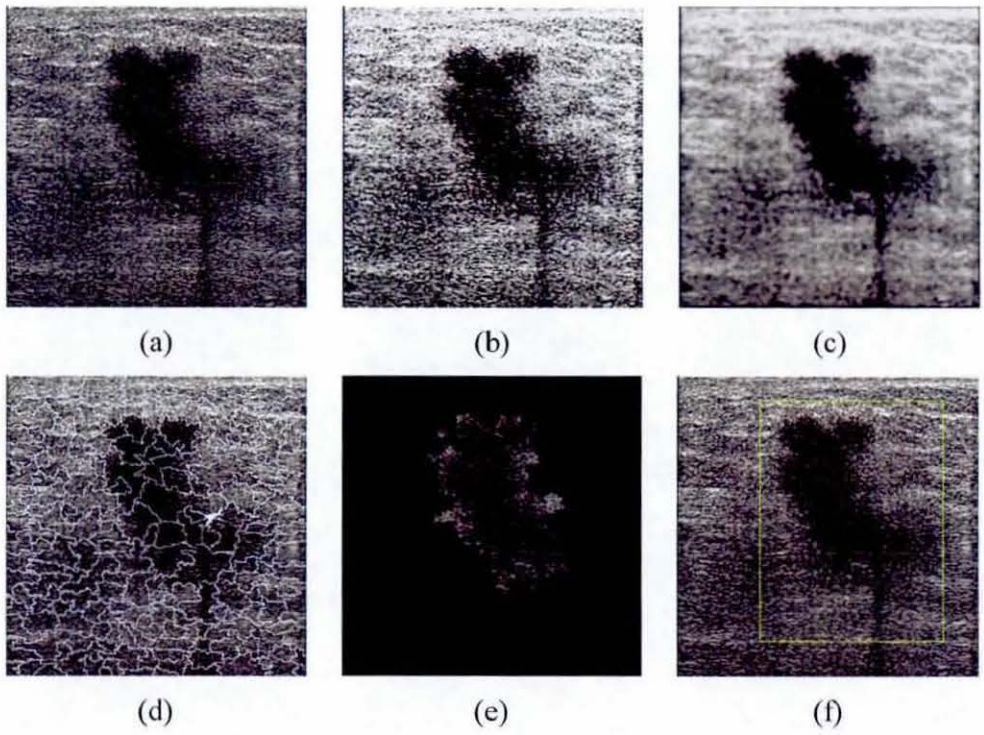


Figure 4-12 Illustration of the operation of the intermediate stages of the proposed method (a) Original image (b) Image after pre-processing (histogram equalization) (c) Image after hybrid filtering (d) Watershed segmentation mapped on to the original image (e) Segments with the same cluster with the initial lesion (f) Labelling of region of interest (ROI) by using the proposed method.

Figure 4-12 illustrates the stage-by-stage operation of the proposed algorithm. The proposed method worked well in detecting cyst lesions, malignant lesions, and some of the fibroadenoma lesions as illustrated in figure 4-13. A comparison with the manual labelling is included.

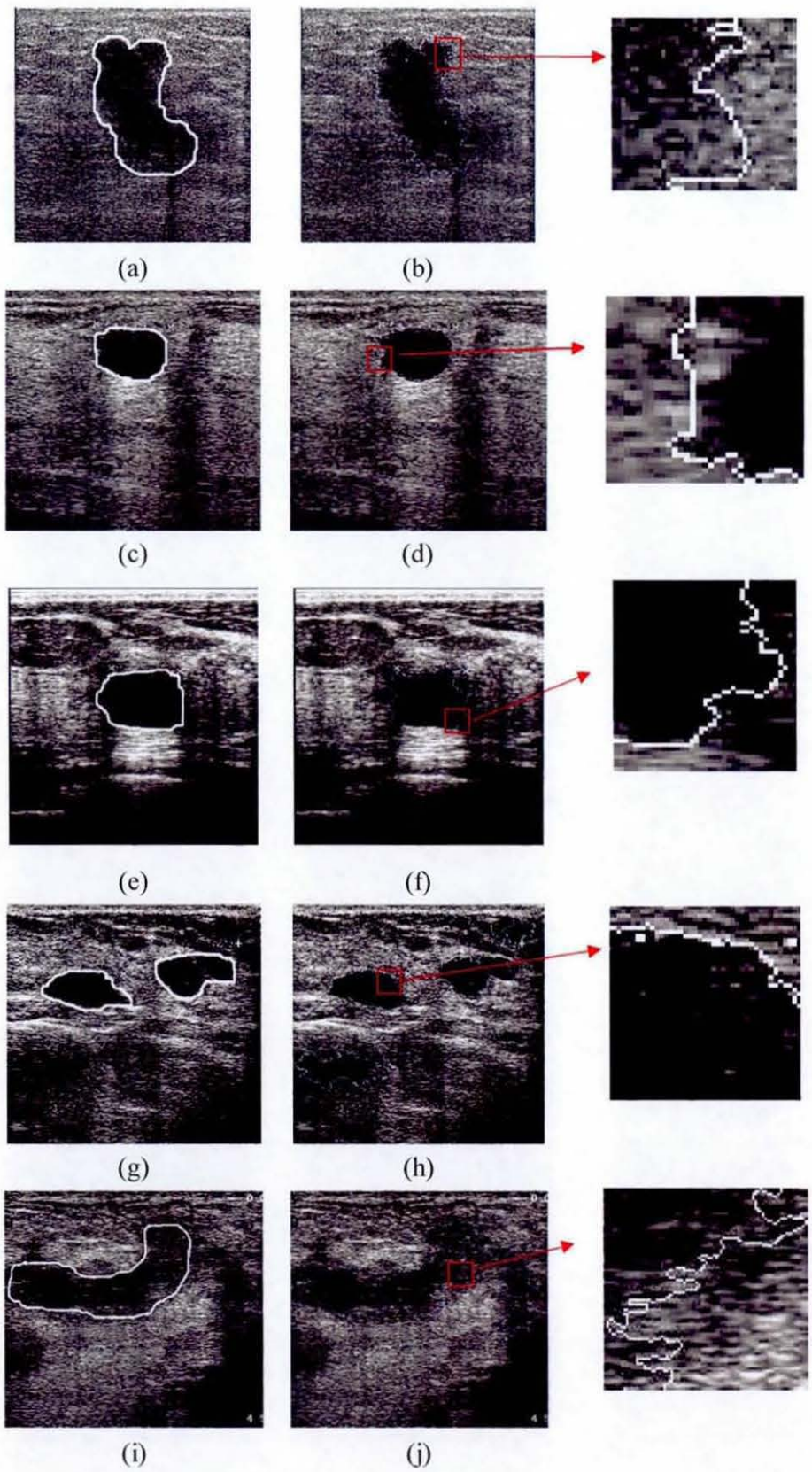


Figure 4-13 (a), (c), (e), (g),(i) Manual labelling of ROI. (b), (d), (f), (h),(j) The results of automated labelling by using our method.

Fibroadenoma are fibre tissues very similar to the normal tissues and was found to be the main cause behind missed detection of ROI. The proposed technique failed to detect some of the suspected Fibroadenoma lesions as illustrated in Figure 4-14.

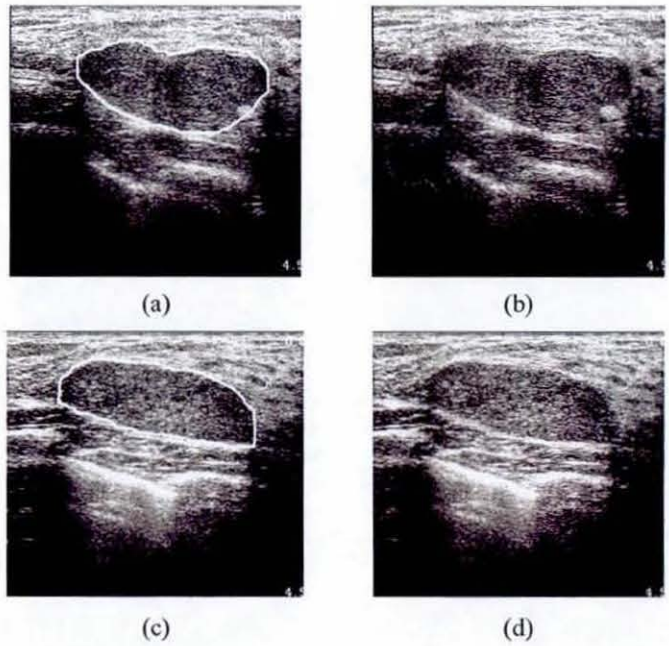


Figure 4-14 (a), (c) Manual labelling of ROI. (b), (d) The results of automated labelling of the proposed method.

4.8 Method-2

Figure 4-15 illustrates a modular block diagram of the second proposed approach to automatic lesion detection. It uniquely combines histogram equalisation, hybrid filtering, multifractal analysis, thresholding segmentation, and a rule-based approach in fully automated ROI labelling.

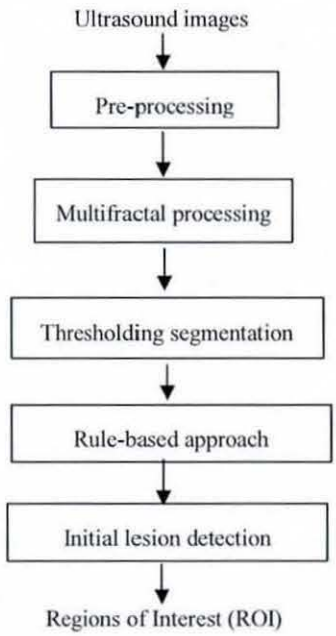


Figure 4-15 Overview of Method-2.

The operation and functionality of the individual stages can be described in detail as follows (note: all algorithms were implemented in LTI-Lib [95], on a Linux platform):

Subsequent to pre-processing stage (see section 4.4), multifractal processing [96] is used to further enhance the partially processed images. In section 4.6.1, experimental results are provided to show that this stage enables better segmentation of lesions as compared to solely applying hybrid filtering. The following section provides a brief overview of multifractal processing and the associated analysis methodology.

4.8.1 Multifractal Processing

A fractal is generally "a rough or fragmented geometric shape that can be subdivided in parts, each of which is (at least approximately) a reduced-size copy of the whole" [97], a property called self-similarity. In analysing the fractal geometry of an image an

attempt is made to exploit self-similarity present. In fractal geometry, the term, fractal dimension, refers to a statistical quantity that gives an indication of how completely a fractal appears to fill space, as one zooms down to finer and finer scales [97]. Multifractal analysis refers to the analysis of an image using multiple fractals (i.e. not just one as in fractal analysis).

The generalized formulation for Multifractal Dimensions (D) of order q can be represented as follows [96]:

$$D_q = \begin{cases} \dfrac{1}{q-1} \lim\limits_{\varepsilon \to 0} \dfrac{\log(x_q(\varepsilon))}{\log(\varepsilon)} & for\ q \in R\ and\ q \neq 1 \\ \lim\limits_{\varepsilon \to 0} \dfrac{\sum\limits_i \mu_i \log \mu_i}{\log(\varepsilon)} & for\ q = 1 \end{cases} \qquad (4\text{-}10)$$

where ε is the linear size of the cells (Note: in our case, we use 3×3 pixel mask, hence, $\varepsilon = 3$, q is the order for cell size ε. Note that when $q<0$, Dq is sensitive to the parts where the measure is very dense. On the other hand, if $q>0$, information on the sparse region can be obtained. In theory, q is in the range $-\infty$ to ∞, and D_q can have an infinite number of values. In practice computing for all values of q is not possible. Hence only four values of q were empirically decided to be used in our experiments, which are, -1, 0, 1, and 2.

The partition function x_q is defined as, $x_q(\varepsilon) = \sum_{i=1}^{N(\varepsilon)} \mu_i^q(\varepsilon)$, where $N(\varepsilon)$ is the total number of cells of size ε, $\mu_i^q(\varepsilon)$ is the measure that is defined on a given set. In this case, the measures are defined as the probability of the grey scale level in the images, where all the grey levels fall in the range of 0 - 1.

To investigate the effect of different q values on ultrasound images, further empirical experiments are carried out.

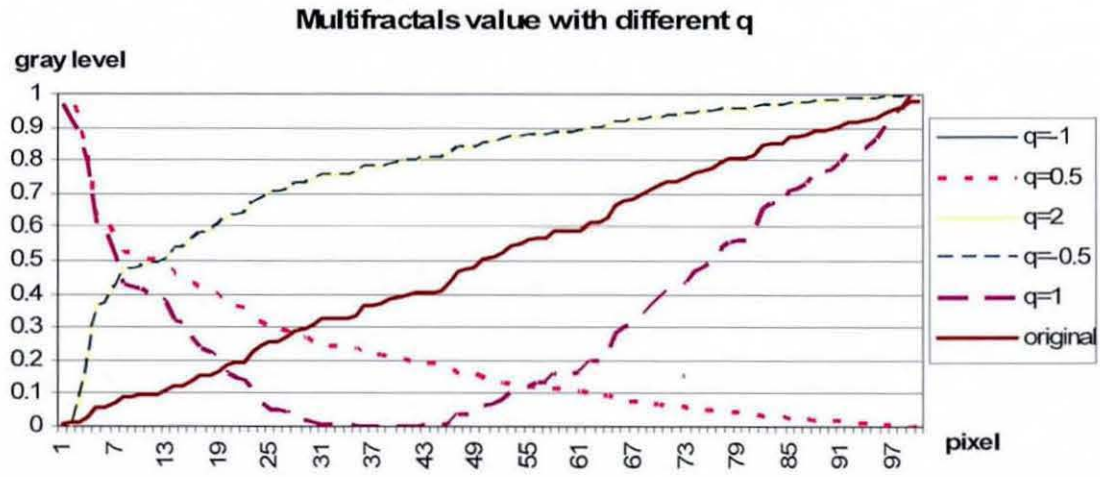


Figure 4-16 Graph plotted for multifractals dimension with different value of q.

The graph in figure 4-16 shows that the multifractals with $-\infty < q < \infty$ yield four types of results. The results can be classified into, $q=0$, $q=1$, $0<q<1$, and ($q<0$ or $q>1$). Note that $0<q<1$ is the inverted value of ($q<0$ or $q>1$). From the results illustrated in figure 4-17, it is seen that any value range $0<q<1$ or ($q<0$ or $q>1$) will help improve the segmentation results. The value $q = -1$, i.e. D_{-1} is chosen as it has the lowest associated computational complexity compared to other values.

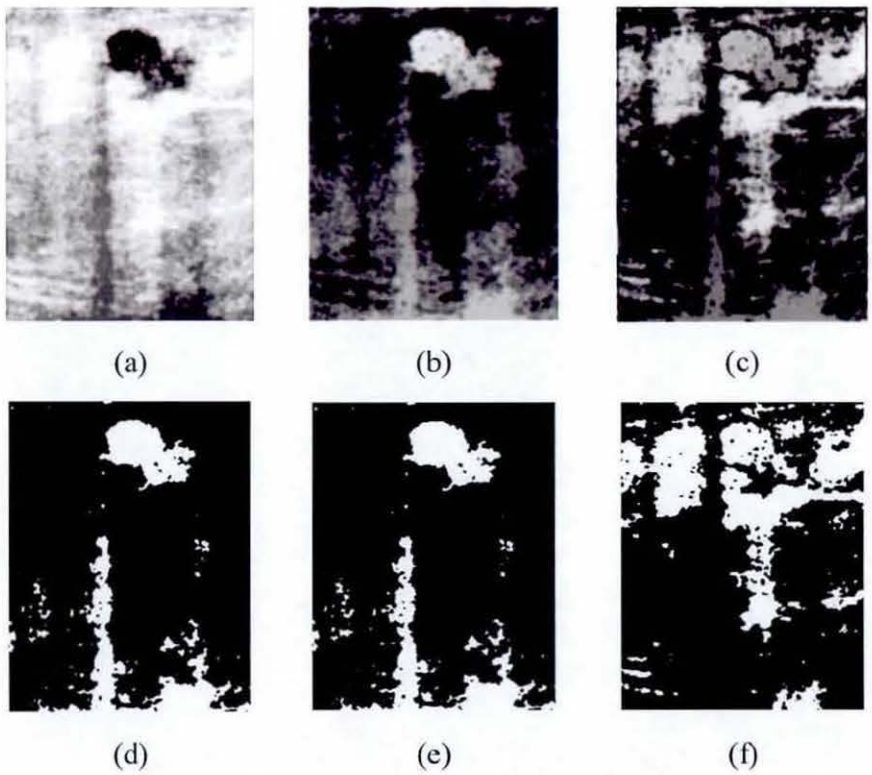


Figure 4-17 Result of multifractal analysis with (a) $q<0$ or $q>1$, (b) $0<q<1$, (c) $q=1$, (d) threshold of (a), (e) threshold of (b), (f) threshold of (c).

To the author's knowledge hybrid filtering has not been used in previous research in breast ultrasound boundary detection or contrast enhancement, and hence is an additional aspect of novelty of this approach.

After the application of hybrid and multifractal filtering the images are ready for lesion segmentation. The following section introduces the reader to the approach adopted.

4.8.2 Thresholding Segmentation

In general, segmentation is a process used to divide an image into its constituent parts. Thresholding segmentation [98] is the most basic, simplest, and fastest algorithm in segmentation. A thresholding procedure attempts to determine an intensity value, named the threshold, which separates the pixels into desirable classes. A parameter θ, named the brightness threshold is chosen and applied to the image as follow:

$$\begin{aligned} &if\ a[m,n] \geq \theta \\ &\quad then\ a[m,n] := 1(object) \\ &else\ a[m,n] := 0(background) \end{aligned}$$

Within the present research context, a fixed threshold is used for segmentation, i.e. a threshold that is chosen independently of the image data. If it is known that one is dealing with very high-contrast images where the objects are very dark and the background is homogeneous and very light, then a constant threshold of 128 on a scale of 0 to 255 was experimentally found to be sufficiently accurate, i.e. the number of falsely-classified pixels is kept to a minimum. The sensitivity of the threshold selection on segmentation accuracy was found to be low, further justifying the use of an image independent fixed threshold.

4.8.3 Rule-based Approach

The thresholding segmentation of section 4.8.2 often leads to the identification of multiple ROIs, of which generally only one or two regions would be of diagnostic importance, i.e. would belong to abnormal lesions. Further the location of the abnormal lesions requires the specification of both position and orientation (assuming 2D ultrasound images). Therefore the use of rule-based approach is proposed to identify these important ROIs, as discussed in below:

The first criterion used for the identification of lesions is the size of the segments. The suspected lesions, are identified as the largest segments out of the likely multiple segments that results from applying the single threshold segmentation. In addition, based on the additional guidance provided in the Ultrasound CD used to obtain the test data, it was observed that 95% of the tumours are located at the upper regions of the images. Hence, a reference point at (x, y), where,

$$x = \frac{image\ height}{3}, and\ y = \frac{image\ width}{2}$$

is chosen as the centre of attention. The lesion that is located closest to the above point, satisfying the above mentioned size related criteria is selected as the final detected lesion. It is noted that the above criteria can be slightly relaxed appropriately if more than one lesion is required to be detected.

4.8.4 Result and Analysis

A detailed presentation of results and an analysis can be given as follows:

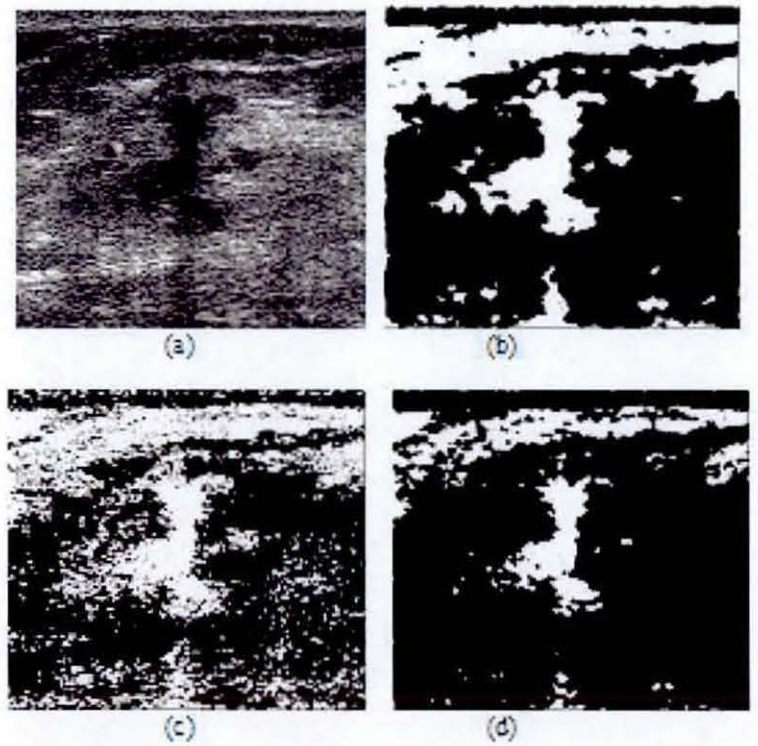


Figure 4-18 Comparison of segmentation approaches (a) original image, when using only (b) Gaussian filtering + Multifractal (c) Non-linear + Multifractal filtering (d) Hybrid + Multi-fractal filtering

Figure 4-18 compares the effects of using different filtering techniques discussed in section-4.4 with multifractal processing. When using Gaussian filtering (see figure 6(b)), due to the smoothing effects introduced at the edges, the noisy regions which should ideally be disconnected, remain connected. This is a problem in the subsequent lesion detection stage as the largest connected region may now not refer to the true lesion that should be detected. The use of Non-linear filtering results in over segmentation and causes many problems, as illustrated by figure 6(c). However, it is noted that the use of hybrid filtering with multifractal processing results in the single

largest segmented area detected, to be the lesion.

Figures 4-19 illustrates the stage-by-stage operation the proposed multifractal based approach to initial lesion detection. A comprehensive performance comparison with the benchmark algorithms, based on the accuracy of the final, detected, lesion regions, is given later (see figure 4-23).

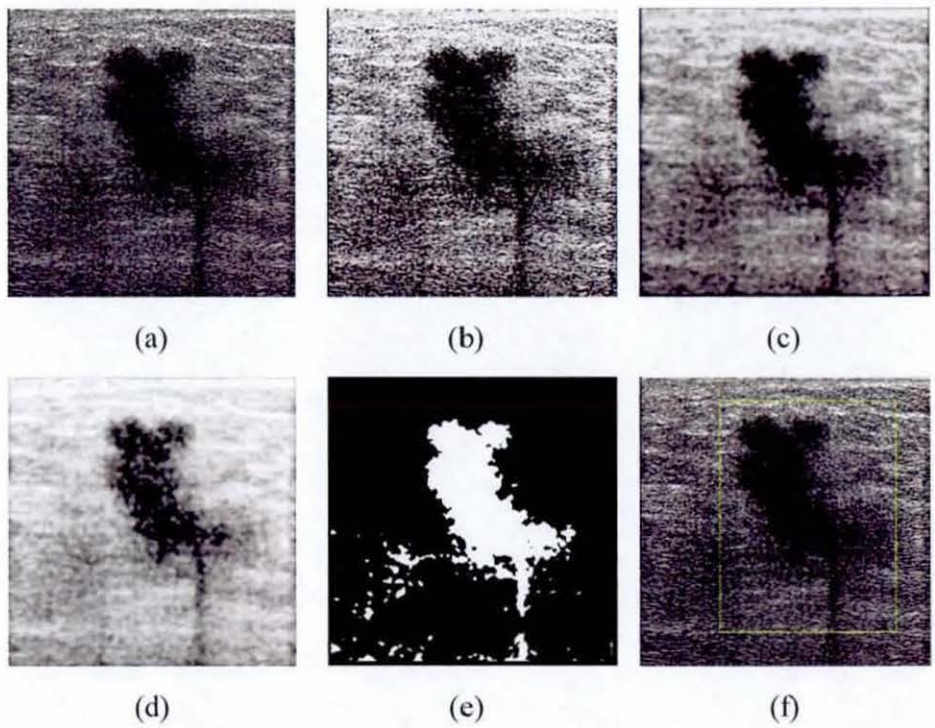


Figure 4-19 Illustration of the operation of the intermediate stages of the proposed algorithm (a) The original image. Image after, (b) pre-processing (histogram equalization) (c) hybrid filtering (d) multifractal processing (e) thresholding segmentation (f) and labelling of region of interest (ROI).

Detailed experiments are performed using the 360 test ultrasound images revealed that the proposed method performs exceptionally well in identifying ROIs of most cyst lesions, malignant lesions, and some of the fibroadenoma lesions. It is noted that due

to the high degree of similarity in texture between normal and fibroadenoma regions, the accurate identification of such regions is always a challenge. A visual comparison of results when using the proposed algorithm on different types of abnormalities is illustrated in figure 4-20. Figure 4-21 illustrates two examples of fibroadenoma ROIs which are not detected accurately.

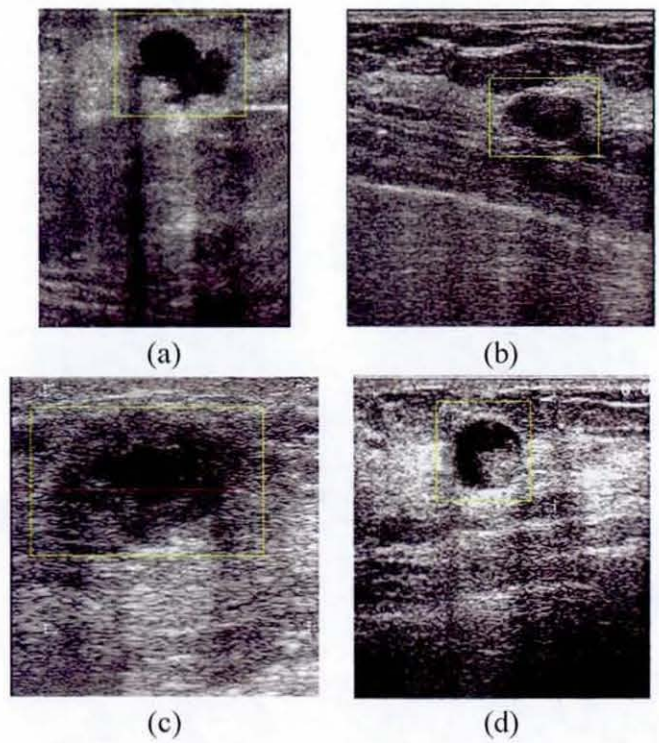


Figure 4-20 The results of automated lesion ROI labelling when using the proposed method, (a) malignant tumour (b) simple cyst (c) fibroadenoma (d) complex cyst.

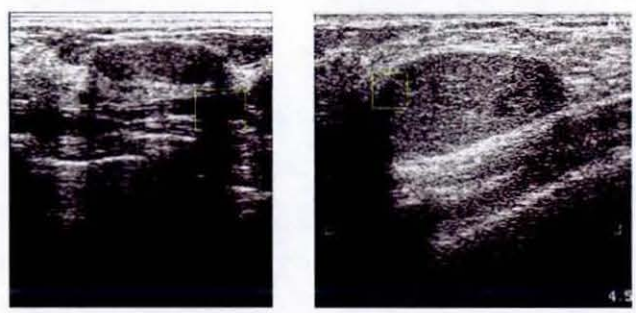


Figure 4-21 Examples of unsuccessful lesion identification for two cases of fibroadenoma.

4.9 Comparison of the proposed methods against benchmark algorithm

Based on the diagnosis performed by radiologists and included in the ultrasound breast imaging CD used in our experiments, the accuracy figures obtained for each type of abnormality for the two proposed methods and the benchmark algorithm of Drukker et.al. [14] are summarized in table 4-4.

Table 4-4. Summary of the ROI detection accuracy for each type of abnormality.

Diagnosis	Total number of images	Accuracy			
		Drukker et al.	Method-1 (statistical analysis)	Method-1 (fractal analysis)	Method-2
Malignant	20	80.00%	65.00%	45.00%	90.00%
Simple Cysts	76	67.11%	60.53%	55.26%	86.84%
Complex Cysts	76	72.37%	71.05%	59.21%	89.47%
Fibroadenoma	58	62.07%	63.79%	46.55%	77.59%
Carcinoma	38	57.89%	78.95%	52.63%	78.95%
Occult lesions	18	94.44%	88.89%	66.67%	88.89%
Adenosis	15	80.00%	73.33%	46.67%	93.33%
Others	59	59.32%	64.41%	54.24%	89.83%
Total	360	67.78%	68.05%	53.89%	86.11%

Figure 13 shows the graphical presentation of the ROI detection accuracy for each type of abnormality of the two approaches presented in this chapter (method-1 and method-2) and that of the benchmark algorithm. Results clearly prove the lesion ROI detection accuracy improvements obtainable by the approach adopted by method-2 are better. The detection accuracy for Fibroadenoma type lesions have generally been the lowest for all methods. However the algorithm of method-2 indicates a 15% improvement in accuracy, in this category as compared to Drukker et. al's [14]

method and that of method-1 (using local mean and the fractal dimension). It is noted that a 90% accuracy is indicated in detecting malignant type lesions when using the proposed algorithm, which is a significant result.

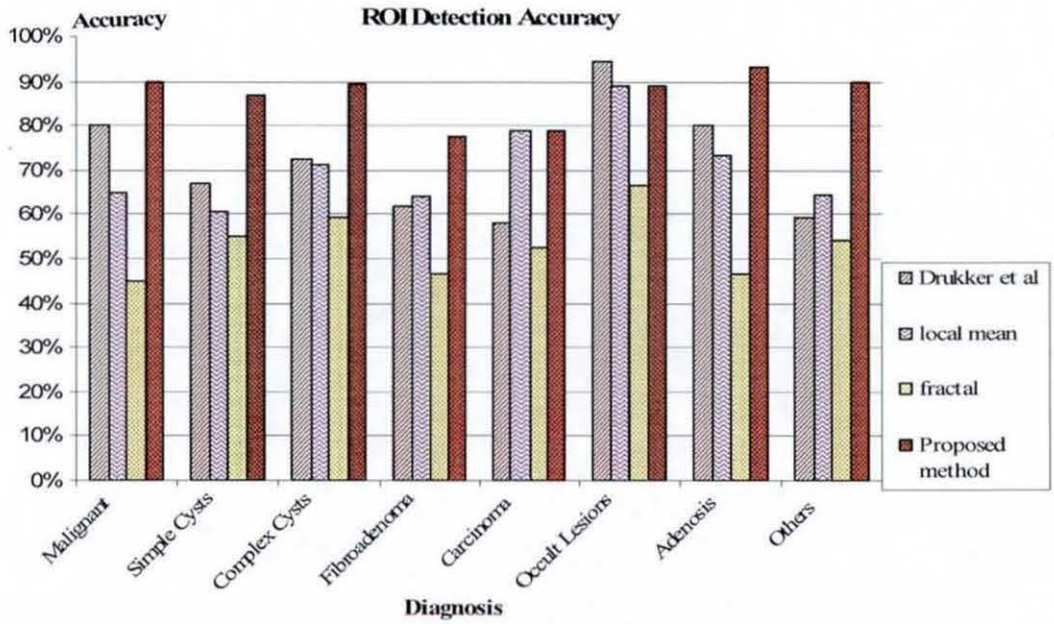


Figure 4-22 Graphical presentation of the ROI detection accuracy for each type of abnormality of Drukker's algorithm, method-1 (with the use of local mean and fractal dimension) and method-2.

Figure 4-23 visually compares the performance of the proposed method-2 to that of method-1 (using both local mean and fractal dimensions) and Drukker et al.'s algorithm. It clearly illustrates the improved accuracy of lesion identification demonstrated by the second approach proposed by method-2. It is seen that the other three algorithms are more likely to identify non-lesion regions or only parts of lesions, as ROIs in the presence of rather challenging diagnostic cases.

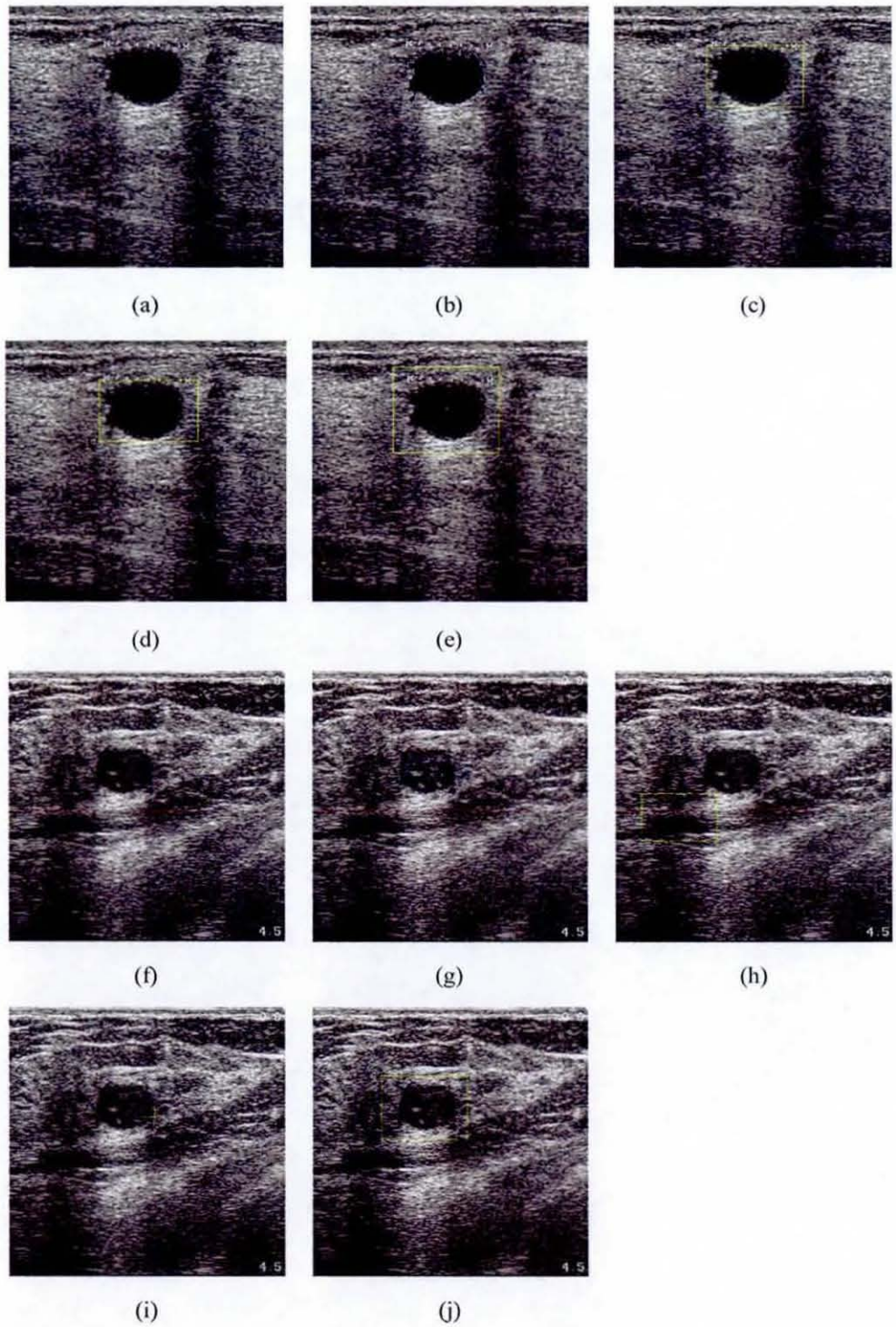


Figure 4-23 Visual performance comparison of method-2 against that of method-1 and the benchmark algorithm. (a),(f) are the original images. (b),(g) are the results of Drukker et al.'s algorithm. (c),(h) are the results of method-1 (using local mean). (d),(i) are the results of method-1 (using the fractal dimension). (e),(j) are the results of the proposed method.

4.10 Summary & Conclusion

The chapter focused on investigating the possibility of fully automatic ROI labelling in ultrasound breast images. The current CAD systems use manual labelling of ROIs by radiologists. Although a number of automatic ROI labelling approaches have been proposed in literature Drukker et al.'s work presented in [14] is one of the best performing state of the art algorithms. Therefore in this chapter we introduced it in detail and have critically evaluated its performance. The results were later used in evaluating the performance of the novel approaches that were presented in the chapter. Further the performance of a hybrid filtering approach that combines two different filters was investigated. It was later utilized for the reduction of speckle noise in the design and implementation of the novel ROI labelling algorithms.

Two novel methods for automatic ROI labelling were proposed. In the first method, it was shown that marker-controlled watershed segmentation can be used to overcome the over-segmentation problem in ultrasound breast image processing. The use of a statistical analysis approach and a fractal analysis approach in separating lesion regions from non-lesion regions were investigated. It was concluded that statistical analysis performs better than fractal analysis. In addition it was shown that *k*-means clustering can be used to further increase the accuracy of the initial lesions detection. However it was observed that *k*-means clustering increases the false positive rate.

Multifractal analysis played an important role in second novel approach proposed for initial lesion detection. It was shown that multifractal analysis enables improved separability of tumour regions from normal regions. A performance comparison

between the two novel approaches proposed for initial lesion detection and the benchmark algorithm, proved that the multi-fractal analysis based approach performed best for the experimental dataset used. It was found that due to the inconsistency issues in ultrasound images, some parameters in the processing stages will have to be adjusted depending on the dataset being experimented.

The multifractal based approach that performed best indicated a 90% accuracy in identifying malignant lesions. It's worst performance was for fibroadenoma, i.e. an accuracy of 77.59%. In chapter 8 we discuss possible improvements to the proposed approaches.

Chapter-5 proposes novel methods for the lesion segmentation, i.e. lesion boundary detection of the ROIs identified in this chapter.

Chapter 5: Boundary Detection

5.1 Introduction

This chapter focuses on the segmentation problem, i.e., obtaining the best boundary detection of lesions in US images. Two methods are proposed. The first method, namely multifractal RGI (Radial Gradient Index) approach, performs lesion segmentation in cropped US images by adopting multifractal processing and region growing segmentation. The performance of this approach is compared with that of the benchmark algorithms, Joo et al. [16], and Kupinski et al. [15]. The second method, namely the multifractal Iso-Gauss approach, is a fully automated segmentation algorithm that can be applied directly on un-cropped US images. The performance of this approach is compared with that of the benchmark algorithm of Drukker et al.'s [14]. Note that the dataset/images used in the experiments of this chapter are from the Breast Ultrasound CD [22], of which a detailed discussion has been provided in Chapter-4.

5.2 Method-1: Multifractal RGI Approach

This sections presents the first novel approach proposed in this chapter for lesion segmentation.

5.2.1 Research Motivation and Benchmark Algorithms

Ultrasound images suffer from speckle noise due to the interference of back scattered signals. This noise significantly degrades the image quality and hinders discriminating the finer details [35]. Due to this relatively poor quality of ultrasound images, their segmentation is a difficult problem. Despite this, segmentation is an important intermediate stage of an ultrasound imaging system as the nature/shape of the boundary of a lesion play an important role in differentiating between malignant and benign tumours [15]. i.e. in many CAD systems, a lesion segmentation step is initially used to obtain an accurate representation of the boundary of a tumour whose features are subsequently determined and used in lesion type classification [16]. According to Joo et al. [16], the reason behind the inaccuracy of many existing ultrasound CAD systems is not the classification stage, but the segmentation stage. Even if corrective adjustments of the labelling were performed by the radiologists, obtaining the perfect boundary of the tumour lesions remains a challenge.

Besides the speckle noise of ultrasound images, the shape of the tumours is another factor that complicates obtaining an accurate boundary. It has been established that the benign lesions of Ultrasound breast images are regular in shape, but the malignant lesions are irregular in shape.

One way to describe the size and the self similarity of a data set is to calculate its fractal dimension [99]. The preliminary research indicated that a single fractal dimension alone is not sufficient to fully discriminate textures and natural surfaces [100]. Based on the above observations, it is crucial to study and further investigate

the practical use of multifractal analysis, which are rotation and intensity invariant. This chapter shows that by processing an ultrasound image in a multifractal dimension, a possibility exists to improve segmentation, which directly results in a refinement of the boundary of lesions.

Two existing, state-of-the-art techniques [15, 16] on boundary detection in US breast images, used as benchmarks to compare the performance of the first approach proposed in this chapter, can be summarised as follows:

In Kupinski et al.'s research work, a radial gradient index (RGI)-based algorithm is used for segmenting lesions in mammographic images. In order to apply the algorithm to ultrasound images, the images are inverted and filtered by median filtering [14]. The centre of the rectangular ROI is chosen as the seed of the algorithm, based on the assumption that the lesion is located at the centre of the ROI. Note that conventional region growing algorithms are constrained by the grey level of the images, but not the shape. In the RGI-based algorithm adopted in [56], in order to incorporate the lesion shape into the creation of the partition, the original image *f(x,y)* is multiplied by an Isotropic Gaussian function centred at the seed point location (μ_x, μ_y) with a fixed variance σ_c^2 as the constraint function [56]. The resulting image, *h(x,y)* is given by:

$$h(x,y) = f(x,y)N(x,y;\mu_x,\mu_y,\sigma_c^2) \qquad (5\text{-}1)$$

The radial gradient segmentation is used to obtain the lesion boundary. The RGI value is calculated from the original image as follows:

$$RGI = \frac{\sum_{(x,y)\in M} \hat{G}(x,y) \bullet \vec{r}(x,y)}{\sum_{(x,y)\in M} \left\| \hat{G}(x,y) \right\|} \qquad (5\text{-}2)$$

where M is the boundary point of the partition, $\hat{G}(x,y)$ is the gradient vector of $f(x, y)$ at position (x, y), and $\vec{r}(x,y)$ is the normalized radial vector at position (x, y). From the definition, due to normalization, RGI values are between -1 and 1, where a RGI value of +1 signifies that along the contour all gradients point radially outwards, and a value of -1 means that all gradients point radially inwards. Actual lesions are expected to have (absolute) RGI values close to 1 [14]. Finally, the partition with the maximum RGI is returned as the final lesion partition.

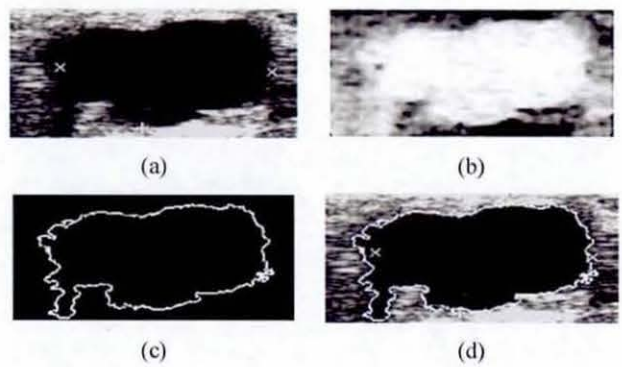


Figure 5-1 RGI based algorithm of Kupinski et al. (a) original image (b) inverted image preprocessed by median filtering. (c) detected boundary. (d) the detected boundary mapped on the original image.

In Joo et al.'s research work [16], median filtering is used for pre-processing. The ROI is processed with a median filter to remove noise and enhance features [16]. To improve the perceptibility of the edges of breast nodules, unsharp masking, which is

well-known in enhancing the structures in images, is used. Then, the contrast of the ROI is enhanced by using a contrast stretching transformation. The ROI is subsequently converted to a binary image by using binary thresholding. The noise generated by binary thresholding is removed by another median filtration stage. Finally, *island areas*, smaller than 500 pixels are removed from the ROI, and an edge detection algorithm is used to detect the final boundary. In the work presented in [16], some of the detected boundaries are manually corrected by a radiologist. Note that in the results shown in this chapter, all presented results are automatically generated by the boundary detection algorithms discussed and are not corrected by a radiologist.

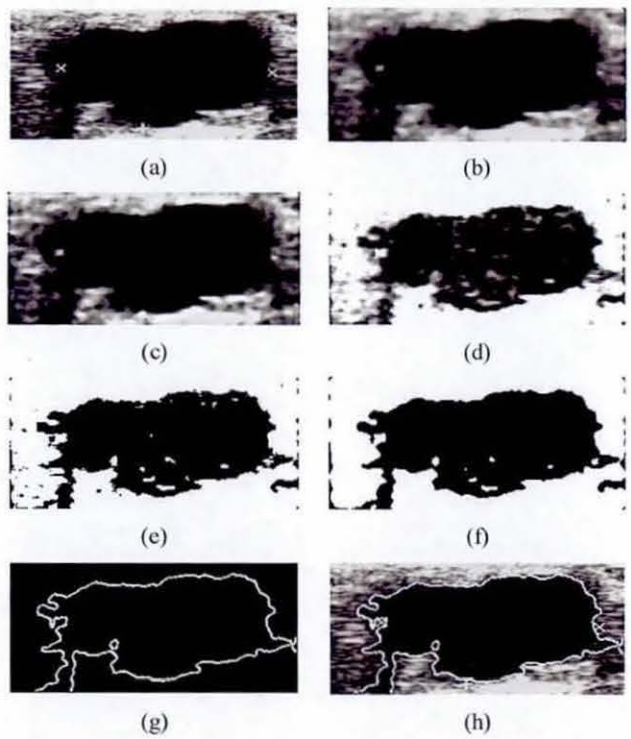


Figure 5-2 Boundary detection algorithm of Joo et al.[16] (a) original image. After, (b) median filtering, (c) unsharp masking, (d) contrast enhancement, (e) binary thresholding, (f) median filtering and (g) the detected boundary (h) the boundary mapped on to the original image.

5.2.2 The Methodology

Figure 5-3 illustrates the block diagram of the first approach proposed, i.e. the multifractal RGI approach. Each section is subsequently discussed in detail.

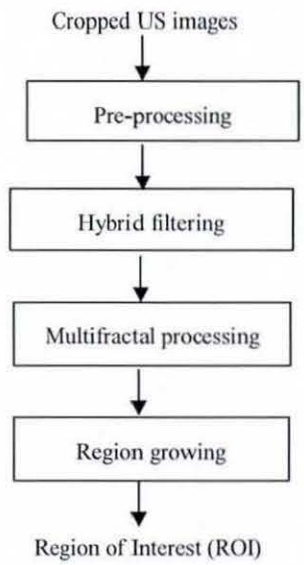


Figure 5-3 Block diagram depicting stages of Approach-1.

The pre-processing, hybrid filtering, and multifractal processing algorithms used within the current context of research were discussed in Chapter 4. Thus only the final stage, i.e. the *region growing algorithm* that leads to the ultimate segmentation of lesion boundaries, is presented in the following section.

5.2.2.1 Region Growing Segmentation

Region growing segmentation is used to detect the boundaries of the images. The thresholds used to control the growing of each partition, is in the range of 128-180. For each of the partitions produced, the RGI value is calculated, as shown in Figure

5-4. The boundary of the partition with the maximum value of RGI is selected as the best boundary [56].

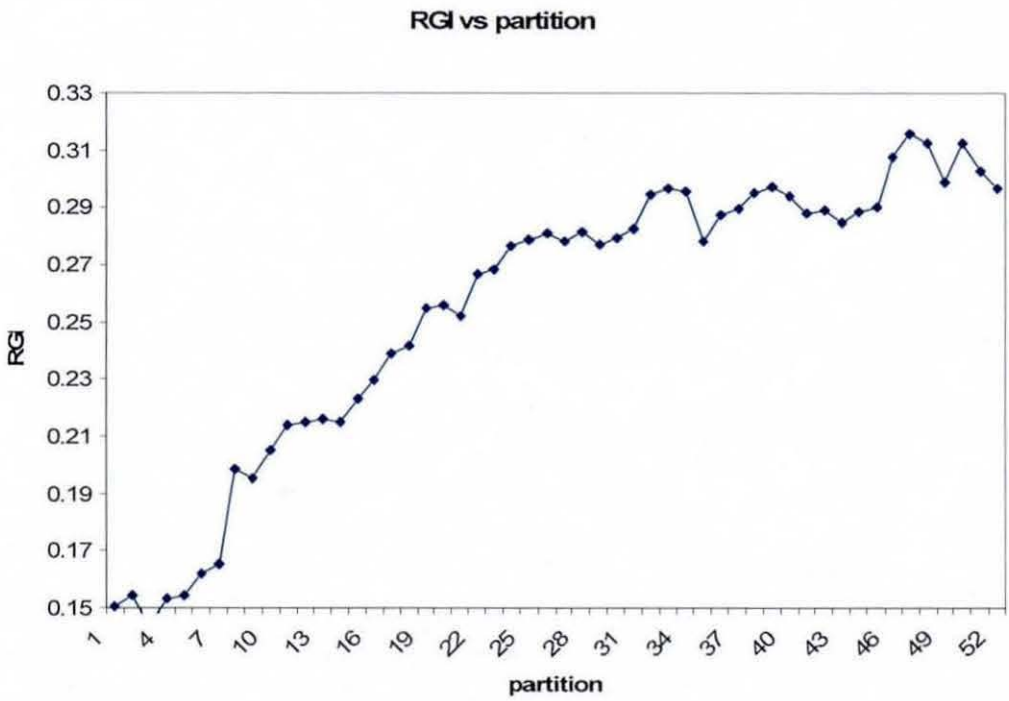


Figure 5-4 The RGI value of each partition. Note that the partitioning is based on the threshold value used. The partition with the largest RGI value is returned as the final lesion partition.

5.2.3 Results and Analysis

The implementation of algorithm proposed was carried out using C++, with the help of the Computer Vision Library, LTI-LIB [95]. Figure 5-5 illustrates the results obtained by the proposed method, in particular the outputs at the various intermediate stages.

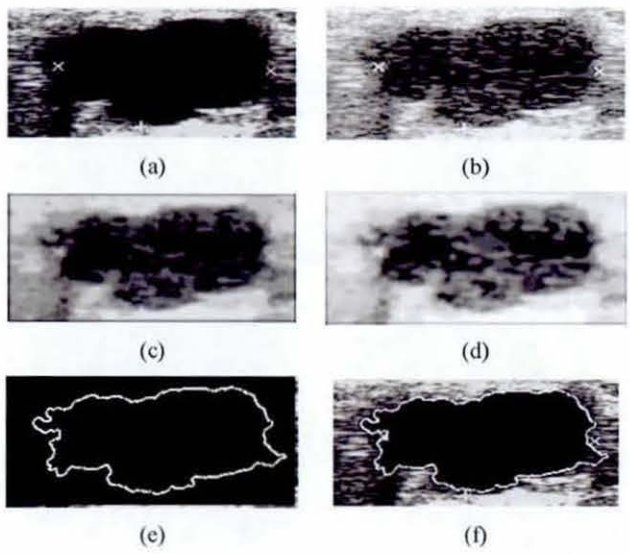


Figure 5-5 Results of the proposed method. (a) The original image. The processed image after, (b) histogram equalization, (c) hybrid filtering, (d) multifractal dimension/analysis, (e) the boundary detected by using the region growing method (Note: the boundary with highest RGI is selected.), (f) the detected boundary mapped on to the original image.

To justify the positive contribution of using multifractal analysis in the proposed methodology, in Figure 5-6 we illustrate the effect of directly using region growing segmentation on the output of the hybrid filtering stage, i.e. on the image illustrated in Figure 5-5(c). In other words, the multifractal analysis stage has been removed from consideration. When the identified region illustrated in Figure 5-6 is compared to that illustrated in Figure 5-5(f), the positive contribution of the multifractal stage is clearly illustrated. It is noted that when multifractal analysis is not used, the lesion includes unwanted noisy areas that should not be a part of the lesion detected.

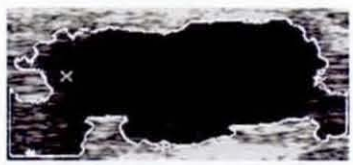


Figure 5-6 Detected boundary when multifractal analysis is excluded.

In Figure 5-7 we illustrate and compare the results obtained by the two benchmark algorithms and the proposed method on five different ultrasound images. It clearly illustrates the improved performance obtainable by using the proposed algorithm.

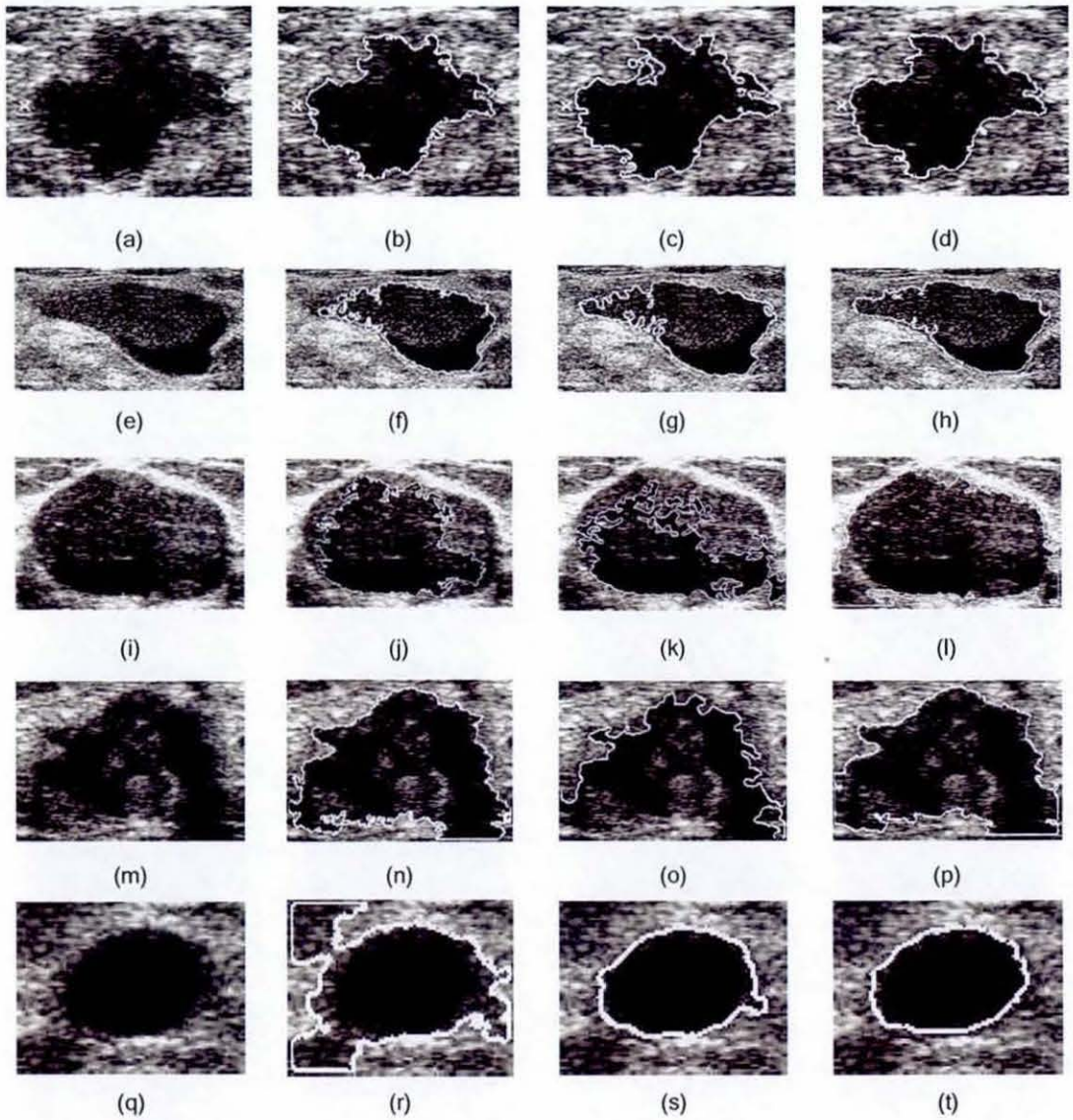


Figure 5-7 Comparison between the three boundary detection methods. (a),(e),(i),(m),(q) are the original images. Results produced by; (b),(f),(j),(n),(r) Kupinski et al.'s[15] method. (c),(g),(k),(o),(s) Joe et al.'s method [16]. And (d),(h),(l),(p),(t) the proposed method.

A careful comparison of the results illustrated in figure 5-7 show that the proposed method is capable of accurately detecting the edges of the lesions. Further the

detected edges are smoother and thus better satisfy acceptability requirements of an automatic detection method. The above final conclusions have been made based on experiments carried out on a larger set of test images.

5.3 Method-2: Multifractal IsoGauss Approach

This section presents the details of the second novel approach proposed in this thesis for lesion segmentation.

5.3.1 Research Motivation and Benchmark Algorithms

At present ultrasound Computer Aided Diagnostic (CAD) systems require the identification of ROI and the initial marking of the centre of interest by an expert radiologist, before being able to automatically produce lesion boundaries. Thus the process of lesion boundary detection and segmentation in existing CAD systems is, semi-automatic. Thus there is a need for a fully automated approach, which will not require human judgment or decision making at an intermediate stage.

In the proposed Multifractal IsoGauss approach, histogram equalization is initially used to pre-process the images followed by hybrid filtering and multifractal analysis stages. Subsequently, a single valued thresholding segmentation stage and a rule-base approach is used for the identification of the lesion ROI and the point of interest that is used as the seed-point. Subsequently, starting from this point, an Isotropic Gaussian function is applied on the inverted, original ultrasound image. The lesion area is then

separated from the background by a thresholding segmentation stage and the initial boundary is detected via edge detection. Finally to further improve and refine the initial boundary, a state-of-the-art active contour method (i.e. gradient vector flow (GVF) snake model) is implemented.

The proposed Multifractal IsoGauss approach (proposed method) is compared with an existing state-of-the-art method, namely the radial gradient index filtering (RGI) technique of Drukker et al. [14] in justifying its robustness and accuracy. The above algorithm has been specifically selected as a benchmark as it is the only existing fully automatic lesion detection and segmentation approach.

5.3.2 The Methodology

Figure 5.8 illustrates the block diagram of the proposed approach. Stage 1 is to obtain the position of the region of interest. In stage 2, we obtain the initial contour of the tumour by using an Isotropic Gaussian function, and enhance the initial contour by using a special type of snake, namely, Gradient Vector Flow (GVF) [101]. The two main objectives of this section are to investigate the effect of Isotropic Gaussian function in obtaining a more accurate initial boundary, and the use of GVF in boundary detection and segmentation. The detailed operation of the proposed algorithm is presented in this section.

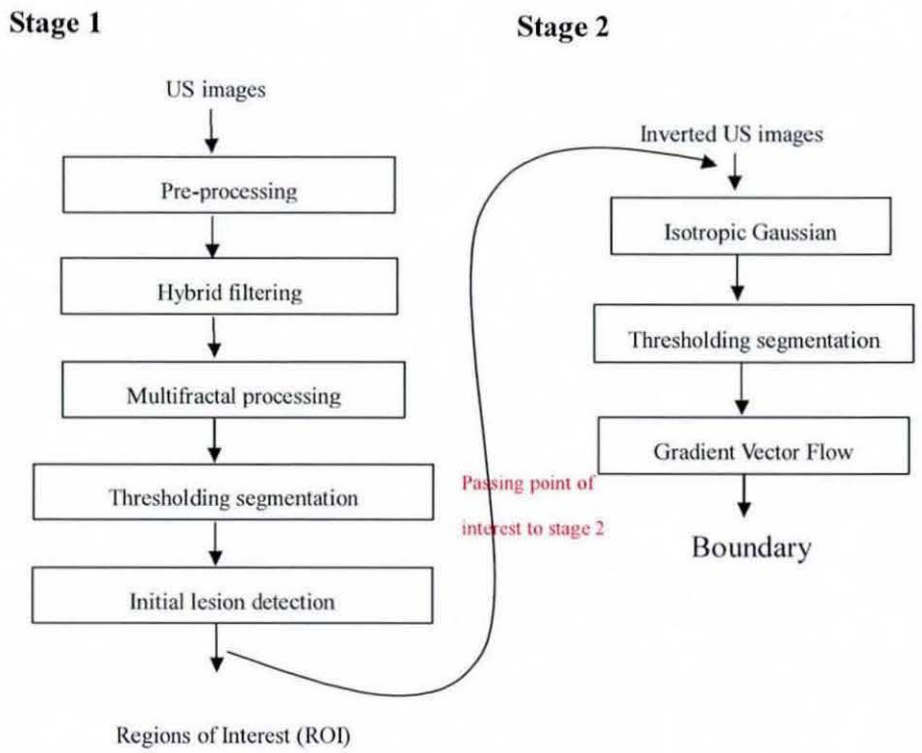


Figure 5-8 Overview of methodology.

5.3.2.1 Ultrasound Images

A total of 360 images from the Breast Ultrasound CD [22] are used in this experiment. Please refer to Chapter 4 for further explanation.

5.3.2.2 Initial Lesion Detection

The detailed discussion on the initial lesion detection stage, i.e. stage 1 of figure 5-8, is presented in Chapter 4, section 4.8, Method-2.

5.3.2.3 Initial Contour Detection

In general, segmentation is a process used to divide an image into its constituent parts. It is credibly accepted that there are no standard segmentation methods that perform

equally well in all images. In medical imaging, more specifically, breast ultrasound images, segmentation is an important intermediate stage of an ultrasound imaging system as the nature of the boundary of a lesion play an important role in differentiating between malignant and benign tumours [15]. In many CAD systems, a lesion segmentation step is initially used to obtain an accurate representation of the boundary of a tumour whose features are subsequently determined and used in lesion type classification [16]. According to Joo et al. [16], the reason behind the inaccuracy of many existing ultrasound CAD systems is not the classification stage, but the segmentation stage. Even if corrective adjustments of the labelling were performed by the radiologists, obtaining the perfect boundary of the tumour lesions remains a challenge.

In order to incorporate the lesion shape into the creation of the partition, the inverted image *f(x,y)* is multiplied by an isotropic Gaussian function centred at the seed point location (μ_x, μ_y) from stage 1, with a fixed variance σ_c^2 as the constraint function [15]. See section 5.2.1 for detailed explanation in isotropic Gaussian function. The isotropic Gaussian images are segmented by thresholding segmentation with a fixed value of threshold. The initial boundary is obtained in a number of steps as illustrated in Figure 5-9.

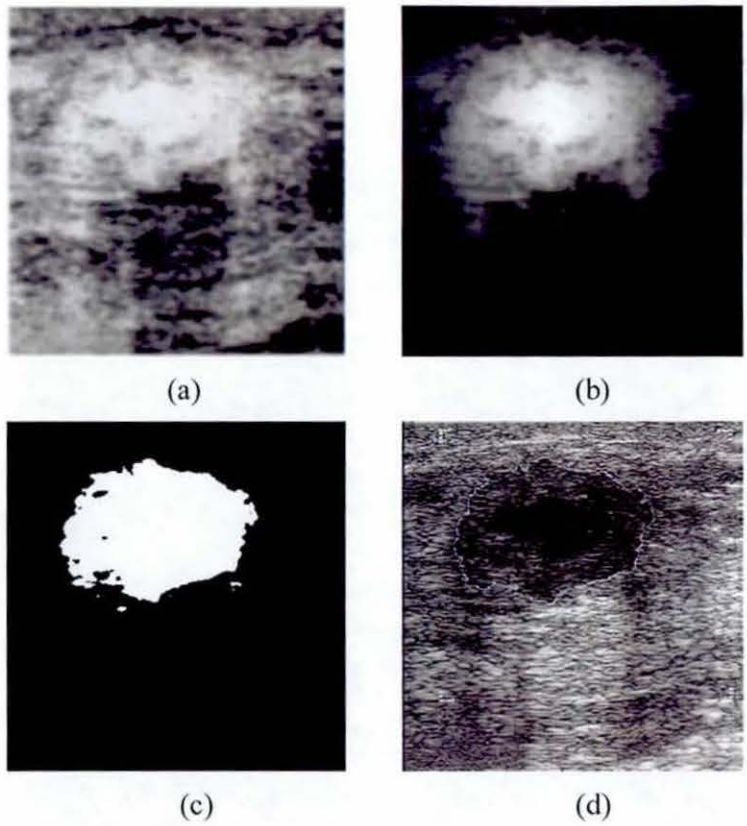


Figure 5-9 Output at various intermediate stages of stage 2: (a) Inverted image, (b) after applying Isotropic Gaussian on the inverted image, (c) after thresholding segmentation and (d) the final lesion boundary.

5.3.2.4 Active Contour Models

Deformable models, snakes, or Active contour models are curves defined within an image domain that can move under the influence of internal forces within the curve and external forces computed from the image data [61]. Yezzi et al. [62] applied the deformable model to segment cysts in ultrasound breast images. In Huang et al.'s [72] research, the initial contour was detected, followed by the implementation of an active contour model to improve the segmentation results.

The traditional active contour model consists of two key shortcomings. The first is that the initial contour must be close to the exact boundary, otherwise it may converge to the wrong result. The second problem is that active contours have difficulties processing into boundary concavities [101]. Xu and Prince [61] introduced the gradient vector flow (GVF) based approach, which they claimed can resolve the above problem. An edge map *f(x,y)* is defined by derived from the image *I(x,y)*. From their definition of GVF in [61], the gradient vector flow field is the vector field $v(x,y)=[u(x,y),\upsilon(x,y)]$ that minimizes the energy functional

$$\varepsilon(u,\upsilon)=\iint_{I}\mu(u_x^2+u_y^2+\upsilon_x^2+\upsilon_y^2)+|\nabla f|^2|v-\nabla f|^2\,dxdy \qquad (5\text{-}3)$$

where parameter μ is a regularization parameter governing the trade off between the first term and the second term in the integrand. μ should be set according to the amount of noise present in the image. Using calculus of variations, it has been shown that the GVF field can be found by solving the Euler Equations [61]:

$$\mu\nabla^2 u-(u-f_x)(f_x^2+f_y^2)=0 \qquad (5\text{-}4a)$$

$$\mu\nabla^2 \upsilon-(\upsilon-f_y)(f_x^2+f_y^2)=0 \qquad (5\text{-}4b)$$

where ∇^2 is the Laplacian operator. The details of the GVF concept and its implementation are well-explained in [61, 101]. The gradient vector flow (GVF) algorithm is a free source available in Internet provided by Xu and Prince [101]. We performed our experiments on the initial boundary using the above implementation. The enhancement of the boundary obtainable is clearly illustrated in figure 5-10.

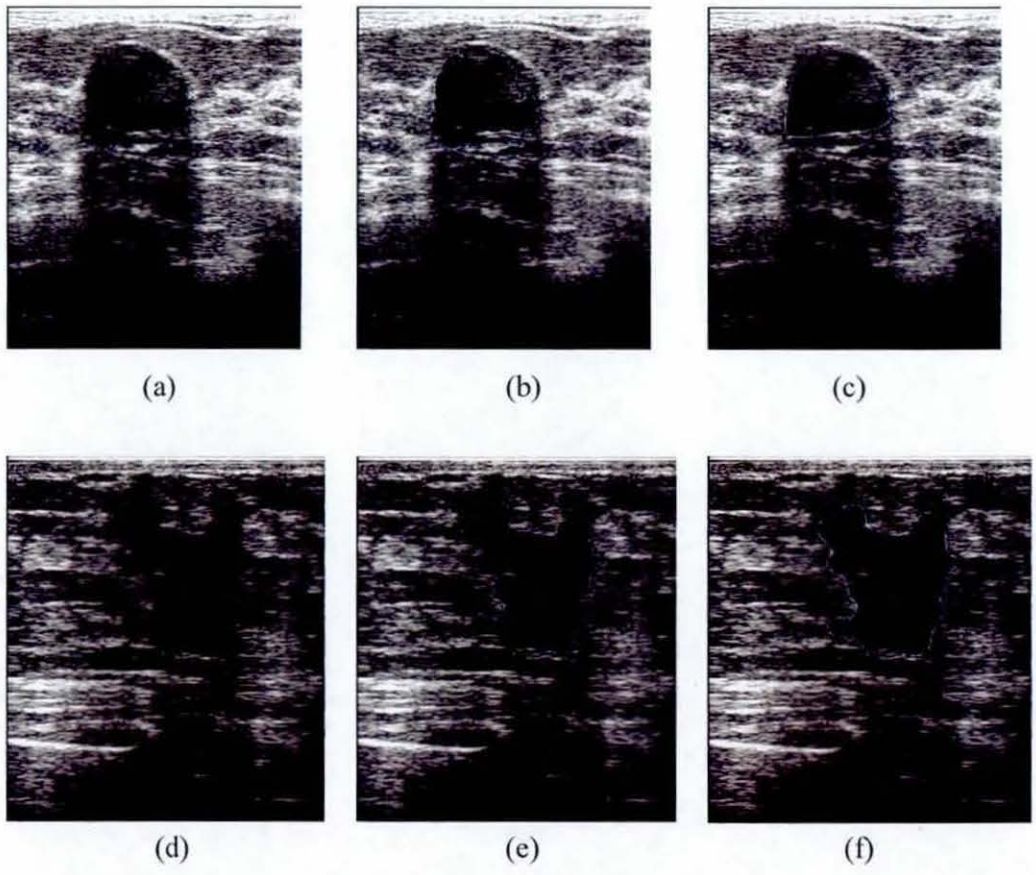


Figure 5-10 (a),(d) Original image (b),(e) Initial boundary (c),(f) Final boundary obtained by the gradient vector flow based approach.

5.3.3 Results & Analysis

In figure 5-11 we illustrate the results obtained when the proposed system is used. In figure 5-12 we compare the results of the proposed algorithm with that of Drukker et al.'s [14] algorithm. A visual comparison of the results illustrated in figure 12 (given the diagnostic results in the Breast Ultrasound CD [22]) demonstrates that the proposed method performs better as compared to the benchmark algorithm.

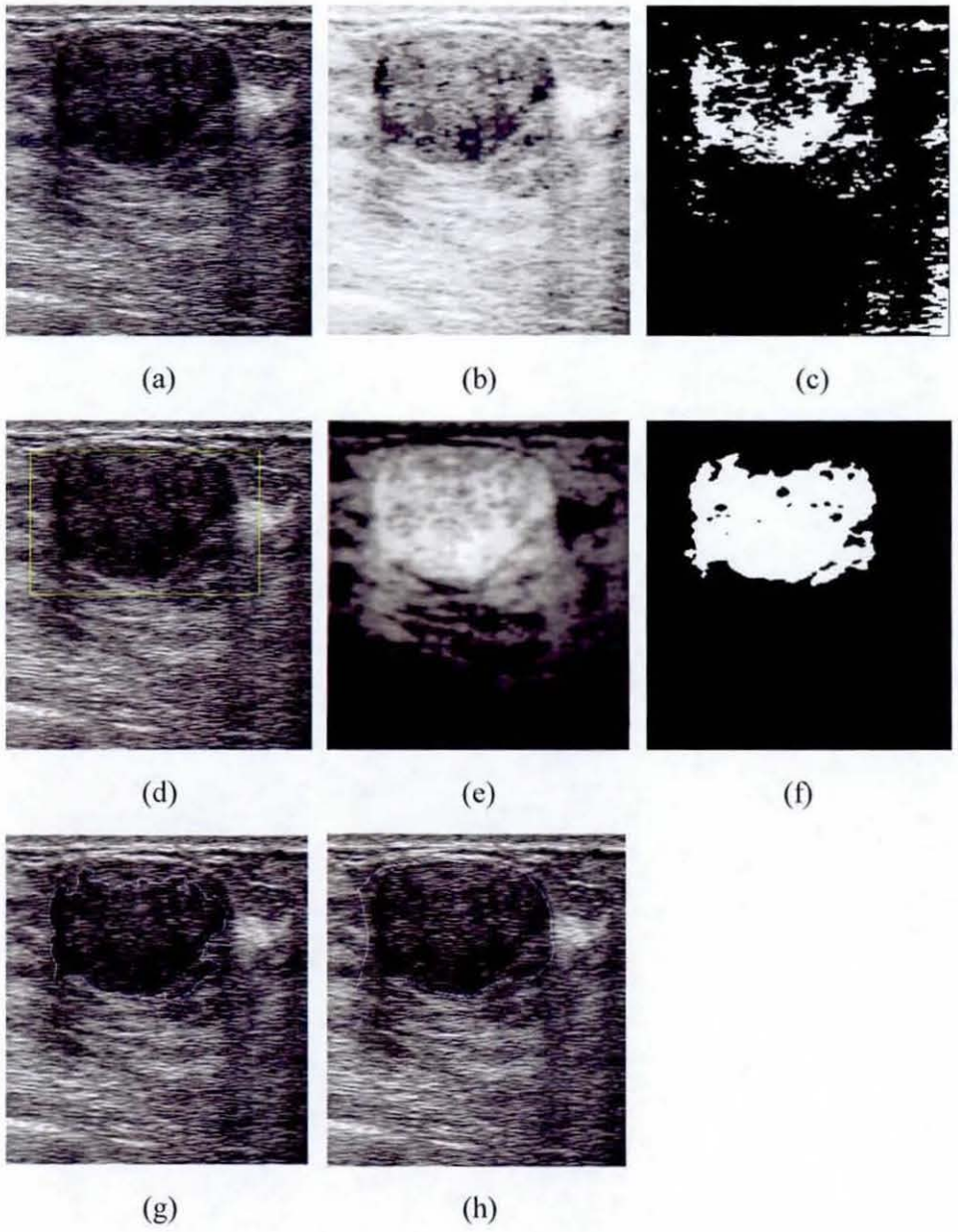


Figure 5-11 Output at various stages: (a) Original image. After; (b) histogram equalization, hybrid filtering and multifractal enhancement, (c) thresholding segmentation, (d) Labelling of region of interest (ROI), (e) Isotropic Gaussian on inverted US images, (f) thresholding segmentation (g), initial boundary and (h) the final detected boundary.

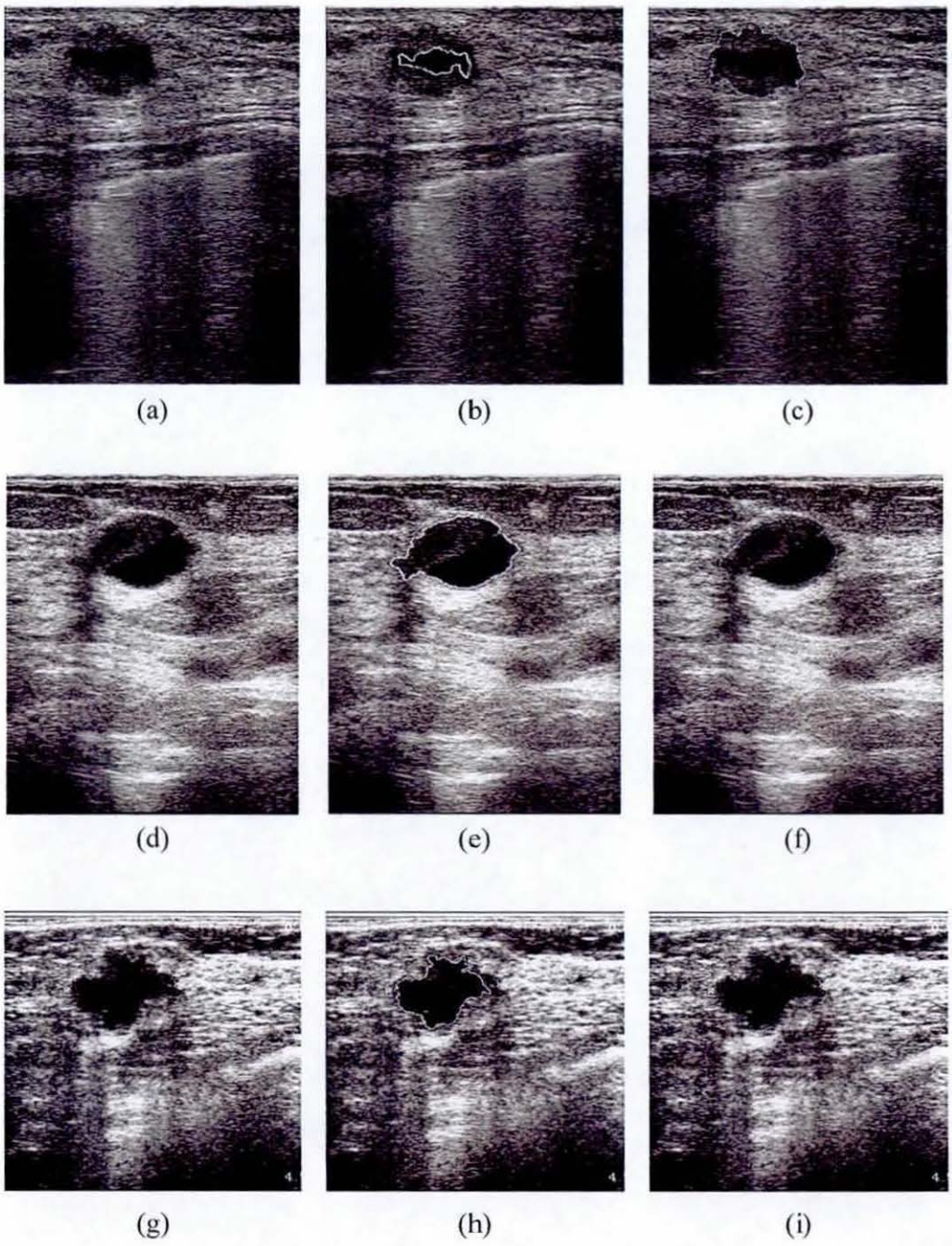
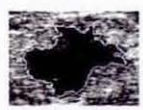
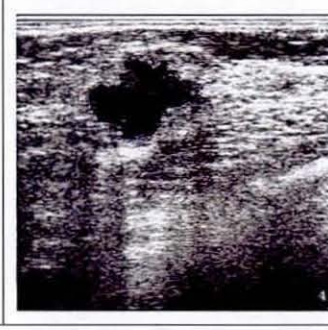


Figure 5-12 Visual performance comparison with benchmark algorithms. (a),(d),(g) The original images. (b),(e),(h) The result of Drukker et al.'s algorithm. (c),(f),(i) The result of the proposed Multifractal IsoGauss approach .

Table 5-1 summarises the novelty of the work presented. It is noted however that fully automated functionality will not completely replace the role of a radiologist in detecting tumours. Nevertheless it provides an alternative way to save the radiologist's time in picking up the suspected regions manually. Although the added capability will improve the functionality of a CAD system, the medical expert/radiologist will take the final decision as to whether or not a tumour has been detected. Another contribution of the work is that it preserves the local and global details. The local details are referred to as the density of the tumour, boundary, width, height, and size of the tumours, while the global details are referred to as the shadowing effects, hyper-echoic, hypo-echoic, and the speckles noise on the full images. The local details are important in the classification of tumours. However with experience some radiologists may learn to identify tumours by only using the global details. Hence, it is important to preserve the global details.

Table 5-1 Novelty of the multifractal IsoGauss approach

Operation	Current CAD methods	Proposed method
ROI Detection	Manually selected by radiologist	Fully automated selection
Segmentation/ Boundary Detection	The boundary is shown on the cropped ROI. Therefore decisions that would need the consideration of global image details cannot be made. The global dimension of the image (context) will also be lost. 	The boundary is detected from the full ultrasound images, which maintain the local details as well as global details. 

5.4 Summary & Conclusion

This chapter proposed two novel algorithms for the segmentation of lesions, based on multifractal analysis of processed ultrasound images. In general it was shown that further processing an ultrasound image in a multifractal dimension, can result in improved segmentation, which directly results in a refinement of the boundary of lesions.

In the first approach proposed, the initial lesions detected were first analysed in multifractal domain. Subsequently region growing segmentation was used to further refine the lesion boundaries identified by multifractal analysis. Our detailed experiments revealed that the direct application of region growing segmentation on initial lesion areas (i.e. ROIs) detected by the method proposed in Chapter-4, results in unwanted noisy areas that are not a part of the lesions being detected as parts of segmented lesions. Further, the presence of the multifractal stage before the region growing segmentation stage, results in smoother more accurate lesion boundary detections. The detailed experiments carried out to compare the proposed approach with two benchmark algorithms proved that the proposed approach performs significantly better than the benchmarks in detecting the lesion boundaries on different types of lesions, especially those of the more complicatedly shaped malignant tumours.

In the second approach proposed, the region growing segmentation stage of the first approach is replaced by the use of an Isotropic Gaussian function to obtain an initial contour of the lesion which is followed by the use of a special type of 'snake', i.e. the GVF to enhance the initial contours. We have shown that this approach to

segmentation can result in a fully automated approach that performs better than the existing state of the art method of fully automatic lesion segmentation proposed by Drukker et al. In future, an aim is to improve the proposed algorithm into a more adaptive algorithm, for instance, the approach should be able to automatically select the optimum parameter set for the snake model. Further, an investigation on the application of implicit active contour models, or level sets, in ultrasound images is proposed.

The accuracies of any subsequent lesion type classification algorithms are highly dependent on the accuracy of the lesion boundary detection algorithm used. Hence, the more accurate of the two novel approaches proposed for lesion segmentation, i.e., the second method based on IsoGauss processing and GVF analysis, is selected as the algorithm that is used to perform lesion boundary detection, which provides input to the classification algorithms investigated in the subsequent, Chapter-6.

Chapter 6: Classification

6.1 Introduction

This chapter aims to identify suitable feature-classifier combinations that will perform optimally for a given dataset of ultrasound images. It is noted that the American College of Radiology (ACR) has introduced a standard for the description of ultrasound images, the Breast Imaging Reporting And Data System (BIRADS) that aims to standardize the reporting of findings related to ultrasound images, clarifies interpretations and facilitates communication between clinicians (see Chapter-3, section 3.2.3). These BIRADS descriptors are widely used by radiologists in manual classification of lesions. Therefore a CAD based system that is capable of providing an automated description of US image lesions can provide vital input to BIRADS and its use by radiologists.

Previous research in US image lesion classification (see Chapter-3, section 3.2.3.2), have focused on finding the optimum feature set that can be used by a selected single classifier. This thesis evaluates the use of ten different classifiers and experimentally determines the optimum feature set for each classifier and make conclusions on the optimum feature-classifier combinations suitable for US image lesions. In particular the performance of a significant number of shape, texture and edge descriptors are compared with the aim of identifying the best feature set to be used in conjunction with the most efficient classifiers.

Initially a detailed study was carried out to evaluate the performance of two commonly used appearance based classification algorithms, based on the use of Principal Component Analysis (PCA) [102], and two dimensional linear discriminant analysis (2D-LDA) [103]. The study showed that these two appearance based classification approaches are not capable of handling the classification of ultrasound breast image lesions. Therefore further investigations in the use of a number of popular feature based classifiers, namely, Bayesian Network (BayesNet) [104], Naïve Bayes (NaiveBayes) [105], Support Vector Machine (SVM) [106], Multi Layer Perceptron (MLP) [104], Radial basis function Network (RBFNetwork) [104], Bagging [107], Adaptive Boosting (AdaBoost) [108], Logistic Regression Boosting (LogitBoost) [109], Random Tree [104], and Random Forest [110] were conducted with appropriate feature extraction phases used as pre-processing stages. Feature extractions stages are divided into three categories, namely shape, texture and edge based descriptors, and are investigated individually using the feature based classifiers. Due to the fact that a combination of features have a better ability to improve classification, detailed investigations into using various feature combinations based on the feature selection tools provided by the widely used WEKA data mining tool [104] is subsequently conducted. Note that in the above investigations, the default parameters of WEKA have been used. Finally a number of well known classification measures such as, success rate, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F-measure and area under ROC curve (A_z) are used to evaluate the performance of different feature set – classifier combinations, making conclusions on the optimum selections/combinations.

For clarity of presentation the chapter is divided into a number of sections: Section

6.2 discusses the use of appearance based approaches, i.e. PCA and 2D-LDA, in lesion classification. Section 6.3 discusses the use of feature-based approaches giving details of various feature detection, selection and classification algorithms. Section 6.4 provides experimental results and analysis giving detailed comparisons of the performance of various feature detection, selection and classification algorithms. Finally section 6.5 concludes the chapter.

6.2 Appearance Based Classification

This section presents two popular appearance based approaches to US lesion classification, namely PCA and 2D-LDA. It further provides information on the dataset used in the evaluation of the two approaches. For the preliminary study of the performance of PCA and 2D-LDA approaches, 40 images from the original test image database was chosen and resized to 128 pixel by 128 pixel. Figure 6-1 illustrates some of the resized images with the corresponding Eigen images.

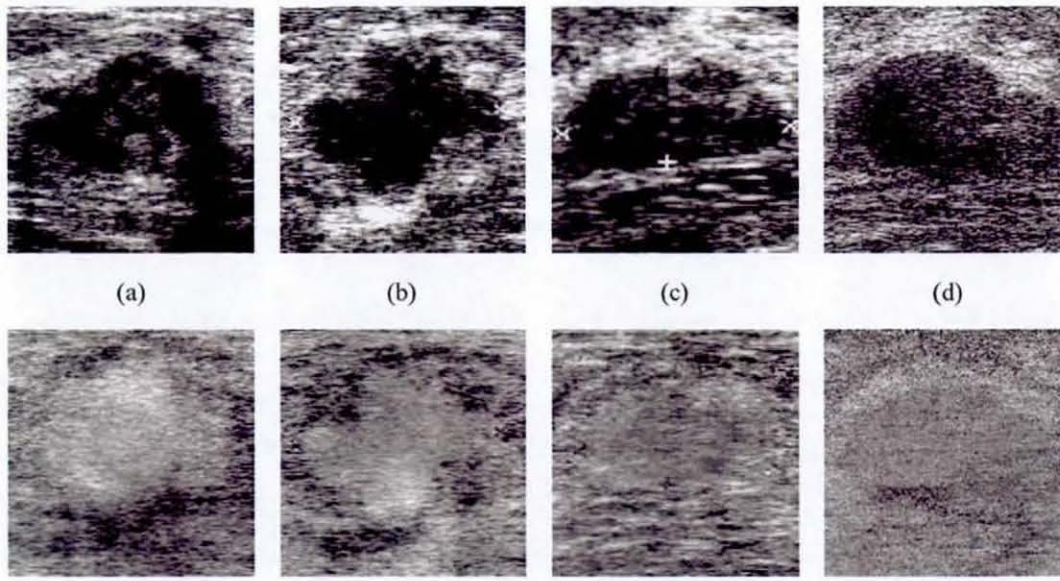
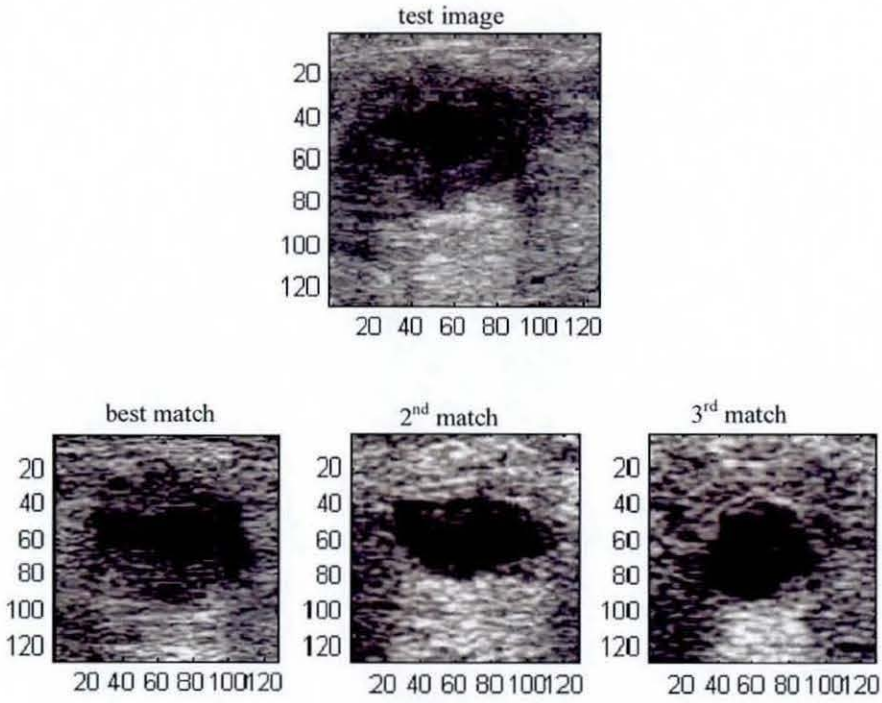


Figure 6-1 Resized test images (a-d) and their corresponding Eigen images (e-h)

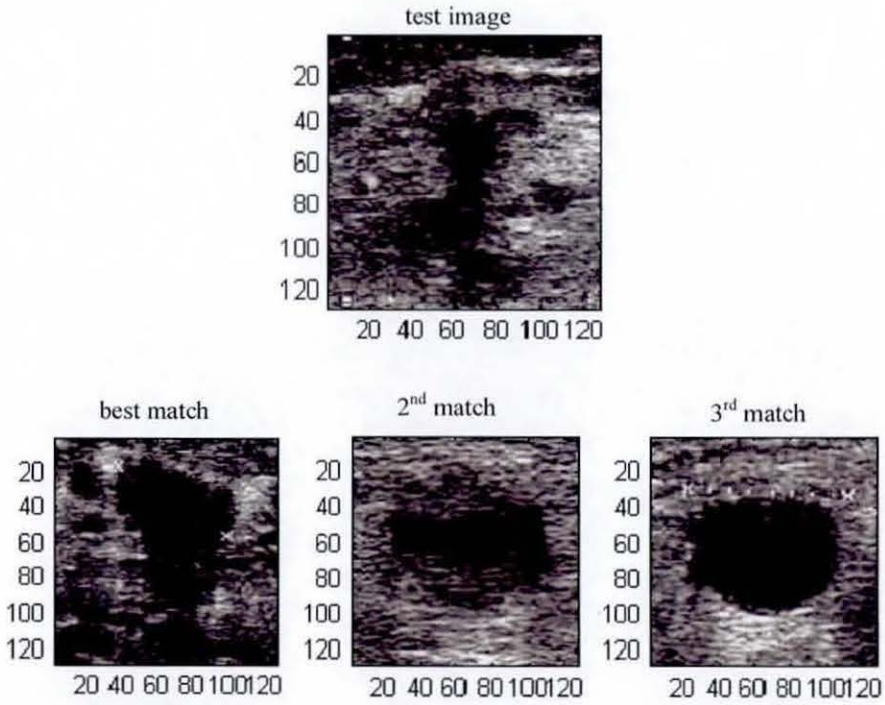
6.2.1 Principal Component Analysis (PCA)

PCA is a powerful tool that is widely used in data reduction. It is a popular unsupervised statistical method that can be used to find useful image representations [102]. PCA is the earliest appearance-based classification approach. It operates directly on the images or appearances of objects and processes the images as two dimensional holistic patterns. In face recognition, facial images are projected to a feature space which best describes the variation amongst known facial images [102]. This feature space is subsequently used in effective facial recognition. Further details of the theory and applications of PCA can be found in [102].

Figure 6-2(i) illustrates the success of the PCA classifier. It illustrates that the diagnosis is accurate. Figure 6-2(ii) illustrates the failure of classification of the PCA classifier. Following the approach proposed in [111] the performance of PCA based lesion classification was evaluated. It was observed that PCA performed poorly in the classification of ultrasound breast images. The results are illustrated in Figure 6-2. The overall success rate obtained was only 55%.

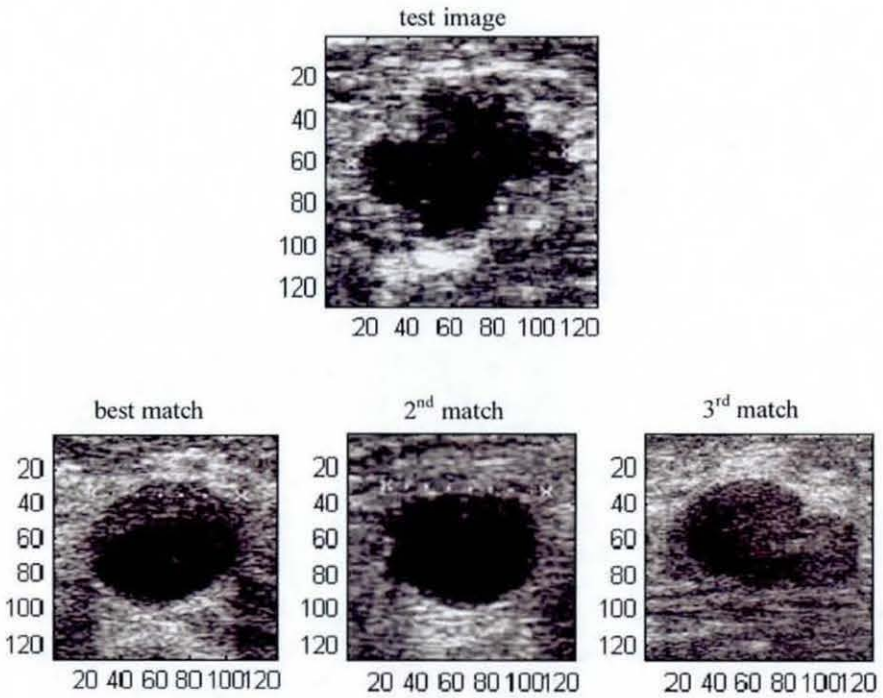


(a)

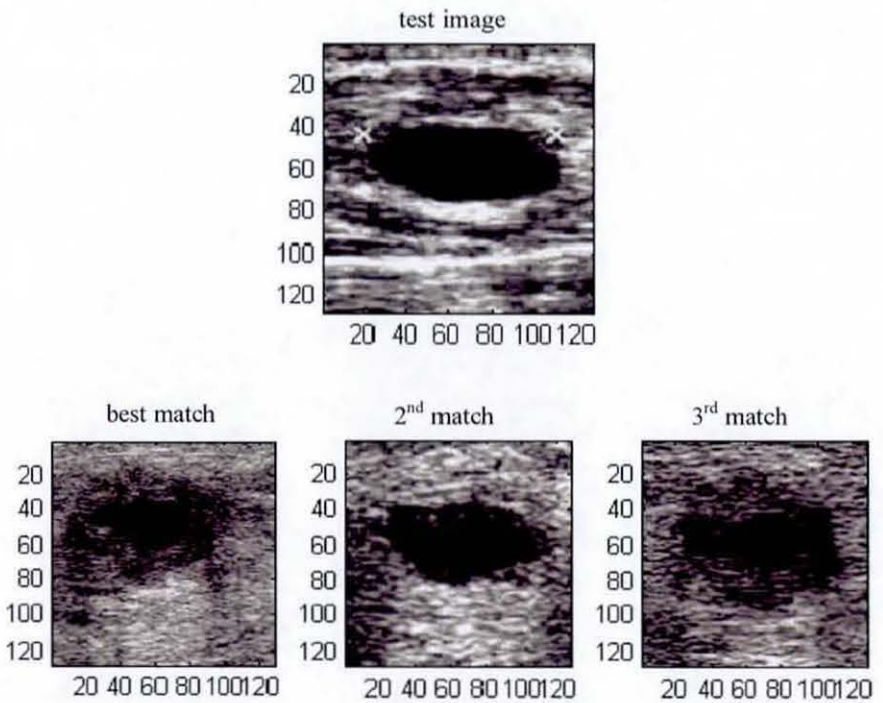


(b)

Figure 6-2(i) Successful classifications in PCA Classifier



(a)



(b)

Figure 6-2(ii) mis-classifications in PCA Classifier

6.2.2 Two Dimensional Linear Discriminant Analysis (2D-LDA)

LDA is a further appearance based approach that may be used in lesion classification. It optimizes the low-dimensional representation of image objects with the aim of most discriminant feature extraction [103]. In LDA, the image is linearly projected into a feature subspace. The projection method is based upon Fisher's linear discriminant and produces well separated classes in a low-dimensional subspace [103]. In pattern recognition research it has been shown that LDA based algorithms outperform PCA-based algorithms [112] in image classification. A variant of LDA, 2D-LDA (Two Dimensional – LDA), proposed by Li et al. [103], directly extracts features from the image matrix, rather than a feature vector created out of an image to compute the between-class and within-class scatter matrices. Interested readers are referred to [112] and [103] for further details of LDA and 2D-LDA. By using the same dataset as used with PCA, the success rate of the 2D-LDA approach to classification of lesions was reported to be 50%.

Therefore it can be concluded that both PCA and 2D-LDA based approaches, i.e. appearance based approaches, are inefficient in the classifications of lesions of US images.

6.3 Feature Based Classification

Feature based approaches to image classification generally consist of two stages, namely, feature extraction and feature classification. Often when the number of features extracted is very high, an additional intermediate stage of feature selection is used.

6.3.1 Features Extraction Approaches

Features can generally be used as a concise representation of an image. In particular feature extraction in ultrasound images is a challenging task due to the speckle noise. For ease of discussion and analysis, we classify the typical features of an ultrasound image into three groups, namely, shape, texture and edge descriptors. These categories are detailed in the following sub-sections.

6.3.1.1 Shape descriptors

Shape analysis has been widely utilised in computer vision and has been intensively developed over the past decades in both theoretical and applied domains [113]. Due to the fact that most US image lesions vary in shape according to their type (e.g. cyst, malignant etc.), shape descriptors have been used in their classification. This section discusses the shape descriptors utilised by previous methods of feature based US lesion classification and a number of further shape descriptors that have been utilised in general within other computer vision applications [112]:

Area to Perimeter ratio (AP): $AP = \frac{area}{perimeter}$, where area is the area of the tumour, and perimeter is the perimeter of the tumour. Tumour perimeter and tumour area are dependent on size of a tumour. Hence as a standalone feature, it does not provide useful information. However, the aspect ratio of the area to perimeter will provide useful information as a malignant tumour has an irregular boundary, which implies a higher value in perimeter.

Convex hull provides important information in the description of shape. Figure 6-3 illustrate the shape of the lesions that correspond to their convex hulls. From the above analysis, the convexity and solidity of the lesions can be calculated as follows:

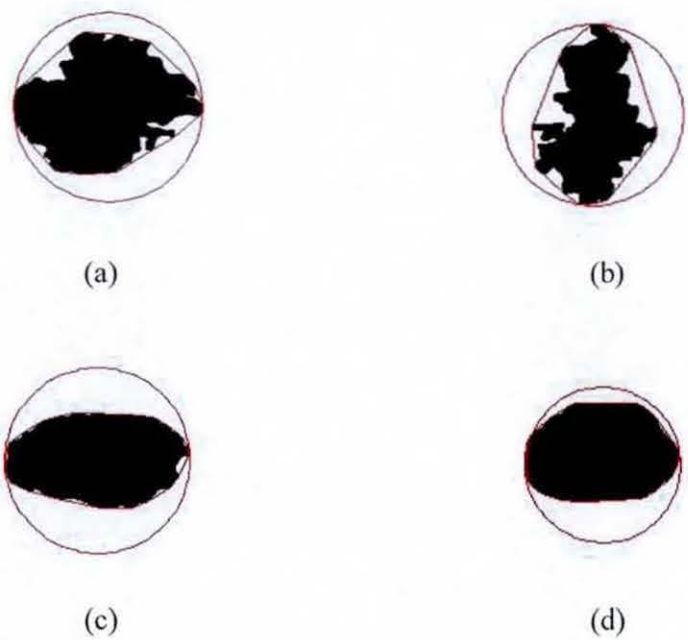


Figure 6-3 Lesions and Convex hull. (a), (b) malignant tumours, (c), (d) benign tumours.

Solidity: $Solidity = \frac{area}{convex\,area}$, where *area* is the area of the tumour region, and *convexarea* is the area of the convex hull. It is clearly shown that the malignant

tumours (as in Figure 6-3 (a) and (b)) has a larger convex hull area as compared to the benign tumours (as in Figure 6-3 (c) and (d)).

Convexity: $Convexity = \frac{convex\ perimeter}{perimeter}$, where *perimeter* is the perimeter of the tumour, and *convex perimeter* represents the perimeter of the convex hull. Figure 6-3 clearly illustrates that a malignant tumour has a higher perimeter value as a result of its irregular nature. Hence, this feature is fairly important in boosting the separability of the lesions.

Elongation: $Elongation = \frac{length}{width}$, where *length* and *width* denotes the length and width of the bounding rectangle as illustrated in the following figure 6-4.

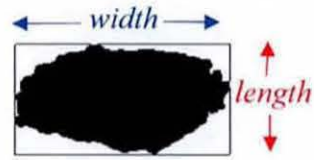


Figure 6-4 Illustration of the bounding rectangle with it's width and length.

Extent: $Extent = \frac{area}{area\ of\ bounding\ \text{rectangle}}$, where area is the *area* of the tumour, and the *area of bounding rectangle* is the area of the rectangle as illustrated in figure 6-4.

Aspect ratio (AR): $\text{AR} = \frac{\text{major axis}}{\text{minor axis}}$, where *major axis* is the line passing through the foci, centre and vertices of the ellipse, and *minor axis* is a line through the centre of an ellipse which is perpendicular to the major axis, as illustrated in figure 6-5.

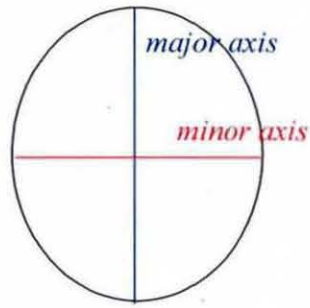
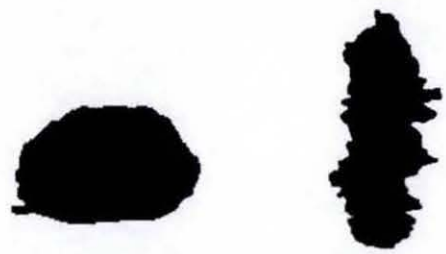


Figure 6-5 illustration of the *major axis* and *minor axis*

Compactness: $Compactness = \frac{\sqrt{4\pi \times area}}{major\ axis}$, where *area* is the area of the tumour, and *major axis* as defined in Figure 6-5.

Roundness: $Roundness = \frac{4 \times area}{\pi \times Max_Diameter^2}$, where *area* denotes the area of the tumour, and *Max_Diameter* denotes the length of the *major axis*. Figure 6-6 illustrates the roundness values for benign and malignant tumours. Note that a benign tumour has a higher roundness value as compare to a malignant tumour.



(a) (b)

Roundness=0.65 Roundness=0.37

Form factor=0.74 Form factor=0.32

Figure 6-6 Illustration of the shape of a: (a) benign tumour, (b) malignant tumour with the respective roundness and form factor values.

Form factor: $Form\ factor = \frac{4\pi \times area}{\sqrt{perimeter}}$, where *area* and *perimeter* denotes the area and the perimeter of the tumour. Figure 6-6 shows that a benign tumour has a higher form factor value than a malignant tumour.

Max/Min radii: $Max/Min\ radii = \frac{Maximum\ radius\ from\ the\ circle's\ centre}{Minimum\ radius\ from\ the\ circle's\ centre}$. The detailed explanation of this parameter is illustrated in figure 6-7.

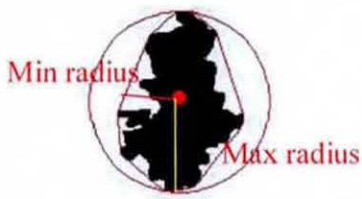


Figure 6-7 Illustration of maximum radius and minimum radius.

6.3.1.2 Texture descriptors

Texture descriptors can be used to differentiate between lesions having differences in internal texture. Figure 6-8 illustrates that texture within tumours vary from benign to malignant types.

(a) (b) (c)

Figure 6-8 Texture within contours of the detected lesions. (a) Cysts, (b) Fibroadenoma (c) Malignant

The texture can be represented in the form of four central moments and by entropy of a pixel value distribution:

Mean/Average Intensity: This refers to the first central moment, i.e. the average intensity of the lesion, in grey-scale.

Standard Deviation: This refers to the second central moment, i.e. a measure of the sparsity of the pixel value intensity within a lesion.

Skewness: This refers to the third central moment, i.e. a measure of the degree of asymmetry of the pixel value distribution.

Kurtosis: This refers to the normalised form of the fourth central moment of a distribution, or the degree of peakedness of a distribution. It measures if the data are peaked or flat relative to a normal distribution.

Minimum Cross Entropy (MCE): The cross entropy of a probability distribution q with respect to a prior distribution p is defined by

$$H(q,p) = \sum_j q_j \log \frac{q_j}{p_j}$$

The idea of *MCE* is to choose the distribution q that has the least cross entropy, with respect to the given prior p [114]. In classifying benign tumours from malignant tumours, a particular challenge met is that typically the pixel intensity value distribution of a fibroadenoma lesion is close in resemblance to pixel intensity value distribution of a normal lesion. Therefore if a normal lesion acts as the prior

distribution p, it is expected that fibroadenoma lesions will have the least *MCE* as compared to the *MCE*s that results from other types of diagnosis. Hence, *MCE* can be used to separate fibroadenoma lesions from lesions of the benign class.

6.3.1.3 Edge descriptors

Polar coordinate analysis, Fourier descriptors, and fractal analysis are widely used in literature as edge descriptors. The following sections define these parameters and summarises their potential use in classification.

Polar Coordinate Analysis: The edge irregularity can be measured by using polar coordinate analysis. However, our observations of the characteristic features of datasets revealed that while some lesions can be analysed perfectly (as in Figure 6-9 (a)) some cannot be analysed by polar coordinate analysis as they consist of multiple edge points on a given diagonal (as in Figure 6-9 (b)).

(a) (b)

Figure 6-9 (a) single point (b) multiple points.

Fourier descriptor: Fourier descriptors are commonly used in pattern recognition applications due to their invariance to the starting point of the boundary and rotation [115].

Figure 6-10 illustrates the efficiency of Fourier descriptors in differentiating malignant tumours from benign tumours. In low frequency domain, it provides the general shape information of an edge, while the details of the edges are represented by the relevant high frequency domain. Therefore it is expected that malignant tumours, with irregular edges, produce higher values in high frequency domain as compared to benign tumours. The Fourier descriptors of a lesion boundary are presented by an array of complex numbers which correspond to the pixels of the object boundary if the image is placed in the complex plane. Fourier descriptors are calculated by combining Fourier transform coefficients of the complex array [115]. It is noted that using a lower number of Fourier descriptors, the general shape can be described (figure 6-10(a) and 6-10(b), 10 descriptors). The detailed information of the shape is described by using the high frequency components of the series (as shown in figure 6-10(a) and figure 6-10(b), 50 descriptors). For consistency, in our preliminary experiments the boundaries are re-sampled to the same size and 100 descriptors are obtained for each. Note that in our experiments the descriptor *Fourier1* represents the summation of the coefficients of real numbers while descriptor *Fourier2* represents the summation of the coefficients of the complex numbers.

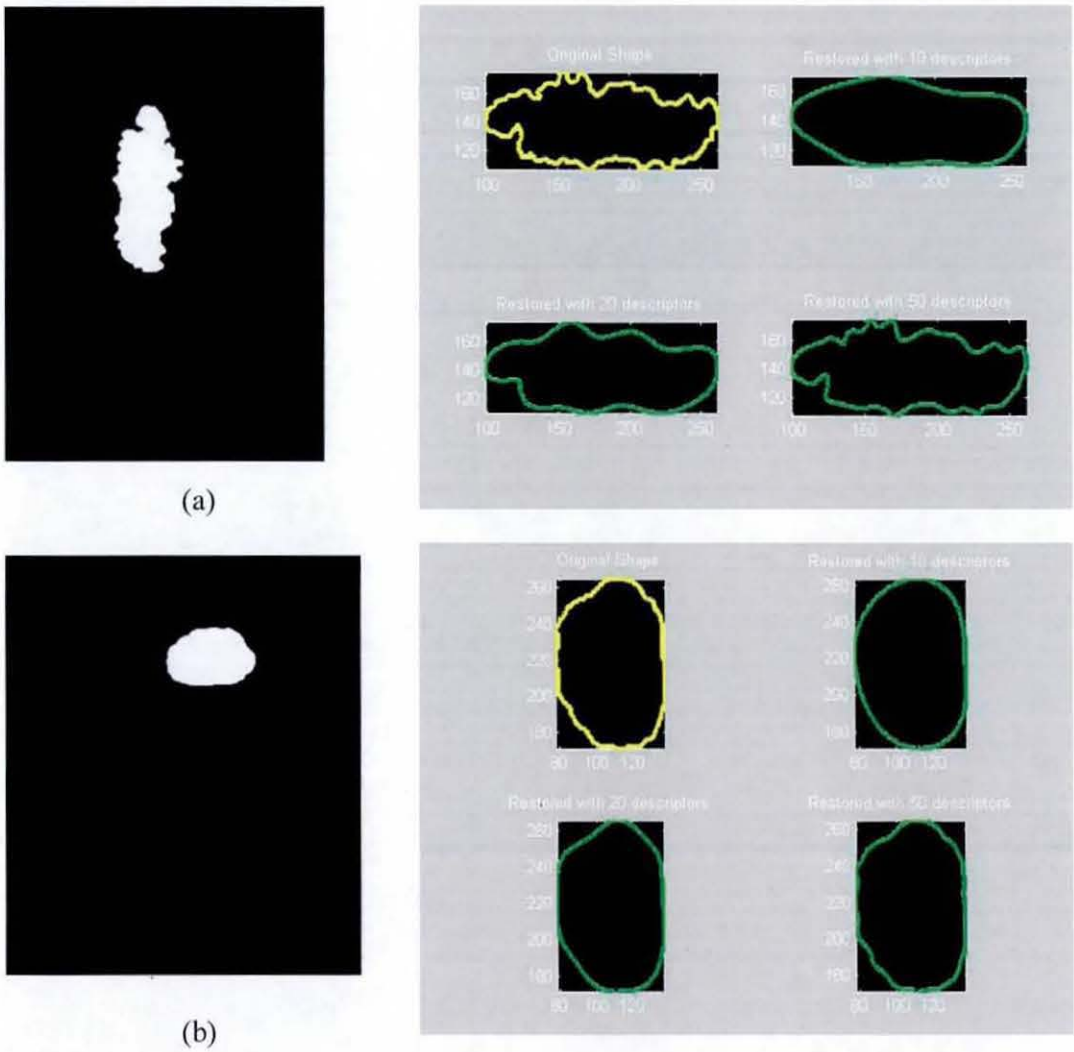


Figure 6-10 Fourier Descriptors for a (a) malignant tumour and a (b) benign tumour, and their corresponding reconstructions of edges using different numbers of descriptors/coefficients.

Fractal Dimension: Chapter-4, section 4.7.2.2 detailed the explanation of fractal analysis. The simple box-counting method is used in extracting the edge information. In the other words, fractal dimension can be used to describe the roughness of the edges. Note that in our experiments, the descriptor, *Fractal1* represents the mean of the fractal dimension, while *Fractal2* represents the standard deviation of the fractal dimension.

6.3.2 Classifiers

There are many choices for an appropriate lesion classifier in US imaging ranging from Bayes classifiers to neural networks. To further investigate the reliability of different feature sets, experiments with different classifiers were conducted. The classification stage is experimented by using WEKA (The Waikato Environment for Knowledge Analysis) [104], the popular, open source, Java based data mining tool. WEKA is a collection of implementations of popular machine learning algorithms for data mining tasks [104] which includes the implementation of a number of different classifiers. WEKA is still in active development (version 3.5.8) at the time of publication of this thesis. Ten classifiers in WEKA were chosen for investigation, namely, BayesNet, NaiveBayes, SVM, Multi Layer Perceptron (MLP), Radial basis function networks (RBF Networks), Bagging, Adaptive Boosting (AdaBoost), LogitBoost, Random Tree, and Random Forest due to their general popularity in a wide range of other applications. A summary of explanation of each of the above classifiers can be presented as follows:

BayesNet: Bayesian approach is based on the probability theory [104]. Given the probability distribution, a Bayes Classifier can achieve optimal results.

NaiveBayes: Naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naïve) assumptions of independence [105]. It is also known as an "independent feature model". For further reading, readers are referred to [105].

SVM: The support vector machines (SVMs) were introduced by Vapnik [106] and are based on statistical learning theory. In pattern classification applications, it has been proved that SVMs provide better generalization performance than the traditional techniques, such as neural networks [104]. The benefits of SVMs include, rapid and excellent classification capability [13] and the ability to generalize in high-dimensional spaces. Hence it is widely accepted as an excellent choice in classification. In addition, the task of classifying malignant tumours from benign tumours can be defined as a so-called binary classification problem. According to [106], SVM is powerful in solving binary classification problems. This thesis implements John Platt's [116] sequential minimal optimization (SMO) algorithm for training a support vector classifier. For further discussions on SVM, readers are referred to [116].

MLP: MLP is a type of artificial neural network that uses back propagation to classify instances. Refer to [104] for further reading.

RBFNetwork: WEKA implements a normalized Gaussian radial basis function network (RBFNetwork). It uses the *k*-means clustering algorithm to provide the basis functions and learns either a logistic regression (discrete class problems) or linear regression (numeric class problems) on top of that. Symmetric multivariate Gaussians are fit to the data from each cluster. Refer to [104] for further details.

Bagging: Bagging also known as bootstrap aggregating, is a meta-algorithm to improve machine learning of classification and regression models in term of stability and classification accuracy [107]. Refer to [107] for further details.

AdaBoost: AdaBoost is a boosting method based on Freund et al.'s [108] original work. This meta-classifier is adaptive in the sense that subsequent classifiers built are tweaked in favour of those instances misclassified by previous classifiers [108].

LogitBoost: LogitBoost is a boosting algorithm formulated by Friedman et al. [109]. It casts the AdaBoost algorithm into a statistical framework.

Random Tree: With k random features at each node, a Random tree is a tree drawn at random from a set of possible trees [104]. In this context "at random" means that each tree in the set of trees has an equal chance of being sampled. Another way of saying this is that the distribution of trees is "uniform". Random trees can be generated efficiently and the combination of large sets of Random trees generally leads to accurate models. Random tree models have been extensively developed and used in the field of Machine Learning in the recent years.

Random Forest: A Random Forest is a meta-learner that comprises of many individual trees. They are designed to operate quickly over large datasets and to be diverse by using random samples to build each tree in the forest. Random forest [110] is an ensemble of unpruned classification or regression trees, induced from bootstrap samples of the training data, using random feature selection in the tree induction process. For further reading refer to [110].

6.3.3 Feature Selection

Although a large number of features can be derived from an image, but not all the features are suitable for classification, some of the irrelevant features will increase complications and can reduce the accuracy of the classifier to be used. Hence, it is important to filter out the irrelevant features, or remove the features that reduce the accuracy of a given classifier. Feature selection is to select a subset of relevant features to build robust classifiers based on measures of separability constructed from the training set [117].

Exhaustive search is the most straightforward and the earliest feature selection approach, which investigates all possible combinations of subsets. However as the number of subsets grow, exhaustive search becomes impractical. BestFirst [104] searches the space of attribute subsets by greedy hillclimbing [104] augmented with a backtracking facility. Linear Forward Selection is an extension of BestFirst, which takes a restricted number of k attributes into account. For further reading, refer to [118]. Genetic Search performs a search using a simple genetic algorithm described in Goldberg [119]. Greedy Stepwise [104] algorithm performs a greedy forward or backward search through the space of attribute subsets. Greedy stepwise algorithm can move either forward or backward through the search space, while Best First algorithm has the option to consider both at any given point in the search. However the Greedy stepwise algorithm is faster but tends to generate errors.

This thesis implements feature selection using WEKA software [104]. WEKA has implementations of a range of feature selection methods. Correlation-based Feature

Subset Selection (CfsSubsetEval class in WEKA) evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with a degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. For further explanations, please refer to [120]. The classifier-based feature selector, namely ClassifierSubsetEval [120], uses a classifier to estimate the "merit" of a set of attributes.

6.4 Experimental Results and Analysis

In line with experiments conducted in relevant literature [8, 41, 68-70, 122], in our experiments, the dataset is divided equally into a training sub-set and a testing sub-set. The training set is used to build the classifier models, while the testing set is used for evaluation of the models.

The measurements used to evaluate the performance includes accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and the predicted A_z value, as implemented by the software tool WEKA. The measure of accuracy is defined as the percentage of correct classifications obtained, which is not suitable as a sole performance measurement measure in lesion classification, for e.g., terms such as false positives can play an important role in the suitability of a classification algorithm [Note: Chapter-3, section 3.2.4 defines explain the remaining statistical measures].

In order to use a single measurement for comparison, most of the medical research projects have used A_z value. However though A_z value is a better single measure of performance as compared to using *accuracy*, due to the prediction approach adopted by WEKA to calculate the A_z value, our experiments revealed that its use is far from being perfect. For binary classifiers, a number of authors of previous literature have used the *success rate* as a single performance metric. It is defined as:

$$success\ rate = \frac{TP + TN}{TP + FP + TN + FN}$$

where, TP (true positives), TN (true negatives), FP (false positives) and FN (false negatives) are defined as in Chapter-3, section 3.2.4.

A further single measure that has been used popularly in characterizing performance of is the F-measure [104], which can be defined as:

$$\text{F-measure} = \frac{2 \times recall \times precision}{recall + precision} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

Where, *recall* and *precision* is defined as in Chapter-3, section 3.2.4.

Due to the fact that a number of different single performance metrics have been used in previous literature for performance characterisation, in this thesis we use all popular measures with the aim of identifying the best metrics for evaluating algorithms tested on ultrasound images.

For clarity of presentation the performance is analysed under a number of sub headings in the following sections. Section 6.4.1 analyses the use of different

performance metrics defined above when using SVM as the classifier and different selections of feature descriptors. Section 6.4.2 continues to analyse the remaining classifiers as defined and implemented by the WEKA tool. Section 6.4.3 analyses different feature selection algorithms as defined and implemented by the WEKA software tool.

6.4.1 Performance evaluation of the SVM classifier

It is important to study the use of individual features in classification before the classification is conducted using combinations of features. This approach will provide a rough idea on how single features perform and how different feature combinations may potentially perform. Figure 6-11 illustrates the performance characteristics when a number of single features are used in classification. It shows that the lesion classes are not ideally separable using any of the single features.

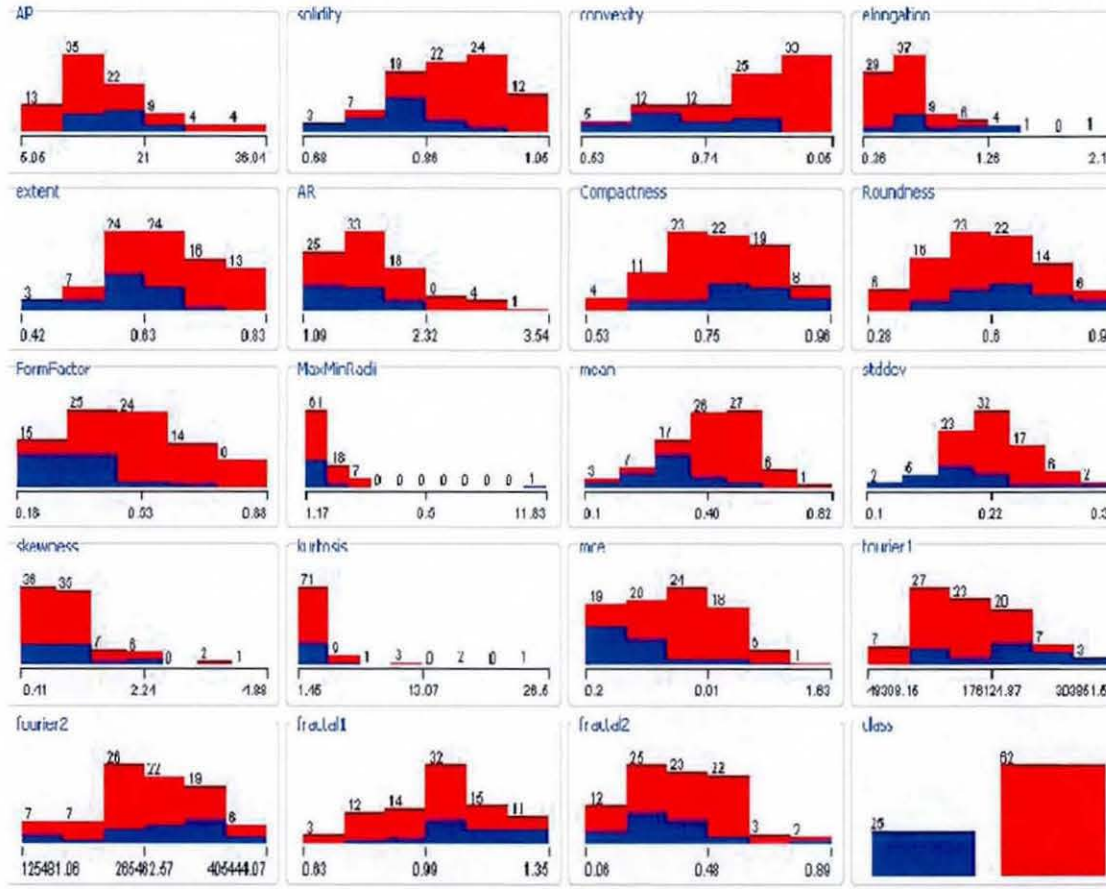


Figure 6-11 Analysis of the use of single different features in classification. The blue and red areas represent malignant and benign classes, respectively. The x-axes represents the value of each parameter represented in feature extraction.

Support Vector Machines (SVM) have been popularly used as a classifier in existing literature [13, 72, 73, 121] due to its general accuracy in classification as compared to other classifiers. Hence we have selected it as the classifier to investigate the use of single features in classification. In table 6-1 the classification metrics accuracy, sensitivity, specificity and A_z obtained when using individual features with SVM as the classifier, is tabulated. It shows that the feature convexity, provides the best separability to the classes as it consistently gives a clear maxima (or minima) for different measures tested. Note the use of additional metrics, positive predictive value (PPV) and negative predictive value (NPV), see section 3.2.4 for the definition of

PPV and NPV.

Table 6-1. Summary statistics in using single feature in SVM based classification

Features	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
(1) AP	0.713	0	1	0	0.713	0	0.500
(2)Solidity	0.839	0.52	0.968	0.867	0.833	0.650	0.744
(3)Convexity	0.862	0.6	0.968	0.882	0.857	0.714	0.784
(4)Elongation	0.724	0.04	1	1	0.721	0.077	0.520
(5)Extent	0.724	0.04	1	1	0.721	0.077	0.520
(6)AR	0.713	0	1	0	0.713	0	0.500
(7)Comp	0.713	0	1	0	0.713	0	0.500
(8)Round	0.713	0	1	0	0.713	0	0.500
(9)Form Factor	0.724	0.04	1	1	0.721	0.077	0.520
(10)Max/min radii	0.713	0	1	0	0.713	0	0.500
(11)Mean	0.713	0	1	0	0.713	0	0.500
(12)Std dev	0.713	0	1	0	0.713	0	0.500
(13)Skewness	0.713	0	1	0	0.713	0	0.500
(14)Kurtosis	0.713	0	1	0	0.713	0	0.500
(15)mce	0.713	0	1	0	0.713	0	0.500
(16)Fourier1	0.713	0	1	0	0.713	0	0.500
(17)Fourier2	0.713	0	1	0	0.713	0	0.500
(18)Fractal1	0.713	0	1	0	0.713	0	0.500
(19)Fractal2	0.713	0	1	0	0.713	0	0.500

Further experiments were performed to evaluate the performance of SVM based classification when all shape, texture and edge descriptors are used as individual groups. Table 6-2 tabulates the results. An important observation is that the use of shape descriptors consistently performed best with all metrics used for performance evaluation. A further observation is that the use of only the shape descriptors is not more efficient than the use of all features. This provides a hint as to the fact that

optimal feature combination may consist of shape descriptors plus a number of other non-shape descriptors. This implies that feature selection is crucial in obtaining the optimal performance of any classification algorithm. This aspect is further investigated in section 6.4.3.

Table 6-2 Summary statistics in using different groups of features in SVM based classification

Features	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
All features	0.931	0.840	0.968	0.913	0.938	0.875	0.904
Shape descriptors	0.931	0.800	0.984	0.952	0.924	0.870	0.892
Texture descriptors	0.851	0.680	0.919	0.773	0.877	0.723	0.800
Edge descriptors	0.712	0	1	0	0.713	0	0.500

6.4.2 Performance evaluation of other classifiers

This section compares the performance of other popular classifiers implemented within the WEKA software tool, when using a combination of 19 different features (see table 6-1 for a list) and using different metrics for the performance evaluation.

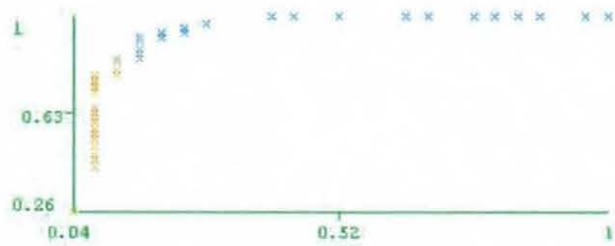
Table 6-3 Summary statistics for different classifiers implemented within WEKA when using all of the 19 features

Classifier	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
BayesNet	0.851	0.880	0.839	0.688	0.945	0.772	0.914
NaiveBayes	0.897	0.880	0.903	0.786	0.949	0.830	0.944
SVM	0.931	0.840	0.968	0.913	0.938	0.875	0.904
MLP	0.897	0.720	0.968	0.900	0.896	0.800	0.944
RBFNetwork	0.908	0.880	0.919	0.815	0.950	0.846	0.955
Bagging	0.839	0.720	0.887	0.720	0.887	0.720	0.899
AdaBoost	0.885	0.720	0.952	0.857	0.894	0.783	0.940
LogitBoost	0.885	0.800	0.919	0.800	0.919	0.800	0.891
Random Tree	0.839	0.760	0.871	0.704	0.900	0.731	0.815
Random Forest	0.862	0.800	0.887	0.741	0.917	0.769	0.926

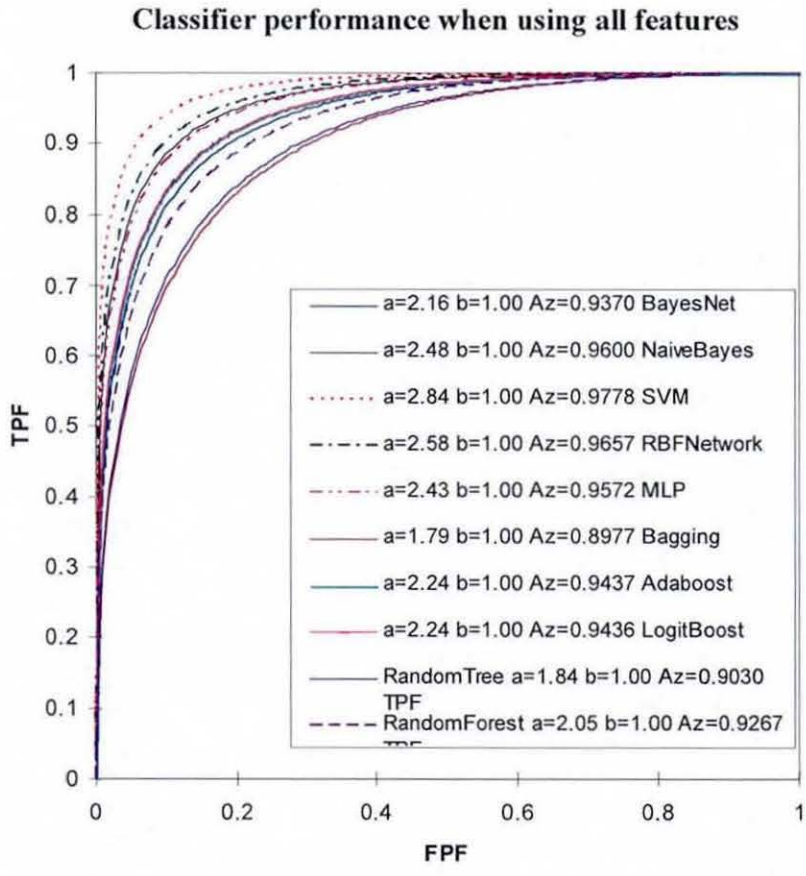
Table 6-3 shows that the 19 features selected provides a good feature combination to achieve consistently high A_z value, success rate, sensitivity and specificity measures in classification. In particular the A_z value illustrates a clear high value for all classifiers. It is also noted that SVM classifier resulted in the highest reading when using four of the eight measures used for evaluation, including the *F-measure* and *success rate*, proving reasons for its popularity in general classification tasks. The results tabulated in table 6-3 proves that classification can be reliably performed using the combination of features, providing proof that the boundary detection step of lesion as proposed in Chapter-5 has provided accurate, classifiable results.

Figure 6-12(b) illustrates the fitted ROC curves obtained when using ten different classifiers, further proving the best performance provided by the SVM classifier. Please note that in binary classification, A_z is a predicted value and therefore depends on the technique used to fit the ROC curve. For example, the predicted A_z values

generated by WEKA and shown in table 6-3 are different from the predicted A_z values generated by ROCKIT [111] and shown in Figure 6-12(b). A visual comparison of the predicted curve for BayesNet is illustrated in figure 6-12(a) and figure 6-12(b).



(a)



(b)

Figure 6-12 Fitted ROC curve shows the performance of ten classifiers with 19 features.

In the following section, we analyse the performance of different classifiers when using all features belonging to different groups namely, shape, texture and edge feature groups. The purpose of this study is investigate towards the possibility of identifying the optimum feature set for different classifiers.

Table 6-4 compares the performance of different classifiers when using all shape based feature descriptors. The results show that the RBFNetwork has the highest *success rate* (i.e. 0.931), *sensitivity*, *specificity* and *F-measure*. SVM performs equally well when using *success rate* as a metric and from *F-measure*, RBFNetwork performs the best.

Table 6-4 Summary statistics for different classifiers when using a combination of all shape features

Classifier	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
BayesNet	0.839	0.84	0.839	0.677	0.929	0.750	0.894
NaiveBayes	0.897	0.88	0.903	0.786	0.949	0.830	0.942
SVM	0.931	0.8	0.984	0.952	0.924	0.870	0.892
MLP	0.897	0.8	0.935	0.833	0.921	0.816	0.928
RBFNetwork	0.931	0.92	0.935	0.852	0.967	0.885	0.926
Bagging	0.862	0.76	0.903	0.76	0.903	0.760	0.898
AdaBoost	0.851	0.76	0.887	0.731	0.902	0.745	0.913
LogitBoost	0.874	0.84	0.887	0.75	0.932	0.792	0.935
Random Tree	0.931	0.84	0.968	0.913	0.938	0.875	0.904
Random Forest	0.828	0.68	0.887	0.708	0.873	0.694	0.901

Table 6-5 compares the performance of different classifiers when using all texture based feature descriptors. Results show that SVM performs best when using *success rate*, *specificity* and *F-measure*. However overall, the combination of texture features performs worse as compared to the combination of shape features, indicating that shape descriptors will play a more important role in lesion classification tasks as compared to texture based features. This is further discussed in section 6.4.3 under feature selection

Table 6-5 Summary statistics for different classifiers when using a combination of all texture features

Classifier	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
BayesNet	0.747	0.76	0.742	0.543	0.885	0.633	0.795
NaiveBayes	0.724	0.880	0.661	0.512	0.932	0.647	0.861
SVM	0.851	0.680	0.919	0.773	0.877	0.723	0.800
MultiLayerPerceptron	0.793	0.76	0.806	0.613	0.893	0.679	0.870
RBFNetwork	0.747	0.680	0.774	0.548	0.857	0.607	0.825
Bagging	0.701	0.800	0.661	0.488	0.891	0.606	0.834
AdaBoost	0.759	0.680	0.790	0.567	0.860	0.618	0.714
LogitBoost	0.632	0.480	0.694	0.387	0.768	0.429	0.718
Random Tree	0.644	0.480	0.710	0.400	0.772	0.436	0.595
Random Forest	0.770	0.720	0.790	0.581	0.875	0.643	0.777

Table 6-6 compares the performance of different classifiers when using all edge based feature descriptors. The results illustrate that none of the classifiers performs consistently well when using different metrics and the overall performance is significantly poor as compared to using all shape or texture features. This implies that a combination of edge features will not provide a viable solution for feature classification. However there may be individual edge features that may help improve the classification when used within an ensemble of other shape and texture based features. This is the focus of investigations carried out in section 6.4.3.

Table 6-6 Summary statistics for different classifiers when using a combination of all edge features

Classifier	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
BayesNet	0.828	0.440	0.984	0.917	0.813	0.595	0.768
NaiveBayes	0.793	0.600	0.871	0.652	0.844	0.625	0.829
SVM	0.716	0	1	0	0.713	0	0.500
MLP	0.793	0.400	0.952	0.769	0.797	0.526	0.788
RBFNetwork	0.793	0.440	0.935	0.733	0.806	0.550	0.759
Bagging	0.816	0.400	0.984	0.909	0.803	0.556	0.814
AdaBoost	0.805	0.520	0.919	0.722	0.826	0.605	0.806
LogitBoost	0.770	0.480	0.887	0.632	0.809	0.545	0.785
Random Tree	0.713	0.640	0.742	0.500	0.836	0.561	0.691
Random Forest	0.713	0.480	0.806	0.500	0.794	0.490	0.732

6.4.3 Comparison of feature selection methods

To obtain more reliable results, a detailed experiment on the use of well-known feature selection algorithms (Best First, LFS, Genetic Search, and Greedy Stepwise) was conducted. Table 6-7 summarises the feature combinations selected when different feature selection algorithms implemented within WEKA were used.

Table 6-7(i) Feature sets selected by different feature selection algorithms implemented within WEKA.

Feature / Method	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
CfsSubsetEval		√		√					√		√	√							
ClassifierSubsetEval																			
BayesNet, Best First		√		√							√								
BayesNet, LFS		√		√							√								
BayesNet, Genetic Search				√	√										√				
BayesNet, Greedy Stepwise		√		√							√								
NaiveBayes, BestFirst	√	√						√				√					√		
NaiveBayes, LFS	√	√						√				√					√		
NaiveBayes, Genetic Search	√			√	√				√			√			√				
NaiveBayes,Greedy Stepwise	√	√						√											
SVM, Best First		√	√	√								√			√				
SVM, LFS		√	√	√								√			√				
SVM, Genetic Search			√	√	√				√			√	√		√		√		
SVM, Greedy Stepwise		√	√	√								√			√				
MLP, Best First	√	√	√	√															√
MLP, LFS		√	√	√					√						√				
MLP, Genetic Search	√			√			√			√	√	√			√	√	√	√	
MLP, Greedy Stepwise	√	√	√	√															√
RBFNet, Best First	√	√		√			√					√				√	√	√	
RBFNet, LFS	√	√		√				√				√				√	√	√	
RBFNet, Genetic Search	√			√		√			√			√	√			√	√	√	
RBFNet, Greedy Stepwise		√		√													√		

Table 6-7(ii) Continue from Table 6-7(i) Feature sets selected by different feature selection algorithms implemented within WEKA.

Feature / Method	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Bagging, Best First		√							√										√
Bagging, LFS		√							√										√
Bagging, Genetic Search	√		√	√								√			√	√	√		
Bagging, Greedy Stepwise		√							√										√
AdaBoost, Best First	√	√	√									√							√
AdaBoost, LFS		√					√		√		√				√		√		
AdaBoost, Genetic Search	√	√		√							√	√		√			√	√	
AdaBoost, Greedy Stepwise	√	√																	√
LogitBoost, Best First	√	√	√	√					√										√
LogitBoost, LFS	√	√	√	√					√										√
LogitBoost, Genetic Search				√	√				√		√	√		√				√	√
LogitBoost, Greedy Stepwise		√		√															
RandomTree, Best First	√																		
RandomTree, LFS												√							
RandomTree, Genetic Search					√			√											√
RandomTree, Greedy Stepwise	√																		
RandomForest, Best First				√															
RandomForest, LFS										√									
RandomForest, Genetic Search					√			√											√
RandomForest, Greedy Stepwise				√															

For example, from table 6-7(i) and table 6-7(ii), it is seen that the feature selection technique, CfsSubsetEval, picked up a feature subset that consists of features, 2 (solidity), 4(elongation), 9 (form factor), 11(mean) and 12(standard deviation) as the best fitting set of features for subsequent classification.

Table 6-8 A comparison of statistics when features 2,4,9,11,12 (CfsSubsetEval selection) are used by different classifiers.

Classifier	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
BayesNet	0.897	0.800	0.935	0.833	0.921	0.816	0.894
NaiveBayes	0.897	0.880	0.903	0.786	0.949	0.830	0.930
SVM	0.912	0.760	0.986	0.950	0.910	0.844	0.872
MLP	0.908	0.800	0.952	0.870	0.922	0.833	0.937
RBFNetwork	0.920	0.840	0.952	0.875	0.937	0.857	0.925
Bagging	0.862	0.720	0.919	0.783	0.891	0.750	0.911
AdaBoost	0.851	0.840	0.855	0.700	0.930	0.764	0.897
LogitBoost	0.885	0.880	0.887	0.759	0.948	0.815	0.940
Random Tree	0.782	0.720	0.806	0.600	0.877	0.655	0.763
Random Forest	0.874	0.800	0.903	0.769	0.918	0.784	0.941

Table 6-8 shows the comparison of statistics by using the subset chosen by CfsSubsetEval, which proves that feature selection has lead to an improvement in classification. RBFNetwork performed the best among the classifiers with a *success rate* of 0.920, *sensitivity* of 0.840, *specificity* of 0.952, and *F-measure* of 0.857. Figure 6-13 visually compares the performance of different classifiers when using the above feature set, by using ROC curves. It illustrates that SVM performs best when using the selected features.

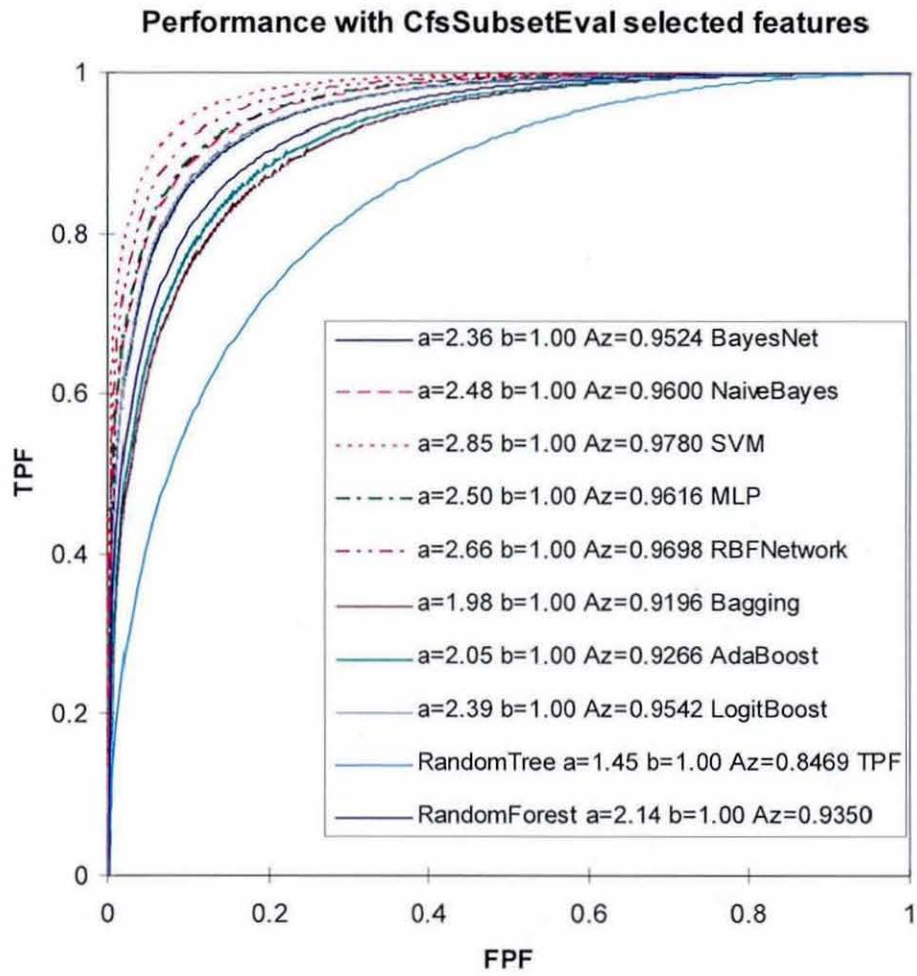


Figure 6-13 Fitted ROC curve illustrates the performance of different classifiers with features selected by CfsSubsetEval

Further performance evaluations under combinations of different features were conducted. It was observed that although the reduction of features resulted in a reduction of complexity, the feature reduction cannot be carried out without limits. In certain cases removal of certain features resulted in an improved performance of classification. However in general the purpose of reducing the features is to find a trade-off between complexity and the cost. In medical analysis, the *sensitivity* and *specificity* is far more important than complexity issue. Although the complexity of a machine is affordable, reduction of *sensitivity* and *specificity* is not acceptable due to the nature of application and impact on human life.

Table 6-9 Results for BayesNet with selected feature sets

Feature Selection Method	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
Best First LFS Greedy Stepwise (2,4,11)	0.885	0.760	0.935	0.826	0.906	0.792	0.886
Genetic Search (4,5,15)	0.874	0.800	0.903	0.769	0.918	0.784	0.898

Table 6-10 Results for NaiveBayes with selected feature sets

Feature Selection Method	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
Best First LFS (1,2,8,12,17)	0.874	0.800	0.903	0.769	0.918	0.784	0.926
Genetic Search (1,4,5,9,12,15)	0.874	0.88	0.871	0.733	.0947	0.800	0.934
Greedy Stepwise (1,2,8)	0.874	0.80	0.903	0.769	0.918	0.784	0.928

Table 6-11 Results for SVM with selected feature sets

Feature Selection Method	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
Best First LFS Greedy Stepwise (2,3,4,12,15)	0.908	0.800	0.952	0.870	0.922	0.833	0.876
Genetic Search (3,4,5,9,12,13, 15,17)	0.908	0.76	0.968	0.905	0.909	0.826	0.864

Table 6-12 Results for MLP with selected feature sets

Feature Selection Method	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
Best First Greedy Stepwise (1,2,3,4,19)	0.897	0.760	0.952	0.864	0.908	0.809	0.951
LFS (2,3,4,9,15)	0.920	0.760	0.984	0.950	0.910	0.844	0.931
Genetic Search (1,4,7,10,11,12 ,15,16,17,18)	0.885	0.800	0.919	0.800	0.919	0.800	0.915

Table 6-13 Results for RBFNetwork with selected feature sets

Feature Selection Method	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
Best First (1,2,4,7,12,16, 17,18)	0.943	0.880	0.968	0.917	0.952	0.898	0.939
LFS (1,2,4,8,12,16, 17,18)	0.943	0.880	0.968	0.917	0.952	0.898	0.948
Genetic Search (1,4,6,9,12,13, 16,17,18)	0.908	0.840	0.935	0.840	0.935	0.840	0.942
Greedy Stepwise (2,4,17)	0.885	0.800	0.919	0.800	0.919	0.800	0.947

Table 6-14 Results for AdaBoost with selected feature sets

Feature Selection Method	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
Best First (1,2,3,12,19)	0.897	0.720	0.968	0.900	0.896	0.800	0.884
LFS (2,7,9,11,15,17)	0.828	0.720	0.871	0.692	0.885	0.706	0.875
Genetic Search (1,2,4,11,12,14 ,17,18)	0.874	0.72	0.935	0.818	0.892	0.766	0.889
Greedy Stepwise (1,2,19)	0.828	0.640	0.903	0.727	0.862	0.681	0.873

Table 6-15 Results for Random Forest with selected feature sets

Feature Selection Method	Success rate	Sensitivity	Specificity	PPV	NPV	F-measure	A_z
Best First Greedy Stepwise (4)	0.701	0.640	0.726	0.485	0.833	0.552	0.770
LFS (10)	0.678	0.400	0.790	0.435	0.766	0.417	0.588
Genetic Search (5,8,19)	0.816	0.640	0.887	0.696	0.859	0.667	0.852

Tables 6-9 to 6-15 summarise the statistical results of classification when using features selected by CfsClassifierSubset in conjunction with each classifier. It is observed that the RBFNetwork with Best First feature selection approach or RBFNetwork with LFS feature selection approach (Table 6-13) resulted in an improved performance and produced reliable results with a *success rate* of 0.943, *sensitivity* of 0.88, *specificity* of 0.968, and *F-measure* of 0.898. By using 8 features (1,2,4,7,12,16,17,18) or (1,2,4,8,12,16,17,18), RBFNetwork performed best with reduced complexity. On the other hand, single feature selection by using Random Tree and Random Forest approaches are not reliable as indicated by the results tabulated in Table 6-15.

6.5 Summary & Conclusion

After obtaining the boundaries of the lesions (Chapter 5), it is important to categorise lesions into being benign or malignant. The focus of Chapter 6 was to identify the best approaches for the above classification task.

The use of appearance based approaches, PCA and LDA in lesion classification was investigated first, proving that they are not suitable for lesion type recognition. Secondly feature based approaches to classification were investigated. A thorough investigation into different types of features, divided into three groups, shape based, texture based and edge based, that can be used in classification was subsequently carried out. Experimental results conclude that in general, shape based features are the most reliable for classification of lesion type, while edge based features are the least reliable. Further investigations revealed that different feature combinations perform best when used in conjunction with different classifiers. Thus identifying the feature-classifier combination that performs optimally is a matter of significant importance. The high accuracy levels in recognition also reflect the accuracy of the lesion boundary detection algorithms proposed in Chapter 5.

It was shown that feature selection algorithms can help reduce the complexity of the classification process and achieve reasonable A_z values in classifiers' performance. In medical applications, a maximum sensitivity and specificity is crucial, as the matter concerns to human life and wellbeing. Hence, a high accuracy algorithms always receives priority in applications even if they are more complicated as compared to others.

In conclusion, the combination of features 1 (area to perimeter ratio), 2 (solidity), 4 (elongation), 8 (Roundness), 12 (standard deviation), 16(fourier1), 17 (fourier2), 18(fractal1) has provided the best A_z value (predicted value of 0.948). This feature combination worked best with RBFNetwork classifier when tested on the given dataset. Due to the difficulties in maintaining the consistency in obtaining ultrasound breast images, it is predicted that this choice of features may vary from database to database. However significant changes are not expected.

Chapter 7 investigates the relationship between human vision and computer vision by investigating the effects of computer processed images in human perception.

Chapter 7:Effects of Computer Processed Images on Human Performance

7.1 Introduction

In Chapter-1 it was discussed that the aim of CAD systems is to provide tools that will assist radiologists and physicians to make more informed and accurate decisions in diagnostics. In this process, CAD systems produce a multitude of processed digital images that will enhance the perceptual quality and/or interpretation of images that would otherwise be difficult to examine due to the presence of imaging artefacts such as speckle noise in US images. Therefore once a CAD system has been designed and implemented it is vital to determine whether the associated processed images have a positive impact on the decision making ability and performance accuracy of its users. Further the detailed analysis of the results of the subjective experiments may provide vital feedback that can be used to further optimize and develop the CAD system and to find a correspondence between the level of experience of an individual and need for particular functionality in the CAD systems, which is vital in developing intelligent tutoring systems for specific professional groups associated with cancer screening.

This chapter presents the findings of a subjective experiment carried out to determine the effects of computer processed images (i.e. using the novel CAD techniques presented in this thesis) on human performance. Finally a detailed analysis of the results is used to draw up conclusions.

7.2 Research Motivation

In the UK, currently there are some 109 cancer screening centres and a growing number of individuals (circa 650) nationally performing approximately 1.5 million examinations per year [3]. As a result there is increasing interest in Intelligent Tutoring Systems (ITS) in the area of medical training [122, 123]. In general the rationale of ITS is based on the assumption that the learner's cognitive processes can be modelled, traced, and corrected in the context of problem-solving [123]. To this effect, Crowley and Medvedeva [124] have worked on the adaptation of ITS and Knowledge-based systems (KBS) for creating intelligent educational systems. However, though ITS have provided a basic pedagogical approach with proven efficacy in domains outside the field of medicine, it was not designed to be used in large, frequently changing, or existing knowledge bases [124] within the area of medical diagnosis.

The aim of the research presented in this chapter is to investigate the effects of using computer processed images alongside non-processed images traditionally used in ITS, particularly within knowledge bases with wide variations. In particular the effects of computer processed images in the judgment of individuals with backgrounds in radiology, computer science, engineering, mathematics and arts are analysed. The results of this study provide valuable input and justifications to the use of CAD/ITS in medical diagnosis, in particular screening of breast cancer.

For Ultrasound images, based on the perspective of human visual system, the abnormalities are difficult to detect if they are above psycho-visual thresholds. If

below threshold, perceptual skills of pattern recognition or matching are needed to detect the abnormality. Once the suspicious region is detected, the cognitive skill of decision making (classification) is needed to group the feature to malignant or benign categories [125].

To become an expert general radiologist, one needs to equip with radiological expertise, thoroughly acquiring knowledge of relevant anatomy, physiology, pathology, physiopathology, projective geometry of radiography, and the essentials of medicine and surgery [123]. Providing additional processed images aims to further assist radiologists in detection and interpretation in ultrasound examinations. It is believed that the additional information will boost the confidence level of the human psycho-visual system in decision making. Within the context of the research presented in this chapter the effects of providing such additional information to non-expert radiologists and non-radiologists is investigated. In the following section, details of the subjective testing methodology adopted to prove the contribution of processed images in ultrasound medical training systems are provided.

7.3 Methodology

Two experiments were conducted to examine the effect of processed images towards two groups of subjects in:

- the detection (perceptual) tasks of non-radiologists and
- the detection and interpretation (perceptual and cognitive) tasks of radiologists.

7.3.1 Experimental design and rationale

The experiments were designed to subjectively measure the level of human perception, perceptual tasks and cognitive tasks, in ultrasound breast imaging. Initially, the experiments were performed on non-radiologists (perceptual task) and were subsequently extended to cover a larger sample of people, with diverse levels of exposure to ultrasound imaging, particularly to those groups who have had experience as radiographers and radiologists (perceptual and cognitive tasks). In particular the interest lies in further analysing the possibility of performance improvement of the latter groups of professionals by the effective use of computer processed images.

The survey involved the following steps:

- Randomly select 5 original ultrasound images. Then select a further five images with similar diagnostic difficulty and obtain their corresponding computer processed (using our work in Chapter-4 , summarised in section 7.3.3) images as sample test images.

- Design questionnaires to impartially investigate subjective judgment.

- Conduct the survey on two distinct subject groups: expert radiologists and non-radiologists.

- Analyse the results using statistical and ROC based methods.

- Draw suitable conclusions.

Each of the steps is described in detail in the following sub-sections.

7.3.2 Test image dataset

The test images used in our work are obtained from a professionally compiled Breast Ultrasound CD [22], in which each image is accompanied by an accurate explanation of its diagnosis from an expert radiologist. A total of 51 clinical ultrasound cases and more than 200 instructional videos are included in the CD ROM. It is particularly noted that the CD only contained images/videos of abnormal breast ultrasound images. In order not to overload the subject [126], the length of the questionnaires was limited to 10 questions. The subjects were required to study and comment on five randomly selected original ultrasound breast images of different diagnostic, and five further specially selected original ultrasound breast images and their corresponding processed images. All subjects were presented with the same set of randomly selected images. However to avoid subjectivity due to performance variations due to human fatigue, the five images were presented to each subject in a random order within the questionnaires.

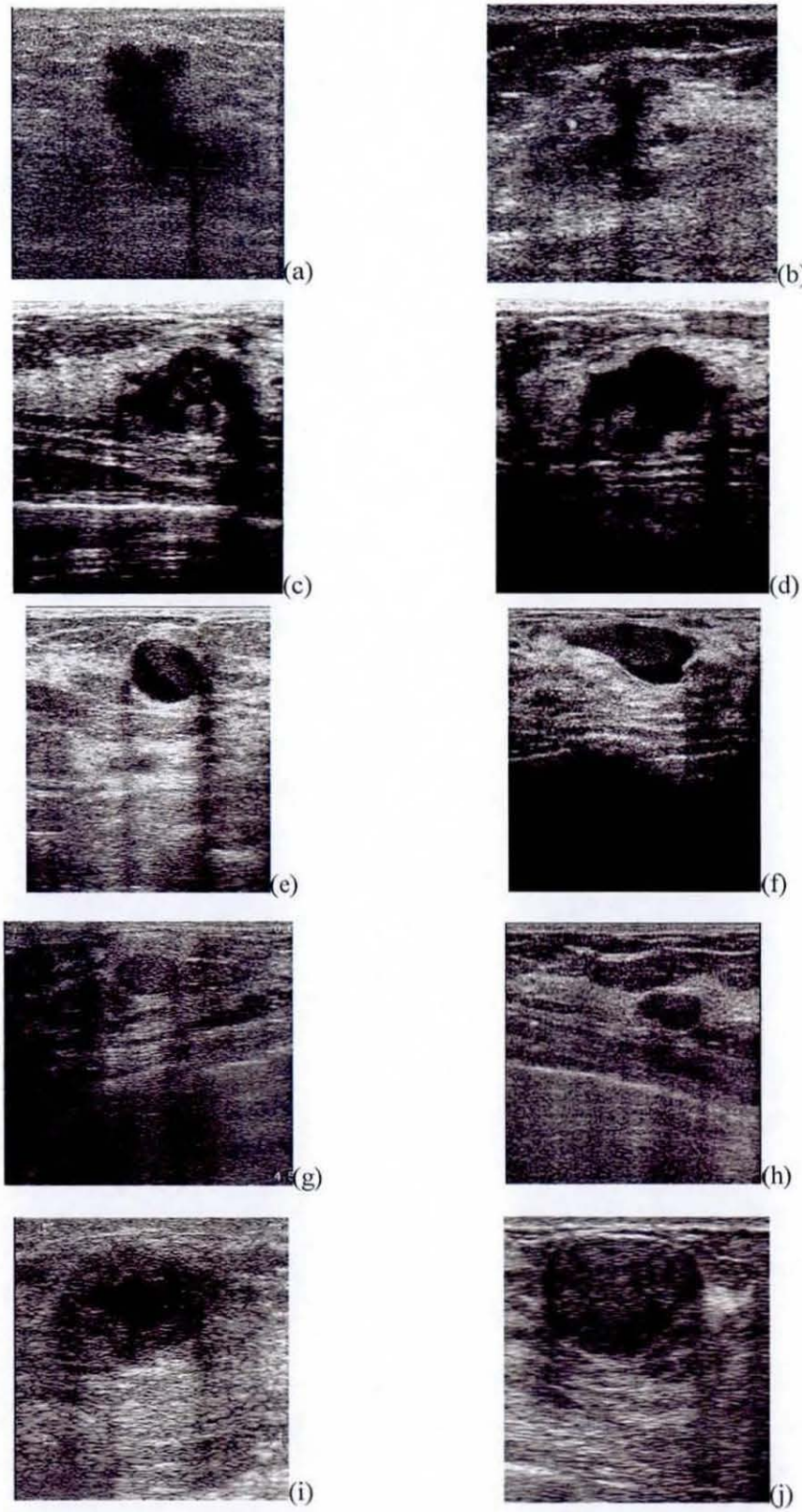


Figure 7-1. (a,b), (c,d), (e,f), (g,h), (i,j) are five pairs of original ultrasound images selected for the survey. (a, c, e, g, i) are selected randomly, while (b, d, f, h, j) are purposely selected to be of same level of diagnostic difficulty.

In order to maintain the consistency of the survey, original ultrasound breast image pairs with similar level of diagnostic difficulty were chosen from the test ultrasound image database. These images are illustrated in Figure 7-1.

7.3.3 Processed images

In medical image analysis, a number of algorithms are available to pre-process ultrasound images to improve chances of subsequent correct diagnosis either by manual, semi-automatic or automatic means. These algorithms use different filtering methods to produce an enhanced image for object segmentation. The degree of enhancement (or the level of visual aid) offered by alternative image processing algorithms may vary. However, human vision has a lot of constraints. Hence, not all the processing/filtering algorithms are suitable for human vision.

Amongst the state-of-the-art pre-processing algorithms, isogauss processing approach has been proven to perform the best in the visibility to human perception (see figure 7-2 for a comparison of performance). Essentially, isogauss processing involves the multiplication of the inverse of the image by an Isotropic Gaussian function (with a fixed variance as the constraint function), centred at a seed point location. Readers interested in more technical details are referred to Chapter-5, section 5.2.1.

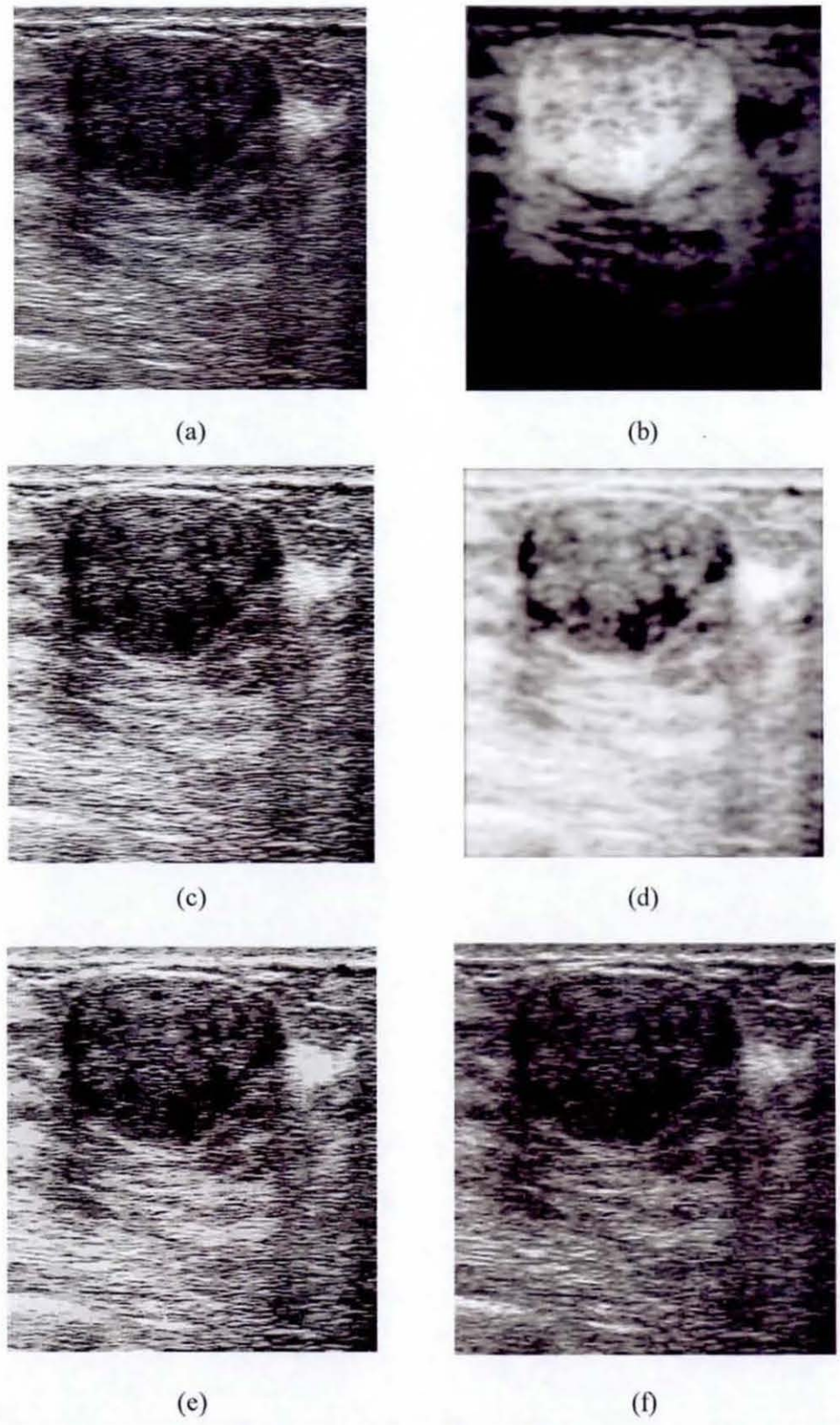


Figure 7-2 (a) Original image (b), (c), (d), (e), (f) Processed images by using different algorithms: (b) isogauss processing, (c) histogram equalization (d) Multifractal filtering (e) Non-linear filtering, (f) median filtering.

7.3.4 Subjects

The Experiments were conducted on two groups of subjects, non-radiologists and radiologists. The non-radiologist group was represented by 40 subjects from different academic backgrounds, i.e. in computer science, engineering, mathematics and arts. One major difficulty in setting up the experiments was the search for participants to represent the group of expert radiologists. For this reason only ten expert radiologists participated in the survey. The convincing, conclusive evidence gathered from the subjective tests confirm this number of subjects as sufficient.

7.3.5 Design of the questionnaire

Two types of questions were designed, one for analysing and commenting on a random set of original images and the second for similar on the set of original-processed image pairs. Due to the differences between the two sets of subject groups (e.g. familiarity with medical terms, abilities etc.), the questionnaires had to be appropriately re-worded for each group. Readers interested in the details of the questionnaires are referred to Appendix A. In the above process, careful consideration has been given to the re-wording to maintain impartiality of responses from the group of subjects. The popular Absolute Category Rating (ACR) [127] method was used to measure the level of perception of subjects. ACR involves a five scale evaluation, in which the subjects are asked to observe the images and determine whether an identifiable object is present in terms of, certainly not exist, probably not exist, not sure, probably exist and certainly exist.

In the case of the volunteer radiologists, the questions were directed in medical terms. First they were asked to decide the suspected region. If either probably exist or certainly exist was selected, the subjects were asked to draw the boundary of the suspected region on the images, and provide the diagnosis of the suspected region with five scale evaluation: benign, probably benign, unknown, probably malignant, malignant. In the case of non-radiologist, some appropriately-guided questions, without medical terms, were presented in the questionnaire. First they were asked to decide the existence of objects in the images. If an object was detected the subjects were further asked to provide the location/size of the suspected region in terms of (x, y) coordinate values and to determine how the width compares with the height. Subjects were further asked to elaborate on the shape of the object by classifying it to be of regular, smooth or unknown shape. The detection was considered to be successful if the subjects managed to accurately detect the presence of an object, while the interpretation was considered to be successful if the subjects managed to provide accurate width, height, and boundary information.

Further subjective experimentation considerations such as, quality of the images presented, their ordering, human fatigue, lighting conditions, time-of-day and age group etc., were also carefully considered.

7.3.6 Receiver Operating Characteristics (ROC) analysis

As a means of illustrating the experimental results, two ROC curves are plotted for each group of subjects with the help of the software package ROCKIT [111]. ROC analysis is a popular tool used to assess, define and compare classification results in medical diagnostics. In medical research, decision threshold represents how often one is right or wrong in a diagnostic test. ROC analysis shows the trade off between the sensitivity and specificity as a decision threshold is varied and uses maximum likelihood estimation to fit a binomial ROC curve [128] to the test data.

The sensitivity is represented by the True Positive Fraction (TPF) and is defined as TPF = (TP / (TP + FN)) x 100, where TP and FN are respectively the number of true positive results and false negative results. The specificity is represented by the False Positive Fraction (FPF) and is defined as FPF = (TN / (TN + FP)) x 100, where TN and FP are respectively the number of true negatives and false positives. In other words the sensitivity is the fraction of cancerous (i.e. malignant) images that were correctly recognized from the total number of cancerous images, whereas specificity is the fraction of images that are recognized as non-cancerous (i.e. benign) from the total number of non-cancerous images. As the non-radiologists group will not be sufficiently experienced to decide whether an abnormality represents a malignant or non-malignant lesion, this decision is deduced by the authors using the answers given by the group members to the additional questions asked in the relevant questionnaire (see section 7.4).

Section 7.4 presents the experimental results and a detailed analysis, leading to conclusions that are presented in section 7.5.

7.4 Results and Analysis

Table 7-1 illustrates the overall results of detection accuracies obtained in the survey, i.e. the perceptual task by non-radiologists. The experimental results are classified according to the subjects' professional background. The computed results are presented in a scale of 0 to 1, where 1 refers to 100%.

Table 7-1 Detection accuracies.

Groups	Images presented to the subject		
	Original images	Processed images	Both images
Computer Science	0.64	0.66	0.76
Mathematics	0.5556	0.4222	0.8667
Engineering	0.56	0.52	0.82
Arts	0.44	0.54	0.76

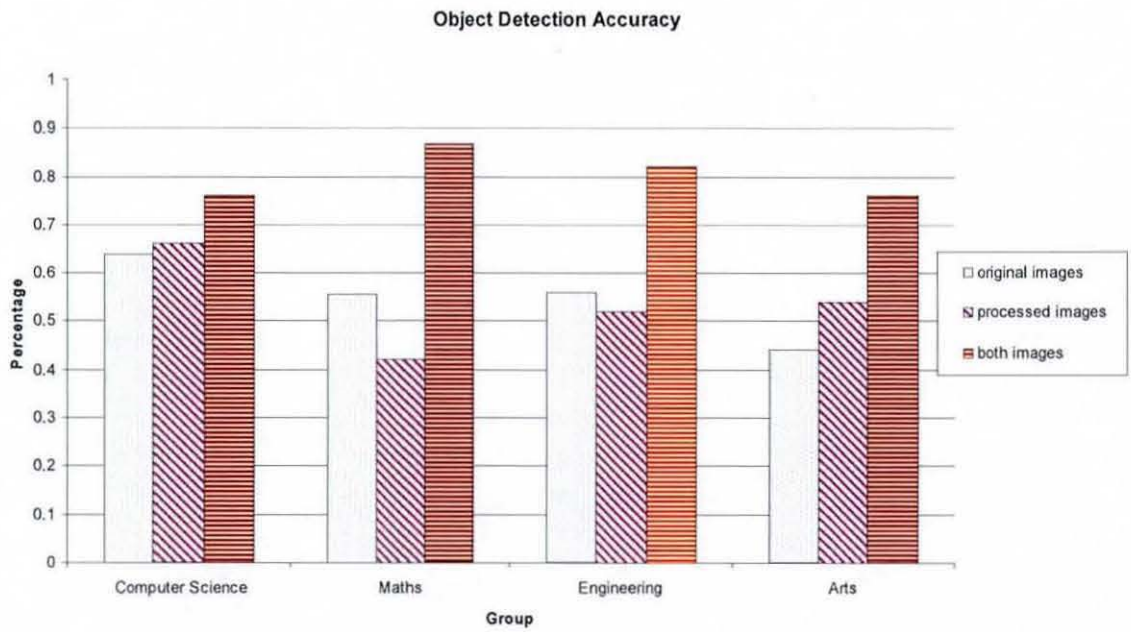


Figure 7-3. Comparison of the detection accuracy between subjects from different professional backgrounds.

Figure 7-3 is provides a graphical presentation of the accuracy results obtained for the different sub-groups of subjects (of the group of non-radiologists). It is noted that subjects having a computer science background, most of whom having expertise in computer vision and image analysis, managed to obtain the highest accuracy rates when individual images, i.e. the original and processed images, were separately shown. When only the original images were shown, subjects with an arts background performed worse and when only the processes images were shown subjects with a mathematical background performed worse. When further queried it was revealed that these results were mostly due to the amount of previous exposure each group has had with ultrasound images and any sort of computer processed images. Therefore it can be concluded that the experience one has had with images play an important role in their ability to use them in perceptual tasks.

The most important observation from figure 7-3 is the ability of all groups of subjects to improve their detection accuracy (perceptual task) significantly when both images are provided simultaneously, as compared to showing them separately. For example the overall average detection accuracy when using original images is 54.89%, which increases to 80.17% when both images were provided simultaneously for inspection. Interestingly the best percentage improvement has also been illustrated by the group of people, mathematicians, who had the least performance accuracy when the images were shown separately.

Figure 7-4 illustrates the results obtained from the ROC tool for the group of non-radiologists. It shows that providing the original and processed images side-by-side allows the non-radiologists to perform substantially better, compared to providing them with only the original images. A lack of experience in identifying abnormalities in ultrasound images results in either: (i) wrong judgements or (ii) more uncertainties in the decision making process, when only original images are examined by this group of people. The processed images provide a clearer segmentation of abnormal regions which increases the accuracy of judgement substantially. Summarising the above observations, it can be concluded that computer processed images, when provided as an aid, alongside original images, increases the perceptual capability of most non-radiologists.

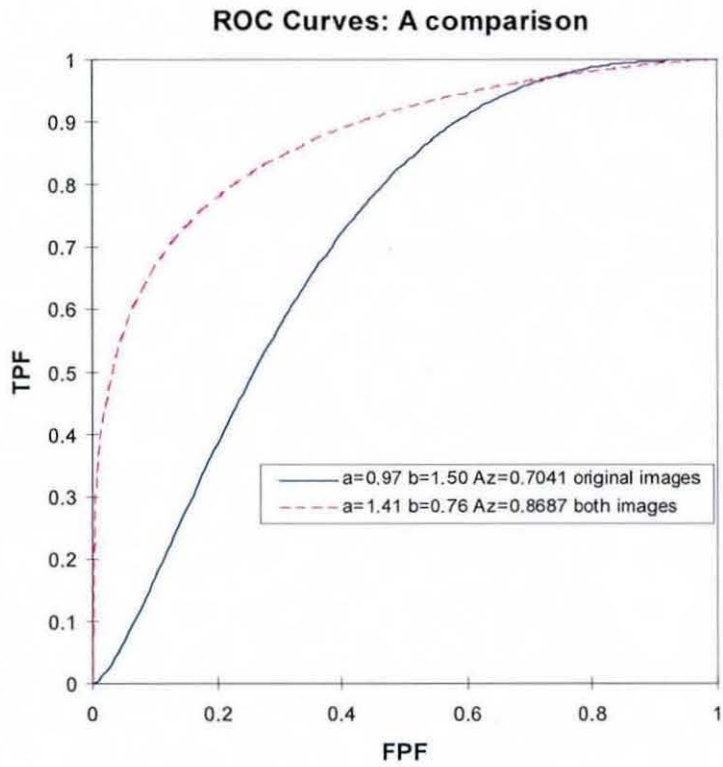


Figure 7-4. Comparison of ROC curves (non-Radiologists)

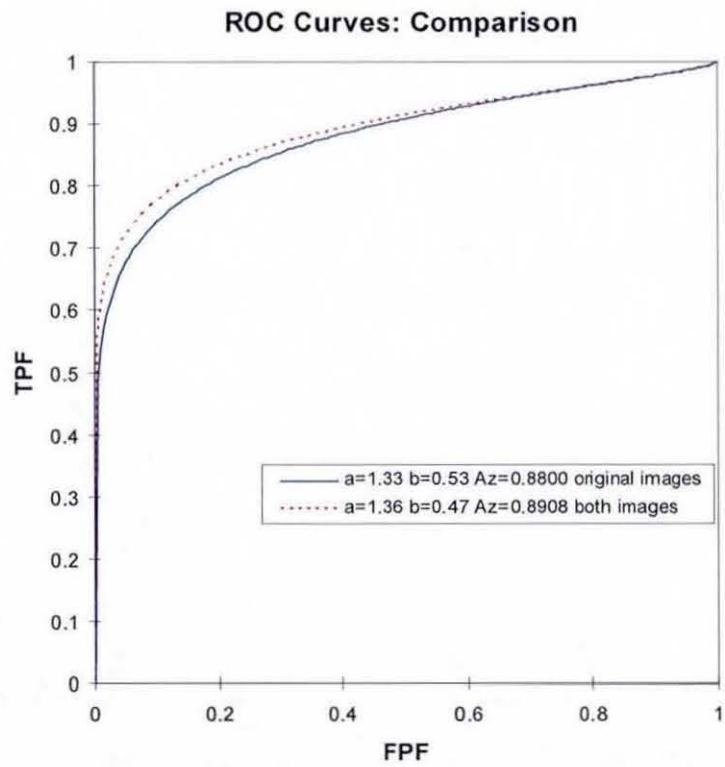


Figure 7-5. Comparison of ROC Curves (Radiologists)

Figure 7-5, illustrates that providing the original and processed images, side-by-side, allows the group of radiologists to perform marginally better. It is noted that subjects in this group, who have expertise in accurately reading ultrasound images in their original format, managed to obtain the high sensitivity and specificity when only the original images were presented.

Given the above observations, it can be concluded that radiological experience with ultrasound images plays an important role in accurate diagnostics, which involved the cognitive process. It can be concluded that the provision of processed images significantly aids the perceptual performance of non-radiologists and marginally improves the performance of experienced radiologists. Given the fact that a trainee radiologists at the start of training, may not be significantly different experience wise as compared to a subject from the group of non-radiologists, it can be concluded that computer processed images may be a useful way to boost their confidence in diagnostics and improve speed of training. Further, despite the marginal improvement in performance accuracy shown by the expert radiologists group, computer processed images can help in decision making, particularly when diagnosing images with noise and artefacts.

7.5 Summary & Conclusion

In this chapter subjective experiments have been carried out to analyse the improvement of human performance, in perceptual and cognitive aspects, that can be obtained by providing computer processed images alongside original images when using ultrasound images in breast cancer screening/detection. It has been shown that the sensitivity and specificity are improved when both images are made available for judgment. It can be concluded that computer processed images, when provided as an aid, alongside original images, increase the perceptual task capability of most individuals, particularly those who are inexperienced radiologists or trainees. These observations clearly prove the effectiveness of using imaging technologies/algorithms in CAD systems. Therefore in an era where there is a shortage of experienced radiologists, CAD systems that provide computer processed images as an aid for diagnostics are highly beneficial. We suggested that processed images could be adapted in the radiologist's training system.

Chapter 8: Summary and Future Work

8.1 Summary

This chapter summarizes the original contributions made by the thesis to the research area of breast cancer detection and recognition in ultrasound images. It further summarises the conclusions made in previous chapters, based on the analysis of the experimental results. It also highlights possible future improvements and directions of research related to the work presented.

This thesis has presented enhanced algorithms in ultrasound breast cancer detection and an initial investigation in the effects of processed images on human performance. It has shown that the proposed algorithms enhanced existing state-of-art computer vision algorithms and the intermediate images from CAD supported human observers in making more accurate decisions in breast cancer detection. The novel ideas in this thesis contribute to the US CAD system and improvement of existing training systems. The contributions relate to the design and development of novel algorithms for initial lesion detection (Chapter 4), segmentation (Chapter 5), classification (Chapter 6), and a study on the effects of processed images on human performance (Chapter 7). The detailed summary of contributions are:

Chapter-4 presents the first application of hybrid filtering and multifractal analysis in ultrasound breast images. Hybrid filtering was implemented at the pre-processed stage while multifractal analysis was useful in separating the segment of interest from the normal lesions. Subsequently, two novel algorithms in initial lesions detection were

proposed with the aim of achieving a fully automated initial lesions detection process. Method-2 had shown promise by increasing the accuracy relative to a benchmark. Overall comparison showed that Method-2 out-performed the other algorithms in initial lesion detection. However, the algorithm has a number of limitations: failure in detecting hyper-echoic lesions, i.e. lesions in cases where suspect regions are brighter than normal tissues; and the algorithm is constrained to a single region of interest in each image.

Chapter-5 has demonstrated an improvement in boundary detection. Two novel algorithms for segmentation were discussed and proposed. It was shown that with multifractal processing, segmentation, i.e. refinement of the boundary of lesions was improved. Further it has shown that the use of an Isotropic Gaussian function with GVF can enhance the results. The output of the segmentation provides the input to the classification algorithms, where the credibility measure was done. From the experimentation in chapter 6, the accuracy of the segmentation algorithm has been deduced. However, they had limitations; they failed to segment some fibroadenoma in cases where the fibroadenoma lesions appeared to be very similar to the appearance in the normal cases.

Chapter-6 featured experiments involving classification algorithms. It was shown that appearance-based classifiers failed to separate benign from malignant tumours, whilst feature-based classifiers were successful in classification of tumours. It was shown that feature selection algorithms can help to reduce the complexity of the classification and achieve reasonable performance. The focus of previous research works were concentrated in finding the optimum feature set that can be used by a

selected classifier. The new breakthrough in this chapter was the attempt in evaluating the use of ten different feature-based classifiers and experimentally determining the optimum feature set for each classifier. From the overall performance, RBFNetwork with feature set {1 (area to perimeter ratio), 2 (solidity), 4 (elongation), 8 (roundness), 12 (standard deviation), 16 (fourier1), 17 (fourier2), 18(fractal1)} out-performed other combinations of feature-classifier, with an achievement of predicated A_z value of 0.948. Within a limited dataset, the feature based classifiers performed reasonably well, but there is still scope for improvement – as implied by data mining principle, "the larger the training set, the better the classifier model".

In a conventional US CAD system, the input to the system is a manually cropped rectangular region of interest (ROI) by a radiologist. As a result, the statistical analysis of the ROI is presented to the radiologists. Computer vision researches have developed various pre-processing algorithms to improve the performance of computer vision. Thus, investigations into the intermediate images in the CAD were undertaken in Chapter-7. Experiments on lesion detection and recognition performance amongst non-radiologists and radiologists were conducted. The results of these experiments show that the algorithms support an improvement in perceptual task, but no significant difference in the cognitive task. The experiment opened a door for researchers to further investigate the relationship between computer vision and human vision. The research in this area could be extended to wider range of subjects and broader examination by applying different algorithms to the pre-processed images. It may be of interest to study which algorithm works the best for human vision and to compare its relative efficiency in computer vision.

8.2 Future Work

Ultrasound breast cancer detection is a complicated and involved multidisciplinary field. It is clear that not all aspects of this task could be encompassed in this research. Some findings of the thesis suggest new areas of investigation. The research presented has resulted in a number of original contributions in the area of ultrasound breast cancer detection algorithms. Arising from these contributions, a number of possibilities exist for the future extension and enhancement of the proposed ideas:

- Detection of hyper-echoic lesions

  The current algorithms performed well in detecting hypo-echoic lesions, this is due to the assumption that the abnormalities would appear of lower intensity when compared to the normal lesions. Rarely, the lesions appeared to be brighter than the normal lesions. Future work involves the adjustment of parameters and alterations to the selection rules to overcome this problem.

- Larger database, to include more normal images

  To further improve the algorithms, a larger database is needed. Normal images are not captured or stored by the sonographer during the diagnostic process, as they are only interested in capturing the abnormal lesions. Consequently, it is common to find paucity of normal images in the database. To overcome this problem, close collaborative work with hospitals and clinics is needed. With a larger training set, a better classification model can be achieved. Ultimately, a better tolerance and consistency in classification might be realised.

- Processed images adaptation in a training tool

  In late 2007 it was announced that breast cancer screening was to be extended again to women aged 47 to 73 years by 2012. Thus, there will be a high demands on radiologists in 2012. Consequently, more training tools are needed to provide options for radiologists. As an important adjunct to mammography, the demand for sonographers is expected to increase in relative proportion. Hence, developing a reliable training tool for US breast cancer detection is important. The potential to adapt processed images in a training tool can only be assessed by empirical experiments. A potential downside to the algorithm is that the reliability of processed images in training might create some decision making dependency. These are research problems that have yet to be addressed.

- Effects of computer vision on human performance

  Chapter-7 demonstrates the study of the effect of computer processed images on human performance; it might motivate researchers in both computer vision and human perception to work together. Further investigation involves examining the relationship between level of training, training methods, age and background of trainees, and the improvement of diagnostic accuracy that can be provided by CAD systems. Additionally, it is worth investigating the effects of different state-of-the art segmentation algorithms on human performance.

- Real time and 3D breast ultrasound

  The initial lesion detection algorithm works on 2D images. In real time scanning, there are series of slices. It is useful to adapt the initial lesions detection algorithm to process the slices simultaneously and provide feedback if there are suspected

lesions in a given slice. Such an application might then raise a real time alert to the radiologist to suggest further inspection of the slice. Currently, research is moving from 2D images to 3D images [129, 130, 131]. Thus, it is imperative that 2D algorithms be adapted or replaced in order that they facilitate lesion detection and classification in 3D breast ultrasound images.

- Multi-modalities

  The field of breast cancer diagnosis involves multiple diagnostic techniques. It is crucial to investigate how US is used alongside other modalities, especially in mammography and Magnetic Resonance Imaging (MRI).

- Contributions to ontological descriptions of ultrasound images

  It has been shown that shape descriptors, texture descriptors and edge descriptors could discern most benign from malignant tumours. These features might be useful in ontological descriptions for data mining purposes.

8.3 Final Conclusions

The enhancement of algorithms for lesion detection and recognition in US breast images remains an active area of research exploration of computer vision algorithms, pattern recognition, data mining, perceptual and cognitive supports for identifying suspect regions in US breast images. However, these computer vision related factors might be considered alongside other refinements; e.g. larger image archives, updated images, frequent perceptual exposure and performance feedback. Additionally, the use

of processed US images might provide valuable insights into the perceptual and cognitive processes underlying the complex search and recognition procedure.

This thesis has provided four significant contributions to US breast imaging: (i) initial lesion detection, (ii) segmentation, (iii) classification algorithms and (iv) assessments of human performance in computer processed images. Experimental results and detailed analyses have been provided to support the novel ideas presented here. These original contributions have been published as a number of journal and conference papers (see Appendix B). An additional research article on the classification has been submitted for publication.

References:

[1] Imaginis Corporation, "Breast Cancer Glossary of Medical Terms," http://www.imaginis.com/glossary/ (accessed on 10/11/2008).

[2] University of Newcastle upon Tyne, Dept. of Medical Oncology, "Online Medical Dictionary," http://cancerweb.ncl.ac.uk/omd/ (accessed on 10/11/2008).

[3] U. K. National Health Service, "National Health Service Breast Screening Programme," http://www.cancerscreening.nhs.uk/breastscreen/index.html, (accessed on 15/01/2008).

[4] R. A. Smith, D. Saslow, K. A. Sawyer, W. Burke, M. E. Costanza, W. P. Evans, R. S. Foster, E. Hendrick, H. J. Eyre, and S. Sener, "American Cancer Society guidelines for breast cancer screening: update 2003," *CA Cancer J Clin,* vol. 53, pp. 141-169, 2003.

[5] Breastcancer.org, 2008, "Mammograms," http://www.breastcancer.org/symptoms/testing/mammograms/, (accessed on 30/10/2008).

[6] S. C. Bushong, *Diagnostic Ultrasound.* United States: McGraw-Hill, 1999.

[7] A. T. Stavros, *Breast Ultrasound.* USA: Lippincott Williams & Wilkins, 2004.

[8] E. A. Ashton and K. J. Parker, "Multiple resolution Bayesian segmentation of ultrasound images," *Ultrasonic Imaging,* vol. 17, pp. 291-304, 1995.

[9] D. I. Hughes and F. A. Duck, "Automatic attenuation compensation for ultrasonic imaging," *Ultrasound Med. Biol.,* vol. 23, pp. 651-664, 1997.

[10] D. Boukerroui, A. Baskurt, N. J. Alison, and O. Basset, "Segmentation of ultrasound images - multiresolution 2D and 3D algorithm based on global and local statistics," *Pattern Recognition Letters,* vol. 24, pp. 779-790, 2003.

[11] D. R. Chen, R. F. Chang, and Y. L. Huang, "Breast Cancer Diagnosis using Self-Organizing Map for Sonography," *Ultrasound in Med. & Biol.,* vol. 26, pp. 405-411, 2000.

[12] R. F. Chang, W. J. Wu, W. K. Moon, and D. R. Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors," *Breast cancer research and treatment,* vol. 89, pp. 179-185, Jan 2005.

[13] Y. L. Huang, D. R. Chen, Y. R. Jiang, S. J. Kuo, H. K. Wu, and W. K. Moon, "Computer-aided Diagnosis Using Morphological Features for Classifying Breast Lesions on Ultrasound," *Ultrasound Obstet.Gynecol.,* 2008.

[14] K. Drukker, M. L. Giger, K. Horsh, M. A. Kupinski, and C. J. Vyborny, "Computerized Lesion Detection on Breast Ultrasound," *Med. Physics,* vol. 29, pp. 1438-1446, 2002.

[15] M. Kupinski, M. L. Giger, P. Lu, and Z. Huo, "Computerized detection of mammographic lesions: Performance of artificial neural network with enhanced feature extraction," *SPIE,* vol. 2434, pp. 598-605, 1995.

[16] S. Joo, Y. S. Yang, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features," *IEEE Transactions on Medical Imaging,* vol. 23, pp. 1292-1300, Oct 2004.

[17] NCI/NIH, 2008, "National Cancer Institute," http://www.cancer.gov/, (accessed on 08/06/2008).

[18] Invision, Radiology Imaging Associates, 2008,"Glossary of Medical Terms," http://www.riainvision.com/invision/patientinfo/resources/patinfo_res_gloss.asp, (accessed on 31/08/2008).

[19] P. Suetens, *Fundamental of Medical Imaging*: Cambridge University Press, 2002.

[10] L. A. Venta, C. M. Dudiak, C. G. Salomon, and M. E. Flisak, "Sonographic Evaluation of the Breast," *RadioGraphics,* vol. 14, pp. 29-50, 1994.

[21] J. M. Blackledge, *Digital Image Processing*. Chichester, West Sussex: Horwood Publishing, 2005.

[22] S. T. Prapavesis, B. D. Fornage, C. F. Weismann, A. Palko, P. Zoumpoulis, and et al., *Breast Ultrasound and US-Guided Interventional Techniques*. Thessaloniki Greece, 2001.

[23] Siemen, "Siemen Medical," http://www.SiemensMedical.com, (accessed on 10/10/2006).

[24] X. Hao, S. Gao, and X. Gao, "A Novel Multiscale Nonlinear Thresholding Method for Ultrasonic Speckle Suppressing," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 787-794, 1999.

[25] N. Friedland and D. Adam, "Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated annealing," *IEEE Trans on Med Imaging*, vol. 8, pp. 244-253, 1989.

[26] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian Multiscale Method for Speckle Removal in Medical Ultrasound Images," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 772-783, 2001.

[27] K. Z. Abd-Elmoniem, A. M. Youssef, and Y. M. Kadah, "Real-Time Speckle Reduction and Coherence Enhancement in Ultrasound Imaging via Nonlinear Anisotropic Diffusion," *IEEE Transactions on Biomedical Engineering*, vol. 49, pp. 997-1014, 2002.

[28] Imaginis Corporation, "Breast Health (Non-cancerous Breast Issues)," http://www.imaginis.com/breasthealth/benign.asp (accessed on 31/08/2008).

[29] S. C. Rankin, "MRI of the breast," *British Journal Radiology*, vol. 73, pp. 806-818, 2000.

[30] Breastcancer.org, "Breast Self-exam (BSE)," http://www.breastcancer.org/symptoms/testing/self_exam/ (accessed on 30/10/2008).

[31] UcsanDiego, "Physical Sciences," http://physicalsciences.ucsd.edu/news_articles/ (accessed on 02/08/2008).

[32] J. J. James, "The current status of digital mammography," *Clin Radiol,* vol. 59, pp. 1-10, 2004.

[33] College of Radiology Malaysia homepage, "Radiology Malaysia," http://www.radiologymalaysia.org (accessed on 08/02/2006).

[34] E. D. Pisano, C. Kuzmiak, and C. Koomen, "Perspective on digital mammography," *Semin Roentgenol,* vol. 36, pp. 195-200, 2001.

[35] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, "Solid Breast Nodules: Use of Sonography to Distinguish between Benign and Malignant Lesions," *Radiology,* vol. 196, pp. 123-134, 1995.

[36] Breastcancer.org, "Magnetic Resonance Imaging (MRI)," http://www.breastcancer.org/symptoms/testing/mri.jsp (accessed on 30/10/2008).

[37] E. Bombardieri and F. Crippa, "PET Imaging in Breast Cancer," *Nucl Med.,* vol. 45, pp. 245-256, 2001.

[38] Review Journal, "Review Journal," http://reviewjournal.com (accessed on 08/07/2006).

[39] Wikipedia, MediaWiki, "Terahertz," http://en.wikipedia.org/wiki/Terahertz (accesed on 08/11/2008).

[40] Bio-medicine, "Terahertz Imaging Promises To Revolutionize Breast Cancer Treatments," http://www.bio-medicine-news/Terahertz-Imaging-Promises-To-Revolutionize-Brast-Cancer-Treatment-9715-1/, (accessed on 10/10/2006).

[41] Y. Yue, M. M. Croitoru, J. B. Zwischenberger, and J. W. Clark, "Nonlinear Multiscale Wavelet Diffusion for Speckle Suppression and Edge Enhancement in Ultrasound Images," *IEEE Trans on Med Imaging,* vol. 25, pp. 297-311, 2006.

[42] D. R. Chen, R. F. Chang, W. J. Wu, W. K. Moon, and W. L. Wu, "3-D Breast Ultrasound Segmentation Usiing Active Contour Model," *Ultrasound in Med. & Biol.*, vol. 29, pp. 1017-1026, 2003.

[43] P. Mrazek, "Nonlinear Diffusion for Image Filtering and Monotonicity Enhancement." vol. PhD Prague: Czech Technical University, 2001.

[44] J. Weickert, "Nonlinear Diffusion Filtering," *Signal Processing and Pattern Recognition,* vol. 2, pp. 424-446, 1990.

[45] D. R. Chen, R. F. Chang, C. J. Chen, M. F. Ho, S. J. Kuo, S. T. Chen, S. J. Hung, and W. K. Moon, "Classification of breast ultrasound images using fractal feature," *Clinical imaging,* vol. 29, pp. 235-245, Jul-Aug 2005.

[46] A. B. Arehart, L. Vincent, and B. B. Kimia, "Mathematical morphology: The Hamilton-Jacobi connection," in *Proc. Fourth Int. Conf. on Computer Vision (ICCV '93)*, Los Alamitos, 1993, pp. 215-219.

[47] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 629-639, 1990.

[48] Y. J. Yu and S. T. Acton, "Speckle Reducing Anisotropic Diffusion," *IEEE Trans. on Image Processing,* vol. 11, 2002.

[49] J. W. Gurney, "Neural Networks at the Crossroads: Caution Ahead," *Radiology,* vol. 193, pp. 27-28, 1994.

[50] J. M. Boone, "Neural Networks at the Crossroads," *Radiology,* vol. 189, pp. 357-359, 1993.

[51] D. R. Chen, R. F. Chang, W. M. Chen, and W. K. Moon, "Computer-aided diagnosis for 3-dimensional breast ultrasonography," *Archives of Surgery (Chicago, Ill. : 1960),* vol. 138, pp. 296-302, Mar 2003.

[52] C. M. Sehgal, T. W. Cary, S. A. Kangas, S. P. Weinstein, S. M. Schultz, P. H. Arger, and E. F. Conant, "Computer-Based Margin Analysis of Breast Sonography for Differentiating Malignant and Benign Masses," *J Ultrasound Med,* vol. 23, pp. 1201-1209, 2004.

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

[54] D. Boukerroui, O. Basset, N. Guerin, and A. Baskurt, "Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation," *European journal of ultrasound : official journal of the European Federation of Societies for Ultrasound in Medicine and Biology,* vol. 8, pp. 135-144, Nov 1998.

[55] R. M. Haralick and L. G. Shapiro, "Image Segmentation Techniques," 1985, pp. 100-132.

[56] M. A. Kupinski and M. L. Giger, "Automated Seeded Lesion Segmentation on Digital Mammograms," *IEEE Trans on Medical Imaging,* vol. 17, 1998.

[57] T. N. Pappas, "An Adaptive Clustering algorithm for image segmentation," *IEEE Trans. on Signal Processing,* vol. 40, pp. 901-914, 1992.

[58] R. Muzzolini, Y. H. Yang, and R. Pierson, "Multiresolution texture segmentation with application to diagnostic ultrasound images," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 12, pp. 108-123, 1993.

[59] V. Grau, U. J. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield, "Improved Watershed Transform for Medical Image Segmentation Using Prior Information," *IEEE Trans on Medical Imaging,* vol. 23, pp. 447-458, 2004.

[60] Y. L. Huang and D. R. Chen, "Watershed segmentation for breast tumor in 2-D sonography," *Ultrasound in medicine & biology,* vol. 30, pp. 625-632, May 2004.

[61] C. Xu and J. L. Prince, "Snakes, Shapes, and Gradient Vector Flow," *IEEE Transactions on Image Processing,* vol. 7, pp. 359-369, 1998.

[62] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Trans on Med Imaging,* vol. 16, pp. 199-209, 1997.

[63] J. M. B. Dias and J. M. N. Leitao, "Wall position and thickness estimation from sequences of echocardiographic images," *IEEE Trans on Med Imaging,* vol. 15, pp. 25-38, 1996.

[64] I. Mikic, S. Krucinski, and J. D. Thomas, "Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow extimates," *IEEE Trans on Med Imaging,* vol. 17, pp. 274-284, 1998.

[65] J. Heinig, R. Witteler, R. Schmitz, L. Kiesel, and J. Steinhard, "Accuracy of Classification of Breast Ultrasound Findings based on Criteria used for BI-RADS," *Ultrasound Obstet.Gynecol.,* 2008.

[66] D. R. Chen, R. F. Chang, and Y. L. Huang, "Computer-aided Diagnosis Applied to US of Solid Breast Nodules by Using Neural Networks," *Radiology,* vol. 213, pp. 407-412, 1999.

[67] D. R. Chen, R. F. Chang, Y. L. Huang, Y. H. Chou, C. M. Tiu, and P. P. Tsai, "Texture Analysis of Breast Tumors on Sonograms," *Seminars in ultrasound, CT, and MR,* vol. 21, pp. 308-316, 2000.

[68] D. R. Chen, R. F. Chang, W. J. Kuo, M. C. Chen, and Y. L. Huang, "Diagnosis of Breast Tumors with Sonographic Texture Analysis Using Wavelet Transform and Neural Networks," *Ultrasound in Med. & Biol.,* vol. 28, pp. 1301-1310, 2002.

[69] B. S. Garra, B. H. Krasner, S. C. Horii, S. Ascher, S. K. Mun, and R. K. Zeman, "Improving the distinction between benign and malignant breast lesions: the value of sonography texture analysis," *Ultrasonic Imaging,* vol. 15, pp. 267-285, 1993.

[70] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Computerized diagnosis of breast lesions on ultrasound," *Medical physics,* vol. 29, pp. 157-164, Feb 2002.

[71] L. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, and M. Helvie, "Classification of Malignant and Benign Masses Based on Hybrid ART2LDA Approach," *IEEE Trans on Medical Imaging,* vol. 18, pp. 1178-1187, 1999.

[72] Y. L. Huang and D. R. Chen, "Support vector machines in sonography application to decision making in the diagnosis of breast cancer," *Journal of Clinical Imaging,* vol. 29, pp. 179-184, 2005.

[73] Wikipedia, MediaWiki, "Support Vector Machine," http://en.wikipedia.org/wiki/Support_vector_machine, (accessed on 10/11/2008).

[74] R. Hopley and J. V. Schalkwyk, "The Magnificent ROC," http://www.anaesthetist.com/mnm/stats/roc/Findex.htm, (accessed on 03/10/2008).

[75] RSNA, "Mammography," http://www.radiologyinfo.org, (accessed on 29/08/2006).

[76] K. C. Young and D. Kitou, "Review of Literature on Digital Mammography in Screening," NHS Cancer Screening Programmes, Sheffield, UK, Report 0502, 2005.

[77] E. D. Pisano, E. B. Cole, and a. et, "Radiologists' preferences for digital mammography display. The International Digital Mammography Development Group," *Radiology,* vol. 216, pp. 820-830, 2000.

[78] M. Kallergi, "Computer-aided Diagnosis of mammographic microcalcification clusters," *Med. Phys,* vol. 31, pp. 314-326, 2004.

[79] W. K. Lim and M. J. Er, "Classification of mammographic masses using generalized dynamic fuzzy neural networks," *Ultrasound in Med. & Biol.,* vol. 31, pp. 1288-1295, 2000.

[80] Y. H. Chang, L. A. Hardesty, and et al., "Knowledge-based computer-aided detection of masses on digitized mammograms: a preliminary assessment," *Med. Physics,* vol. 28, pp. 455-461, 2001.

[81] H. J. Scott and A. G. Gale, "Breast Screening: PERFORMS Identifies Key Mammographic Training Needs," *British Journal of Radiology,* vol. 79, pp. S127-S133, 2006.

[82] M. H. Yap and A. G. Gale, "Grid-enabled Mammographic Auditing and Training System," *Proc. SPIE*, vol. 6919, 69190A, March 2008.

[83] I. Foster, C. Kesselman, J. Nick, and S. Tueke, "The physiology of the Grid - An Open Services Grid Architecture for Distributed Systems Integration," 2002.

[84] D. Rogulin, F. Estrella, T. Hauer, R. McClatchey, S. R. Amendolia, and A. Solomonides, "A Grid Information Infrastructure for Medical Image Analysis," in *DiDaMIC 2004 workshop*, 2004.

[85] A. Solomonides, M. McClatchey, M. Odeh, M. Brady, M. Mulet-Parada, D. Schottlander, and S. R. Amendolia, "MammoGrid and eDiamond: Grids Applications in Mammogram Analysis," in *Proceedings of the IADIS International Conference: e-Society 2003*, pp. 1032-1033, 2003.

[86] P. Wintz and R. Gonzalez, "Digital Image Processing." vol. 2nd, Anonymous, Ed., pp. 146-152, 1987.

[87] P. R. Hill, C. N. Canagarajah, and D. R. Bull, "Image Segmentation using a Texture Gradient Based Watershed Transform," *IEEE Transactions on Image Processing,* vol. 12, pp. 1618-1633, 2003.

[88] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing using Matlab*: Pearson Prentice Hall, 2004.

[89] B. M. T. Romeny, *Front-End Vision and Multiscale Image Analysis*: Kluwer Academic Publishers, 2002.

[90] M. F. Barnsley, R. L. Devaney, B. B. Mandlebrot, H. Peitgen, D. Saupe, and R. F. Voss, *The Science of Fractal Images*: Springer-Verlag, 1988.

[91] M. R. Schroeder, *Fractals, Chaos, Power Laws: Minutes From an Infinite Paradise*: W H Freeman & Co (Sd), 1991.

[92] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*: Englewood Cliffs: Prentice Hall, 1988.

[93] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An Efficient *k*-means clustering Algorithm: Analysis and Implementation," *IEEE Transactions on PAMI,* vol. 24, pp. 881-892, 2002.

[94] Wikipedia, "*k*-means clustering tutorial," http://en.wikipedia.org/wiki/K-means_algorithm, (accessed on 10/11/2008).

[95] LTI-LIB, "Computer Vision Library," http://ltilib.sourceforge.net/doc/homepage/index.shtml (accessed on 08/10/2005).

[96] G. Du and T. S. Yeo, "A Multifractal Approach for Auto-segmentation of SAR Images," 2001.

[97] B. B. Mandlebrot, *The Fractal Geometry of Nature*: W.H. Freeman and Company, 1982.

[98] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, "A Survey of Thresholding Techniques," *Computer Vision, Graphics and Image Processing,* vol. 41, pp. 233-260, 1988.

[99] H. David, *Multifractals - Theory and Application*: Chapman & Hall/CRC, 2001.

[100] B. B. Mandlebrot and N. Sarkar, "Texture segmentation using fractal dimension," *IEEE Trans on Pattern Analysis and Machine Intell.,* vol. 17, pp. 72-76, 1995.

[101] C. Xu and J. L. Prince, "Generalized gradient vector flow external forces for active contours," *Signal Processing,* vol. 71, pp. 131-139, 1998.

[102] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience,* vol. 3, pp. 71-86, 1991.

[103] M. Li and B. Yuan, "2D-LDA: A Statistical Linear Discriminant Analysis for Image Matrix," *Pattern Recognition Letters,* vol. 26, pp. 527-532, 2005.

[104] I. H. Witten and E. Frank, *Data Mining - Practical Machine Learning Tools and Techniques*. United States of America: Morgan Kaufmann, Elsevier, 2005.

[105] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo, pp. 338-345, 1995.

[106] V. Vapnik, *The Nature of Statistical Learning Theory*: Springer-Verlag, 1995.

[107] L. Breiman, "Bagging Predictors," *Machine Learning,* vol. 24, pp. 123-140, 1996.

[108] Y. S. Freund, R.E., "A Short Introduction to Boosting," *Journal of Japanese Society for Artificial Intelligence,* vol. 14, pp. 771-780, September, 1999.

[109] J. Friedman, T. Hastle, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics,* vol. 28, pp. 337-407, 2000.

[110] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, pp. 5-32, 2001.

[111] C. D. Metz, B. A. Herman, and C. A. Roe, "Statistical Comparison of Two ROC Curve Estimates Obtained from Partially-paired Datasets," *Med. Decis. Making,* vol. 18, pp. 110-121, 1998.

[112] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face Recognition Using LDA-Based Algorithms," *IEEE Trans. on Neural Networks,* vol. 14, pp. 195-200, 2003.

[113] J. C. Russ, *The image processing handbook, 5th Edition.* Boca Raton, Fla: CRC Press, 2007.

[114] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 971-987, 2002.

[115] L. D. F. D. Costa and R. M. Cesar, *Shape Analysis and Classification: Theory and Practise, 1st Edition*. Boca Raton, FL, USA: CRC Press Inc, 2000.

[116] J. Platt, B. Scholkopf, C. Burges, and A. Smola, "Fast Training of Support Vector Machines using Sequential Minimal Optimization," *Advances in Kernel Methods - Support Vector Learning*, Anonymous, Ed.: MIT Press, 2006.

[117] T. Mu, A. K. Nandi, and R. M. Rangayyan, "Classification of Breast Masses Using Selected Shape, Edge-sharpness, and Texture Features with Linear and Kernel-based Classifiers," *Journal of Digital Imaging*, vol. 21, pp. 153-169, 2008.

[118] M. Guetlein, "Large Scale Attribute Selection Using Wrappers," 2006.

[119] D. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning*. Reading, MA: Addison Wesley, 1989.

[120] M. A. Hall, "Correlation-based Feature Subset Selection for Machine Learning." vol. PhD Hamilton, New Zealand: University of Waikato, 1998.

[121] R. F. Chang, W. J. Wu, W. K. Moon, and D. R. Chen, "Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis," *Ultrasound in medicine & biology*, vol. 29, pp. 679-686, May 2003.

[122] R. S. Crowley and O. Medvedeva, "A General Architecture for Intelligent Tutoring for Diagnostic Classification Problem Solving," *AMIA. Annu. Symp. Proc.*, pp. 185-189, 2003.

[123] R. Azevedo and S. P. Lajoie, "The Cognitive Basis for the Design of a Mammography Interpretation Tutor," *International Journal of Artificial Intelligence in Education*, vol. 9, pp. 32-44, 1998.

[124] R. S. Crowley and O. Medvedeva, "An Intelligent Tutoring System for Visual Classification Problem Solving," *Artif.Intell.Med.*, vol. 36, pp. 85-117, 2006.

[125] A. G. Gale, "PERFORMS - a self assessment scheme for radiologists in breast screening," *Seminar in Breast Disease: Improving and Monitoring Mammographic Interpretation Skills,* vol. 6, pp. 148-152, 2003.

[126] Market Research Society (MRS), "Questionnaire Design Guidelines," Report May 2006 , 2006.

[127] S. Moller, "Chapter 3," in *Assessment and Prediction of Speech Quality in Telecommunication*, Anonymous, Ed.: Springer, 2000.

[128] S. Carsten, W. Sebastian, S. Tania, and J. Klaus, "Comparison of Eight Computer Programs for Receiver-Operating Characteristc Analysis," *Clinical Chemistry,* vol. 49, pp. 433-439, 2003.

[129] Y. H. Chou, C.M. Tiu, J. Chen, and R.F. Chang, "Automated Full-field Breast Ultrasonography: The Past and The Present," *J Med Ultrasound,* vol. 15, pp. 31-44, 2007.

[130] Breast Cancer Source, "3D Ultrasound distinguishes malignant from benign breast masses," AstraZeneca, 2008, http://www.breastcancersource.com/78454?itemId=3959767, (accessed 20/11/2008).

[131] B. Sahiner, H.P. Chan, M.A. Roubidoux, et al. "Malignant and Benign Breast Masses on 3D US Volumetric Images: Effect of Computer-aided Diagnosis on Radiologist Accuracy," *Radiology,* vol.242, Number 3, pp. 716-724, March 2007.

Appendix A: Sample Questionnaires

Questionnaire (Radiologist) *reference:*

Do you have any knowledge in ultrasound images?

Not at all ☐ beginner ☐ intermediate ☐ good ☐ expert ☐

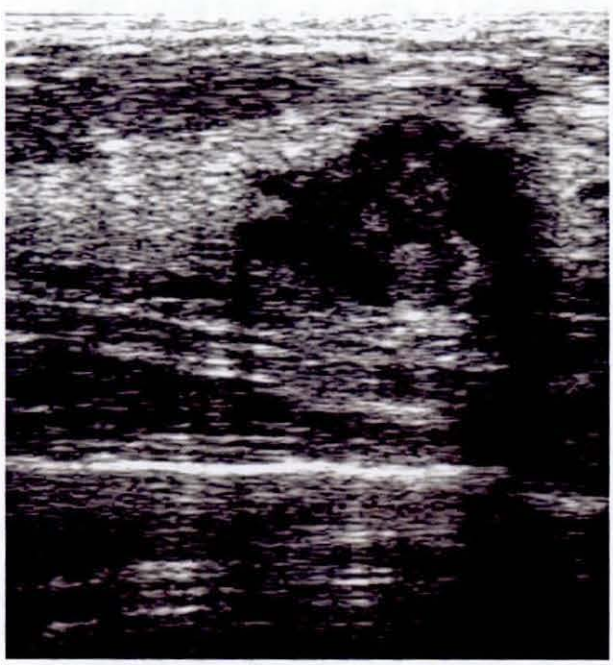
What is your background?

Physician ☐ Resident ☐ Technologist ☐ other (please state): _____

Your contact (optional):_____

The Objective of this research is to develop an efficient training system in breast cancer detection. Thanks for your kind contribution (to save life)!

Look at the images below and select the answer that best fit.

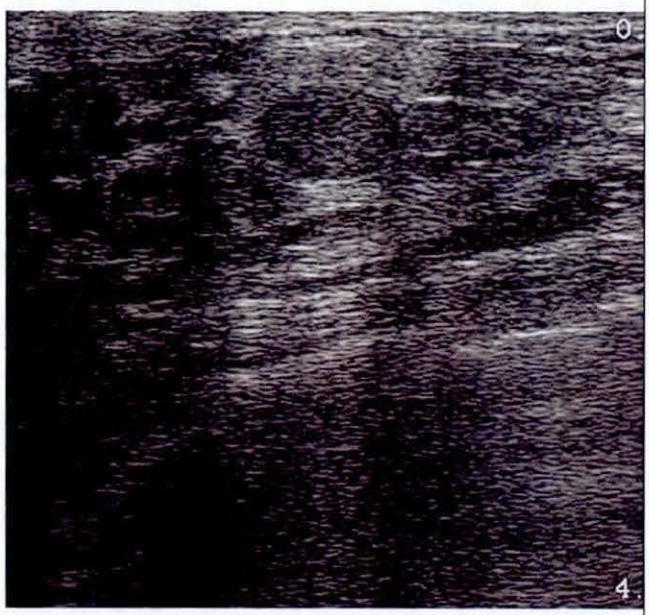


Any suspected abnormal region?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5,</u> please <u>draw the boundary</u> of the abnormal lesion, then give the diagnosis:

1 – benign ☐
2 – probably benign ☐
3 – unknown ☐
4 – probably malignant ☐
5 – malignant ☐



Any suspected abnormal region?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5,</u> please <u>draw the boundary</u> of the abnormal lesion, then give the diagnosis:

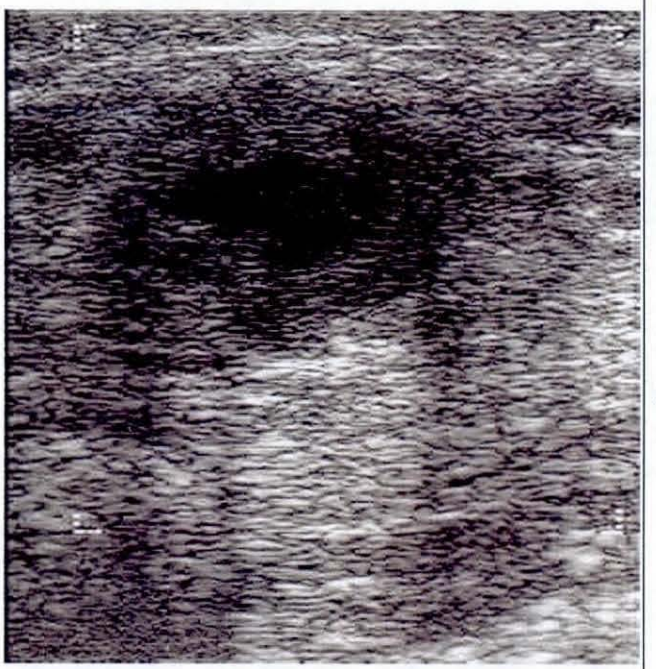
1 – benign ☐
2 – probably benign ☐
3 – unknown ☐
4 – probably malignant ☐
5 – malignant ☐

Any suspected abnormal region?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5,</u> please <u>draw the boundary</u> of the abnormal lesion, then give the diagnosis:

1 – benign ☐
2 – probably benign ☐
3 – unknown ☐
4 – probably malignant ☐
5 – malignant ☐

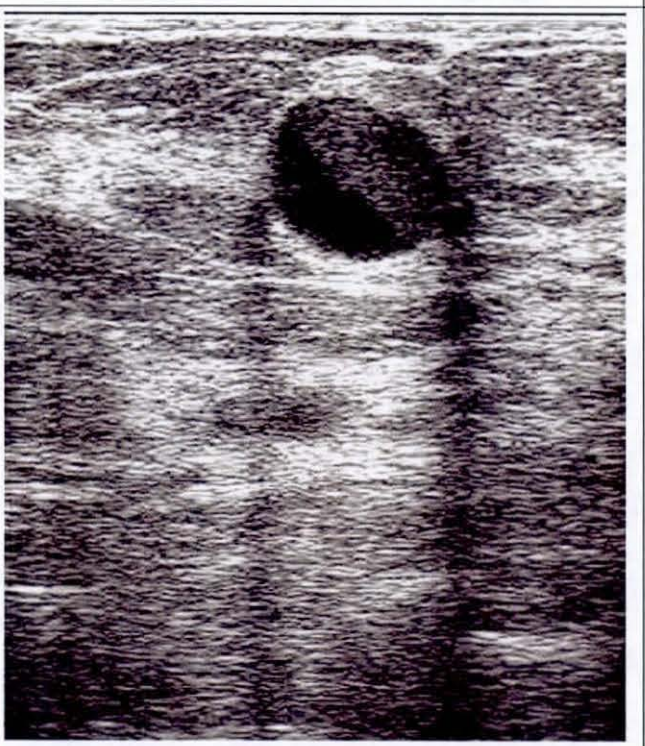


Any suspected abnormal region?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5,</u> please <u>draw the boundary</u> of the abnormal lesion, then give the diagnosis:

1 – benign ☐
2 – probably benign ☐
3 – unknown ☐
4 – probably malignant ☐
5 – malignant ☐

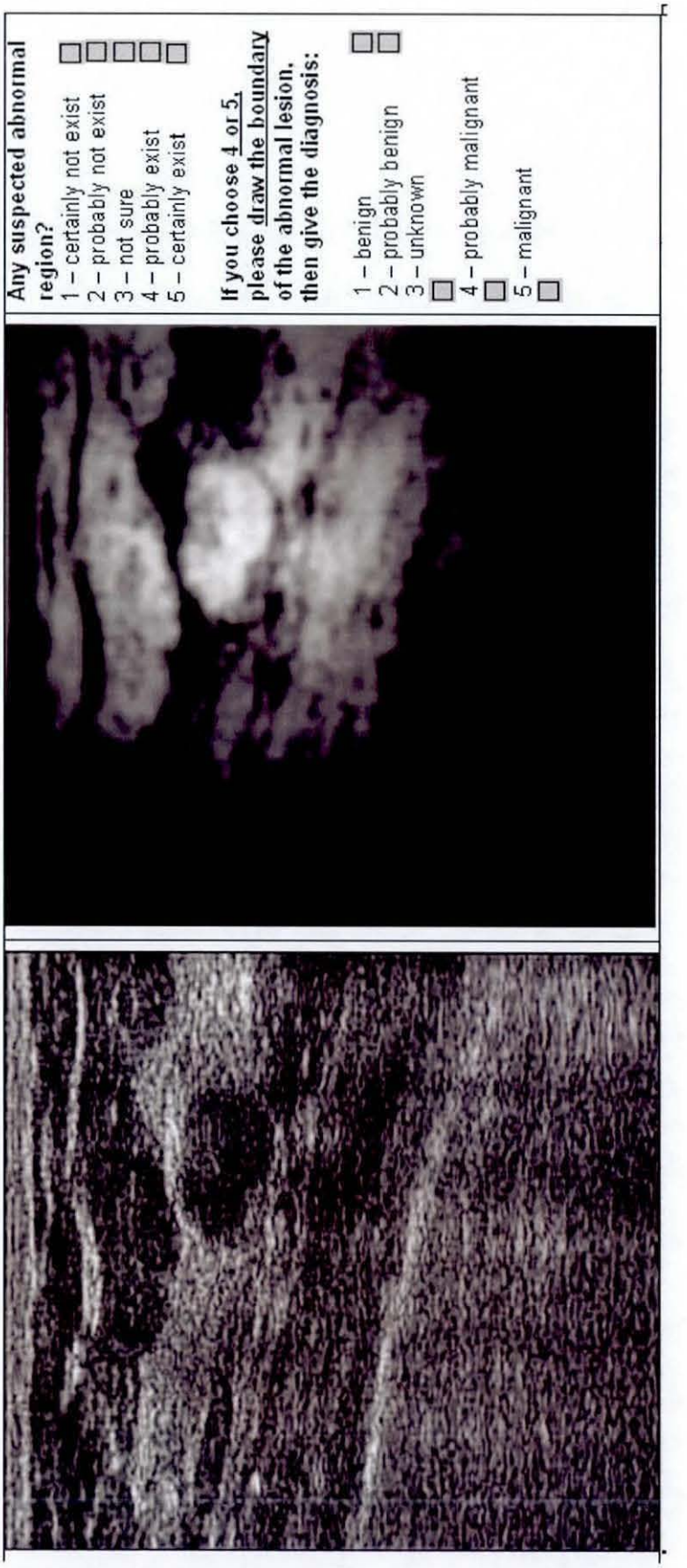


Any suspected abnormal region?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5,</u> please <u>draw the boundary</u> of the abnormal lesion, then give the diagnosis:

1 – benign ☐
2 – probably benign ☐
3 – unknown ☐
4 – probably malignant ☐
5 – malignant ☐


Any suspected abnormal region?
1 – certainly not exist
2 – probably not exist
3 – not sure
4 – probably exist
5 – certainly exist
If you choose 4 or 5, please draw the boundary of the abnormal lesion, then give the diagnosis:
1 – benign
2 – probably benign
3 – unknown
4 – probably malignant
5 – malignant

Any suspected abnormal region?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5</u>, please <u>draw the boundary</u> of the abnormal lesion, then give the diagnosis:

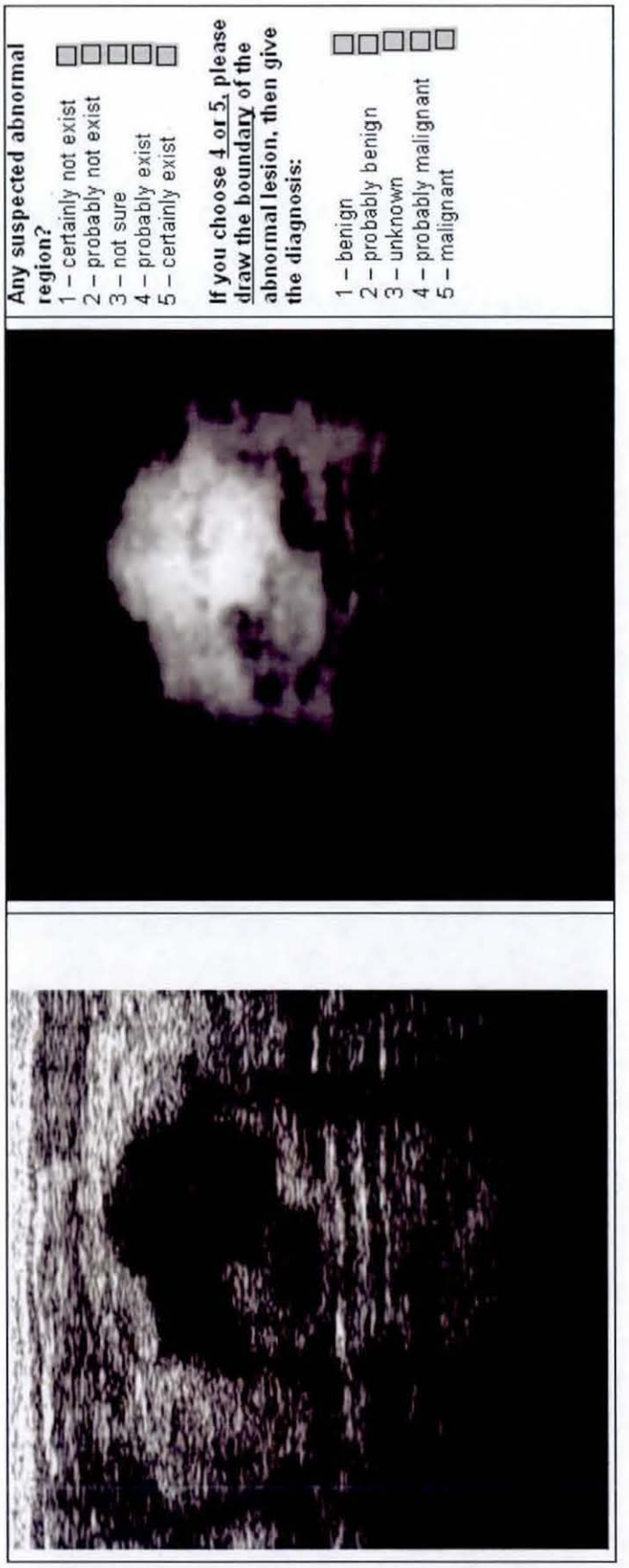
1 – benign ☐
2 – probably benign ☐
3 – unknown ☐
4 – probably malignant ☐
5 – malignant ☐

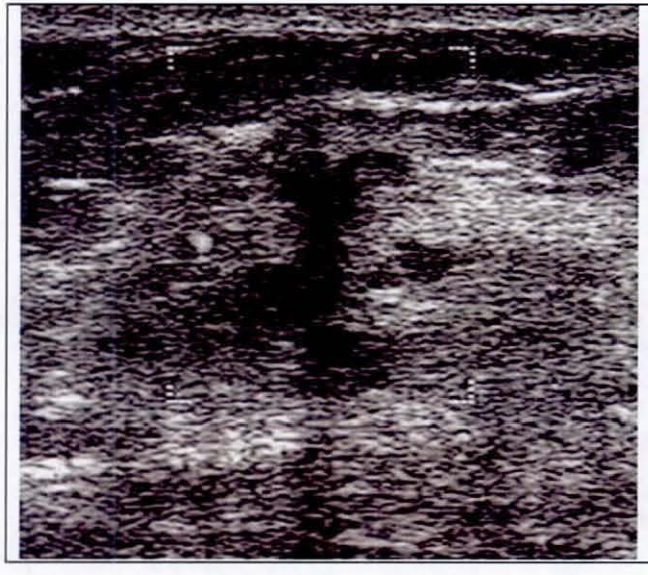
Any suspected abnormal region?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5,</u> please <u>draw the boundary</u> of the abnormal lesion, then give the diagnosis:

1 – benign ☐
2 – probably benign ☐
3 – unknown ☐
4 – probably malignant ☐
5 – malignant ☐


Any suspected abnormal region?
1 – certainly not exist
2 – probably not exist
3 – not sure
4 – probably exist
5 – certainly exist
If you choose 4 or 5, please draw the boundary of the abnormal lesion, then give the diagnosis:
1 – benign
2 – probably benign
3 – unknown
4 – probably malignant
5 – malignant



Any suspected abnormal region?

1 – certainly not exist ☐
2 – probably not exist ☐☐
3 – not sure ☐☐
4 – probably exist ☐☐
5 – certainly exist ☐

If you choose 4 or 5, please draw the boundary of the abnormal lesion, then give the diagnosis:

1 – benign ☐
2 – probably benign ☐☐
3 – unknown ☐☐
4 – probably malignant ☐☐
5 – malignant ☐

Do you think the processed image assist you in detecting the ROI? Yes/No

Do you think the processed image assist you in classification? Yes/No

Additional Comments: _____

_____

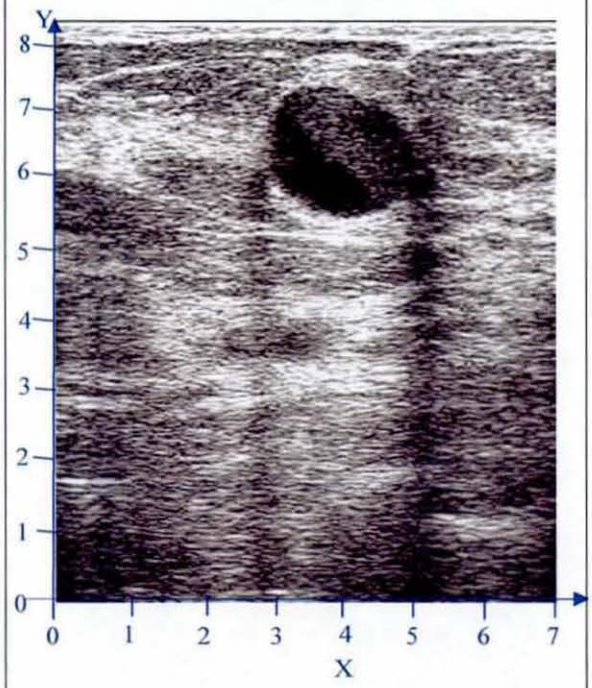
_____

Questionnaire (Non-radiologists)

Do you have any knowledge in ultrasound images?
Not at all ☐ beginner ☐ intermediate ☐ good ☐ expert ☐

What is your background?
Computer science ☐ mathematics ☐ Ergonomics ☐ psychology ☐
management ☐ medical ☐ other (please state): _____

Look at the images below and select the answer that best fit.



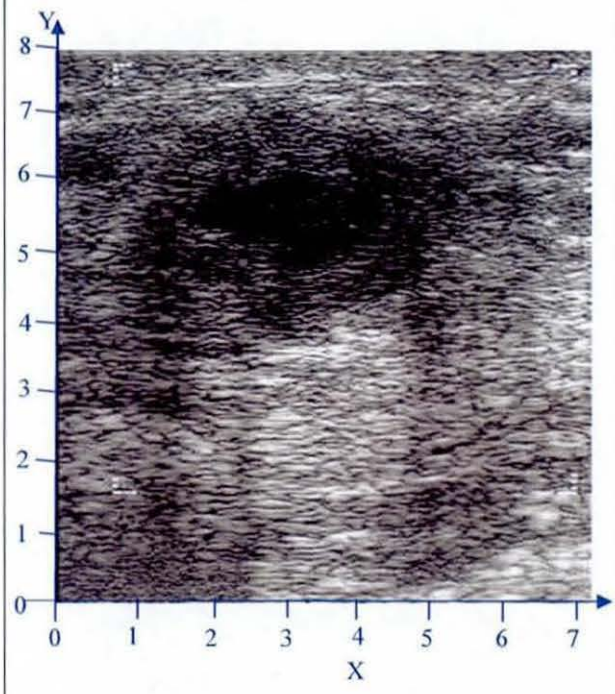
Any object in the image?
1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5,</u> please describe the features:

Object location(approximation):
X: to
Y: to

Object size:
Height > width ☐
Width > height ☐
Width = height ☐
Unknown ☐

Object Boundary:
Smooth ☐
Irregular ☐
Unknown ☐



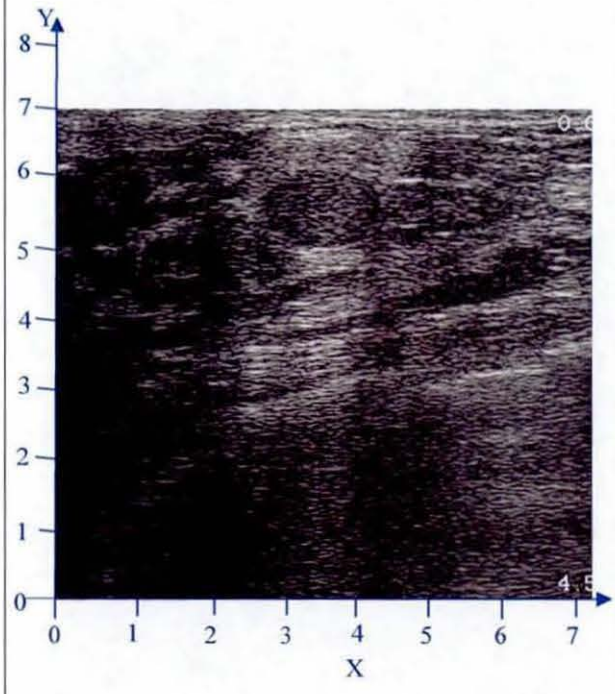
Any object in the image?
1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5</u>, please describe the features:

Object location(approximation):
X: to
Y: to

Object size:
Height > width ☐
Width > height ☐
Width = height ☐
Unknown ☐

Object Boundary:
Smooth ☐
Irregular ☐
Unknown ☐



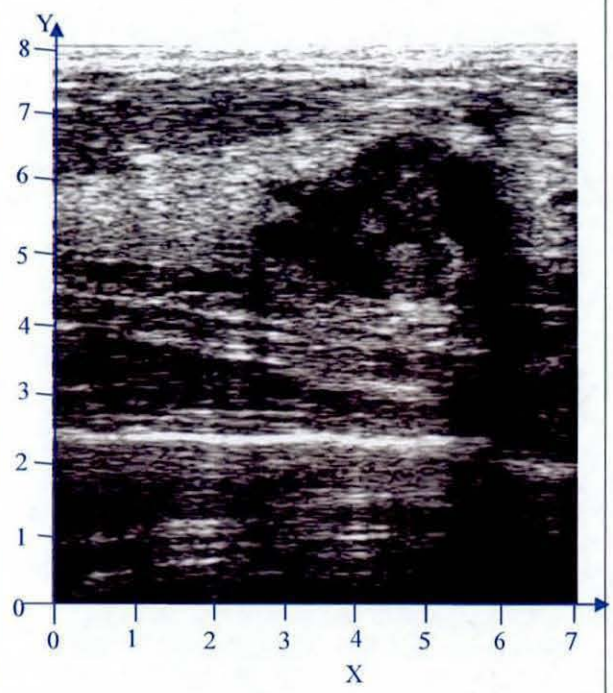
Any object in the image?
1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5</u>, please describe the features:

Object location(approximation):
X: to
Y: to

Object size:
Height > width ☐
Width > height ☐
Width = height ☐
Unknown ☐

Object Boundary:
Smooth ☐
Irregular ☐
Unknown ☐



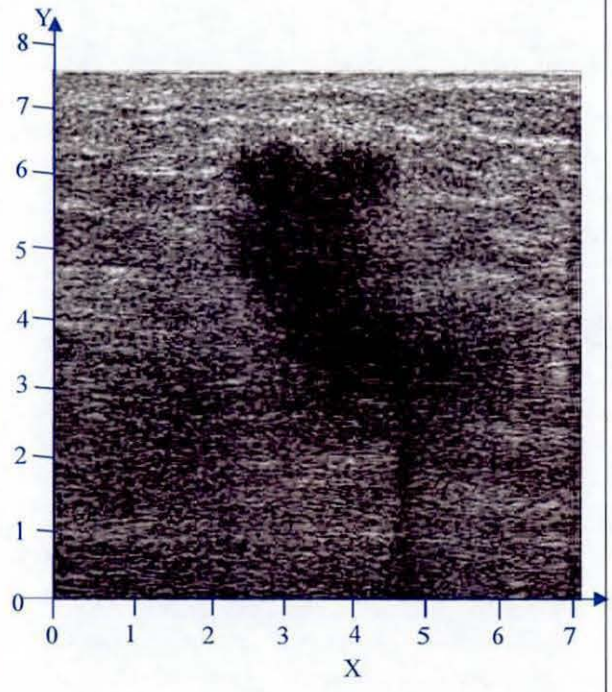
Any object in the image?
1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose <u>4 or 5</u>, please describe the features:

Object location (approximation):
X: to
Y: to

Object size:
Height > width ☐
Width > height ☐
Width = height ☐
Unknown ☐

Object Boundary:
Smooth ☐
Irregular ☐
Unknown ☐



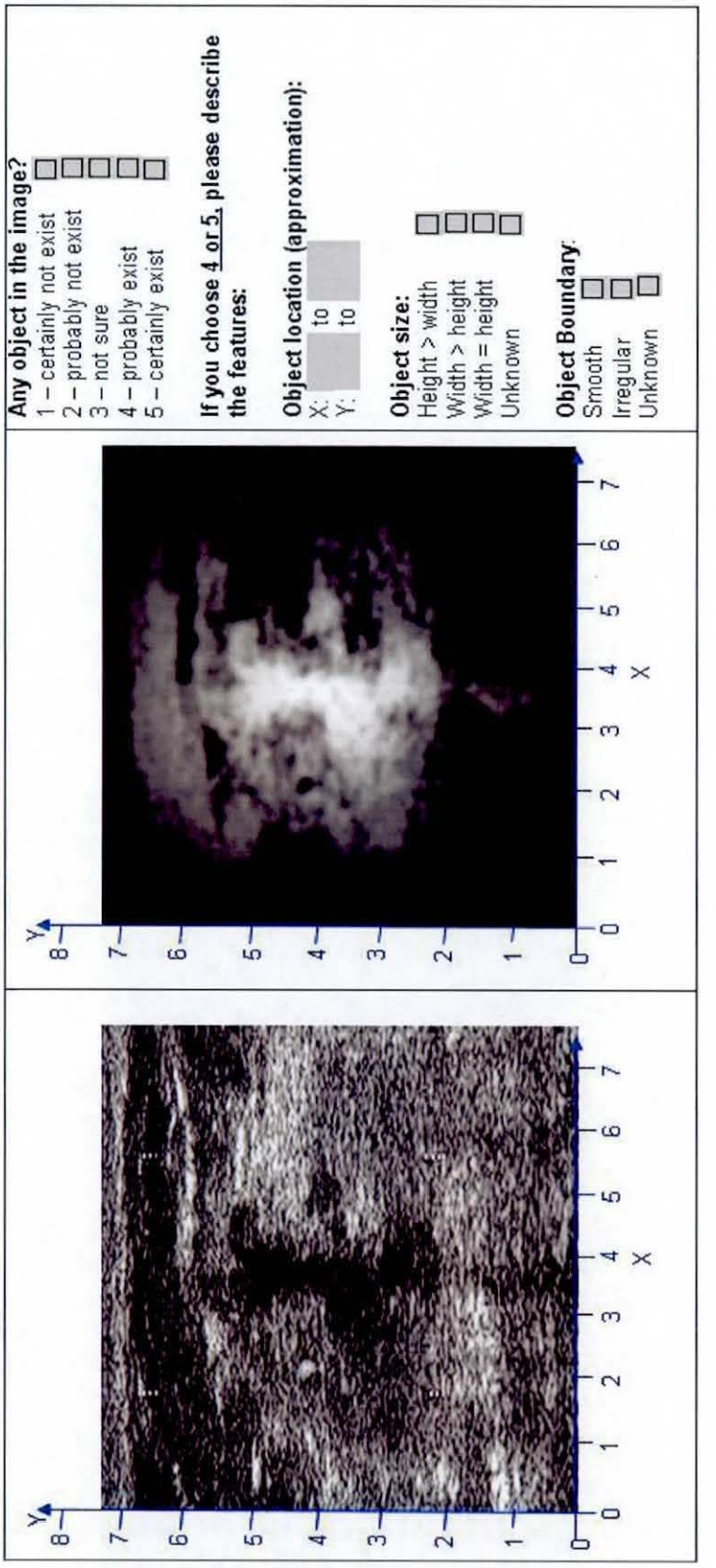
Any object in the image?
1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

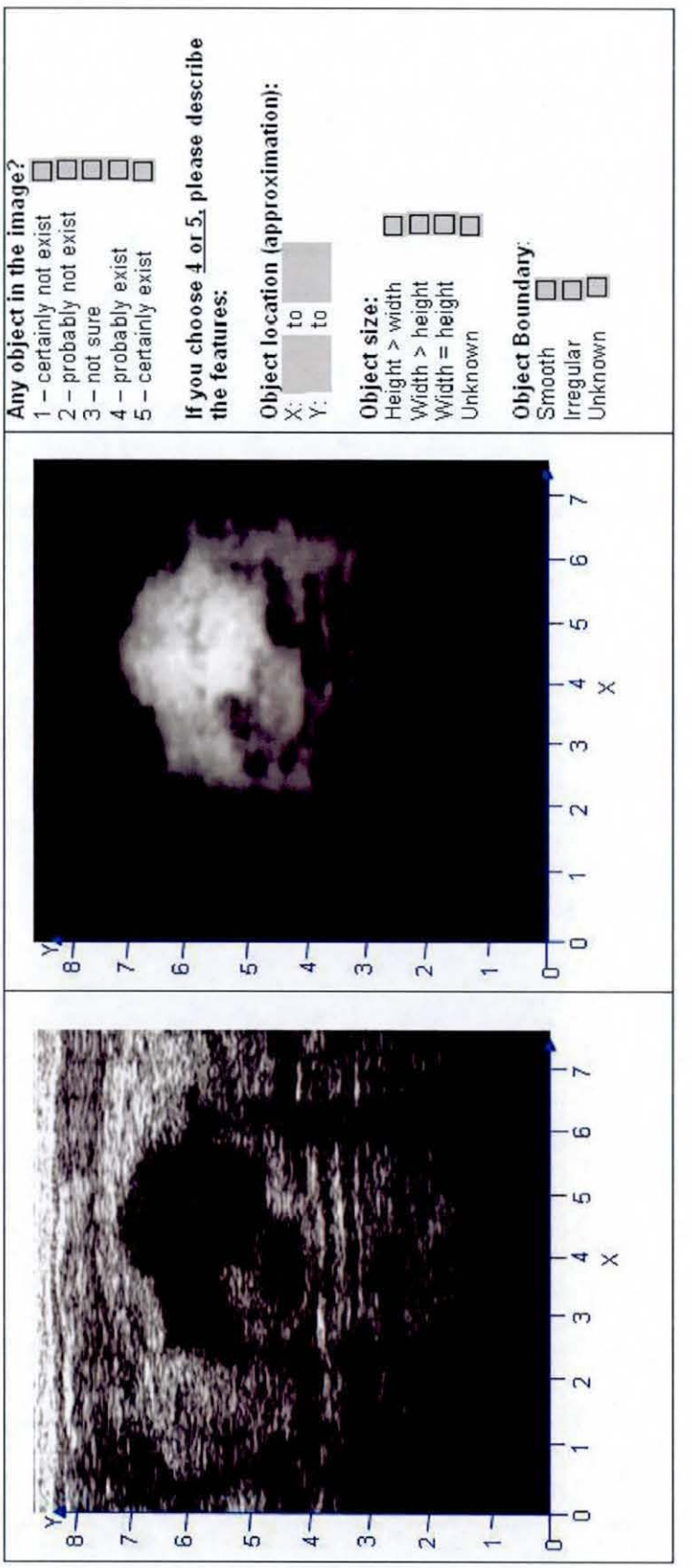
If you choose <u>4 or 5</u>, please describe the features:

Object location (approximation):
X: to
Y: to

Object size:
Height > width ☐
Width > height ☐
Width = height ☐
Unknown ☐

Object Boundary:
Smooth ☐
Irregular ☐
Unknown ☐


Any object in the image?
1 – certainly not exist
2 – probably not exist
3 – not sure
4 – probably exist
5 – certainly exist
If you choose 4 or 5, please describe the features:
Object location (approximation):
X: to
Y: to
Object size:
Height > width
Width > height
Width = height
Unknown
Object Boundary:
Smooth
Irregular
Unknown
Y
X
0 1 2 3 4 5 6 7 8


Any object in the image?
1 – certainly not exist
2 – probably not exist
3 – not sure
4 – probably exist
5 – certainly exist
If you choose 4 or 5, please describe the features:
Object location (approximation):
X: to
Y: to
Object size:
Height > width
Width > height
Width = height
Unknown
Object Boundary:
Smooth
Irregular
Unknown
Y
X
0 1 2 3 4 5 6 7
1 2 3 4 5 6 7 8

Any object in the image?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose 4 or 5, please describe the features:

Object location (approximation):

X: ____ to ____
Y: ____ to ____

Object size:

Height > width ☐
Width > height ☐
Width = height ☐
Unknown ☐

Object Boundary:

Smooth ☐
Irregular ☐
Unknown ☐

Any object in the image?

1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose 4 or 5, please describe the features:

Object location (approximation):

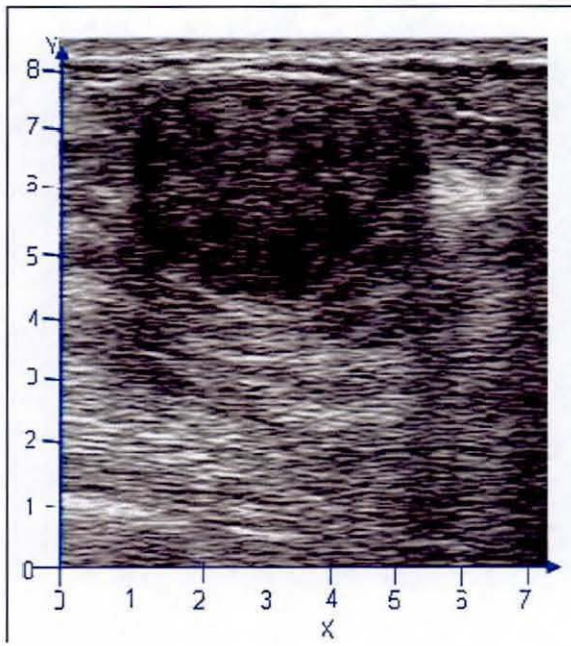
X: ____ to ____
Y: ____ to ____

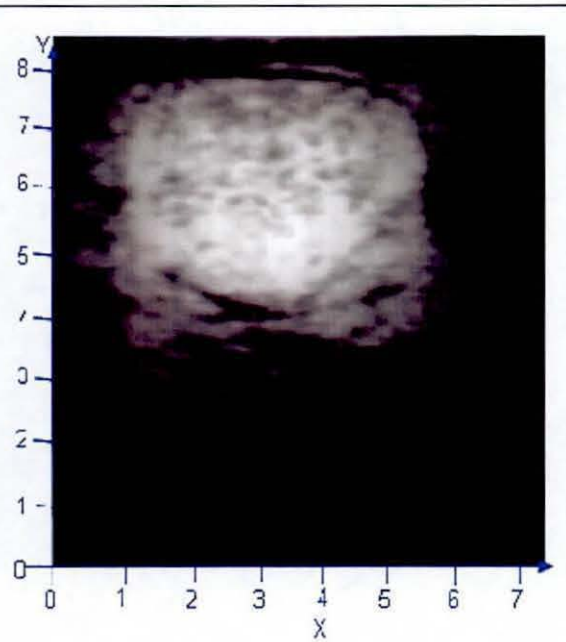
Object size:

Height > width ☐
Width > height ☐
Width = height ☐
Unknown ☐

Object Boundary:

Smooth ☐
Irregular ☐
Unknown ☐





Any object in the image?
1 – certainly not exist ☐
2 – probably not exist ☐
3 – not sure ☐
4 – probably exist ☐
5 – certainly exist ☐

If you choose 4 or 5, please describe the features:

Object location (approximation):
X: ____ to ____
Y: ____ to ____

Object size:
Height > width ☐
Width > height ☐
Width = height ☐
Unknown ☐

Object Boundary:
Smooth ☐
Irregular ☐
Unknown ☐

This research is to develop an efficient training system in breast cancer detection. Thanks for your kind contribution (to save life)!

Appendix B: List of Publications

Journal papers:

[P1] M.H. Yap, E.A. Edirisinghe, H.E. Bez. "A Novel Algorithm for Initial Lesion Detection in Ultrasound Breast Images", Journal of Applied Clinical Medical Physics, vol. 9, issue 4, pp 181-199, Fall 2008, ISSN: 15269914.

[P9] M.H. Yap, E.A. Edirisinghe, A.G.Gale, H.E. Bez."Processed Images in Human Perception: A Case Study in Ultrasound Breast Imaging". European Journal of Radiology (EJR), Elsevier. In press.

Conference papers:

[P3] M.H. Yap, E.A. Edirisinghe, H.E. Bez. "Object Boundary Detection in Ultrasound Images", crv p.53,Third Canadian Robotics and Machine Vision Conference (CRV'06), 2006.
(DOI: http://doi.ieeecomputersociety.org/10.1109/CRV.2006.51)

[P4] M.H. Yap, E.A. Edirisinghe, H.E. Bez. "Initial Lesion Detection in Ultrasound Breast Images", Visualization, Imaging, and Image Processing, VIIP 2006 (28-30th Aug), pp 215-220.

[P5] M.H. Yap, E.A. Edirisinghe, H.E. Bez, H.T. Ewe, "Initial Lesion Detection and Region of interest Labeling in Ultrasound Breast Images", International Conference on Vision Information Engineering, VIE 2006 (26-28th Sep), pp 333-338, ISBN 0 86341 671 3/9.

[P6] Yap, M.H., Edirisinghe, E.A., Ewe, H.T. and Bez, H.E., "Multifractal Analysis: A Novel Approach to Ultrasound Breast Imaging", Proceedings of the 12th Chinese Automation & Computing Society Conference in the UK, 12th Chinese Automation & Computing Society Conf. in the UK, Loughborough, UK, 2006, pp 150-155, ISBN 0 9533890 9 X.

[P7] M.H. Yap, E.A. Edirisinghe, H.E. Bez. "Fully Automatic Lesion Boundary Detection in Ultrasound Breast Images", SPIE Medical Imaging Conference,651265123I, 17th-22nd February,2007, San Diego, CA, US.

[P8] M.H. Yap, E.A. Edirisinghe, A.G..Gale, H.E. Bez. "Evaluation of the Effects of Processed Images in Human Performance: A Case Study in Ultrasound Breast Imaging", IET The 4th Visual Information Engineering Conference, 25-27th July 2007, London, ISBN 978-0-86341-830-3.

[P9] M.H. Yap, E.A.Edirisinghe, A.G..Gale, H.E.Bez. "The Effects of Computer Processed Images in Human Performance", Radiological Society of North America (RSNA 2008, LL-IN2068-D08), 1st December 2008, Chicago.

[P10] M.H.Yap, E.A.Edirisinghe, H.E.Bez, "A Comparative study in Ultrasound Breast Imaging Classification". SPIE Medical Imaging Conference, 7th-12th Feb 2009. Accepted.

Journal paper under preparation:

[P11] M.H.Yap, E.A.Edirisinghe, H.E.Bez. "Enhanced algorithms for Lesion Detection & Recognition in ultrasound images", Medical Image Analysis, Elsevier.

Appendix C: Journal Papers

JOURNAL OF APPLIED CLINICAL MEDICAL PHYSICS, VOLUME 9, NUMBER 4, FALL 2008

A novel algorithm for initial lesion detection in ultrasound breast images

Moi Hoon Yap, Eran A. Edirisinghe, and Helmut E. Bez
Department of Computer Science, Loughborough University, Loughborough, U.K.

M.H.Yap@lboro.ac.uk

Received 6 August 2007; accepted 3 June 2008

This paper proposes a novel approach to initial lesion detection in ultrasound breast images. The objective is to automate the manual process of region of interest (ROI) labeling in computer-aided diagnosis (CAD). We propose the use of hybrid filtering, multifractal processing, and thresholding segmentation in initial lesion detection and automated ROI labeling. We used 360 ultrasound breast images to evaluate the performance of the proposed approach. Images were preprocessed using histogram equalization before hybrid filtering and multifractal analysis were conducted. Subsequently, thresholding segmentation was applied on the image. Finally, the initial lesions are detected using a rule-based approach. The accuracy of the automated ROI labeling was measured as an overlap of 0.4 with the lesion outline as compared with lesions labeled by an expert radiologist. We compared the performance of the proposed method with that of three state-of-the-art methods, namely, the radial gradient index filtering technique, the local mean technique, and the fractal dimension technique. We conclude that the proposed method is more accurate and performs more effectively than do the benchmark algorithms considered.

PACS numbers: 87.57.Nk

Key words: medical image analysis, ultrasound imaging, region-of-interest, hybrid filtering, multifractals

I. INTRODUCTION AND SCOPE

Breast cancer is the most common of all cancers affecting women in the developed countries.[1] More middle-age women die of breast cancer than of any other single cause.[1] In the United Kingdom, more than 41 000 cases are diagnosed annually, and it is predicted that 1 in every 9 women will develop breast cancer at some point in life.[2] Early detection plays a significant role in the fatality of breast cancer. Technologies that aid in the early detection of cancers have therefore attracted much attention from the research community.

Mammography and ultrasound imaging are the standard technologies used in cancer screening. Mammography is accepted as the "gold standard" for breast imaging. It is widely used as the primary tool for cancer screening. However, in diagnostic workup, mammography and breast ultrasound are often used as complementary investigations. Mammography has been shown to cause high false-positive rates in diagnosis, and the radiation dose to the breast is harmful.[3] Further, cost considerations have resulted in most countries choosing to use screen film mammography instead of a digitized version. However, the inability to change image contrast or brightness, problems in detecting subtle soft-tissue lesions (dense glandular tissues), and difficulties with archiving have limited the application of screen film mammography.

[a] Corresponding author: Moi Hoon Yap, Department of Computer Science, Loughborough University, Holywell Park, U.K.; phone: +44(0)15-0963-5739; fax: +44(0)19-0963-5709; email: M.H.Yap@lboro.ac.uk

181

Ultrasound studies have been shown to be good at picking up many of the cancers missed by mammography, especially in women who have dense breasts. In addition, ultrasound is noninvasive, portable, and versatile. Further, it does not use ionizing radiation, and it is relatively lower in cost. However, ultrasound images have a major disadvantage: poor quality because of multiplicative speckle noise that results in artifacts. Segmentation of lesions in ultrasound images is therefore a challenging task that remains an open problem despite many past research efforts.

In most existing breast screening approaches, the "initial lesion"—that is, a suspect region—is manually located by a trained radiologist in a pre-processing stage by marking its topmost, leftmost, bottommost, and rightmost boundary limits with crosses. These crosses (and hence the initial lesion) are then manually encompassed within a rectangular region of interest (ROI)[4] by the radiologist and subsequently presented to a computer-aided diagnosis (CAD) system for further analysis leading to the segmentation and classification of the tumor. Because the selection of extreme points (that is, the crosses) and the rectangular region both require human intervention, these steps are open to subjectivity and human error. As a result, a well-trained and experienced examiner with knowledge of the normal echo anatomy of the breast and the changes caused by pathology is required for accurate breast cancer screening. A failure to include an entire lesion within the ROI or to outline ROIs containing lesions (among other errors) can severely undermine the performance of the CAD system. Further, it has also been shown that radiologists with varying training backgrounds and experiences often reach rather different results in the reading of sonograms.[5]

Our current research focus is to provide the radiologist with an automated tool that can effectively assist in the selection of the ROI and in improving the consistency of interpretation. However, it is worthwhile noting that the automatic detection of ROIs is not meant to replace the radiologist, but to provide a tool to reduce the radiologist's ROI labeling time (see Section II) and to warn of possible ROIs that might otherwise be missed because of the poor quality of the ultrasound image.

Evaluation of the performance of the proposed algorithm requires a suitable test image database, a suitable evaluation metric, and a design goal. Because of the practical difficulties in obtaining databases with ultrasound images of normal and nearly normal breasts, previous state-of-the art algorithms[6,7] of initial lesion detection have used U.S. image databases that consist solely of malignant and benign tumors. In our experiments, we used a U.S. image database of 360 malignant and nonmalignant images that additionally contains location information concerning the extreme points of tumors (marked with crosses by a number of expert radiologists). The metric used to evaluate the performance of the algorithm on the database is an "overlap" figure[6] defined as the ratio of the intersection and the union of the two lesion areas that were manually identified by the radiologists and by the computer-based algorithm. Specifically, in work by Drukker et al.[6] and Yap et al.,[7] the design goal has been to achieve an overlap value of in excess of 0.4, which can be represented as

$$overlap = \frac{X \cap Y}{X \cup Y} \geq 0.4, \tag{1}$$

where X is the lesion area extracted by the computer-based algorithm, and Y is the lesion labeled manually by the radiologists. In other words, the assumption has been that, if an overlap of 0.4 or more results, then the computer-based algorithm has been successful in accurately and automatically performing the otherwise manual task of initial lesion identification. A secondary metric, "accuracy," can therefore be defined as the percentage of experiments obtaining an overlap value of beyond 0.4. In the present research context, we use "accuracy" as the objective metric to evaluate and compare the results of the performance of the proposed algorithm with that of state-of-the-art methods. Further subjective results have been illustrated for visual comparison.

The present paper is divided into five sections for clarity of presentation. Section II discusses our research motivation and the existing solutions to the problem. Section III introduces the proposed approach and provides a detailed discussion of each stage. Section IV sets out the experimental results and discusses the results in detail. Finally, Section 5 concludes with further directions for improvement and research.

II. MOTIVATION AND RESEARCH BACKGROUND

Many ongoing ultrasound breast imaging research projects are focused on creating CAD tools that have high sensitivity, specificity, and consistency in lesion classification. Examples include Boone,[8] Gurney,[9] Boukerroui et al.,[10] Chen et al.,[5] and Sehgal et al.[11] Unfortunately, these systems are based on the assumption that the ROI will be pre-selected by a radiologist and that the analysis will be performed only on the cropped ROI. This requirement improves accuracy in detecting lesion shape because the noisy, dark, poor-quality surrounding areas can be excluded from consideration because of the manual selection of the specific ROI. It therefore follows that a fully automated CAD tool used in cancer screening will require a preprocessing stage that is capable of automatic ROI labeling. Our work is focused on providing an effective solution to that problem (see Fig. 1).

Two existing state-of-the-art techniques for automatic ROI labeling of existing breast images, used as benchmarks to compare the performance of the proposed technique, can be summarized as follows:

- Drukker et al.[6] investigated the use of the radial gradient index (RGI) filtering technique to automatically detect lesions on breast ultrasound images. In ultrasound images, lesions are almost invariably darker than the background. Thus, in Drukker's work, the grayscale of the original ultrasound images is initially inverted. Subsequently, images are pre-processed by a

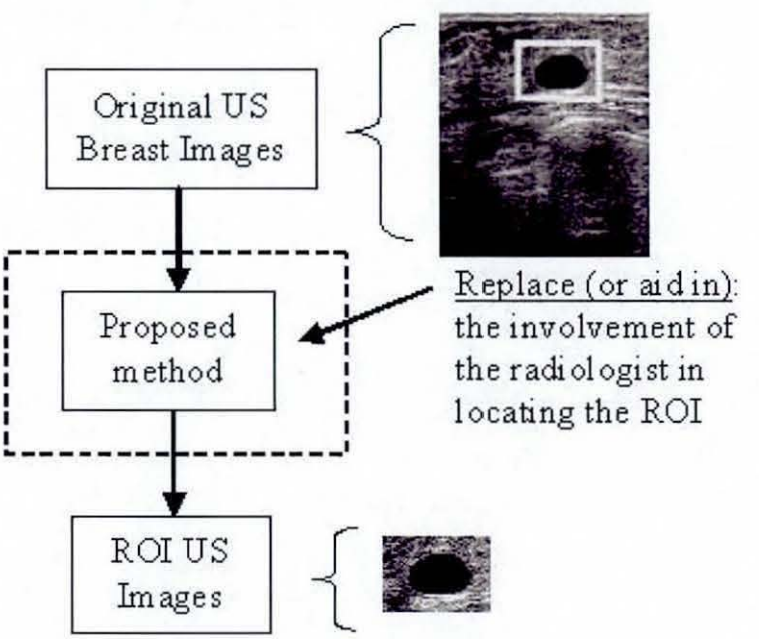


FIG. 1. Use of the proposed solution. US = ultrasound; ROI = region of interest.

median filter to remove speckle noise. The resulting image is fed to a RGI filter.[6] In RGI filtering, the images are sub-sampled by a factor of 4. A brightness threshold (see Subsection III.E) for the RGI-filtered image is varied iteratively from 0.74 to 0.62 until either at least one lesion of interest is detected. The detected areas smaller than 5 mm^2 are discarded. Lesion candidates are segmented from the background by maximizing an average radial gradient (ARD) index for regions grown from the detected points. According to Drukker et al., maximizing the ARD index is more accurate than maximizing the RGI index.[6] At an overlap level of 0.4 with lesions outlined by a radiologist, 75% accuracy of lesion detection was reported for the test set of ultrasound images used.

- Yap et al.[7] analyzed the use of statistical methods (for example, local mean) and values of fractal dimensions in initial lesion detection. The images were preprocessed using histogram equalization,[12] and then hybrid filtering (see Subsection III.C) and marker-controlled watershed segmentation were applied. (A "watershed" is the ridge that divides areas drained by different river systems. A catchments basin is the geographic area draining into a river or reservoir. The watershed transform applies these ideas to grayscale image processing in a way that can be used to solve a variety of image segmentation problems.) Marker-controlled watershed segmentation is an approach based on the concept of markers to control oversegmentation in watershed transform.[13] The minimum local mean and the minimum of the fractal dimensions (see Subsection III.D) of the identified segments were then used to identify the initial lesion. Subsequently, neighborhood segments are identified, and these are finally combined to form the ROI. The accuracy of the automated ROI labeling is measured by an overlap of 0.4 between its lesion outline and the lesions labeled by the radiologists. The accuracy of ROI detection when using local mean was reported to be 69.21%; fractal dimension was 54.21%.

The proposed approach detailed in the next section intends to further extend the accuracy of ROI detection by following an effective multistage algorithm.

III. PROPOSED APPROACH

Fig. 2 shows a modular block diagram of the proposed technique. It uniquely combines histogram equalization as a preprocessing stage, followed by hybrid filtering, multifractal analysis, thresholding segmentation, and a rule-based approach in fully automated ROI labeling.

The operation and functionality of the individual stages are described in detail in the subsections that follow. All algorithms were implemented in LTI-Lib[14] on a Linux platform.

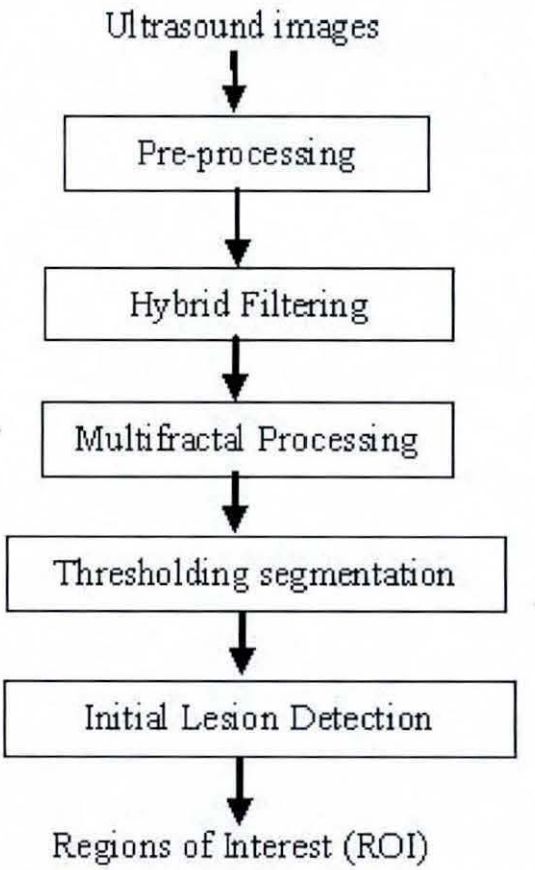


FIG. 2. Overview of the methodology.

A. Ultrasound images

In general, ultrasound images are complex because of data composition, which can be described in terms of speckle information. Upon visual inspection, speckle noise consists of a relatively high-level gray intensity, qualitatively ranging between hyperechoic (bright) and hypoechoic (dark) domains.[15]

Notably, any automatic system that is designed to detect abnormal lesions in ultrasound images should, in the end, be verified or compared with the judgment of a medical expert or radiologist. The test images used in our work were obtained from a professionally compiled compact disc (CD) of breast ultrasound images,[16] which consists of explanations and verifications from several qualified expert radiologists. We selected 360 images from the CD for our experiments. Of the 360 images, 20 were malignant, 76 were simple cysts, 76 were complex cysts, 58 were fibroadenomas, 38 were carcinomas, 18 were occult lesions, 15 were adenosis lesions, and 59 were a combination of other diagnoses. Each image had been manually processed by an expert radiologist, and the extreme points of the suspected lesions were already marked with crosses. Because the objective metric of the experiment (see Section I) required identification of the lesion boundary marked by a radiologist (rather than the extreme points or ROI), we obtained the services of an expert radiologist to mark these boundaries manually and to verify the extreme points given in the ultrasound CD.[16]

B. Preprocessing

As mentioned previously, the credibility of a high-quality breast ultrasound examination depends on the scanner (that is, the quality of the original image) and the experience of the examiner. The preprocessing stage deals with the issue of guaranteeing the homogeneity of the original ultrasound images, which thus subsequently improves the chances of lesion ROI detection being more accurate. In the proposed approach, we use a histogram equalization strategy tested in earlier experiments[17] as a preprocessing stage to achieve the homogeneity guarantee.

Histogram equalization[12] is similar to contrast stretching, in that it attempts to increase the dynamic range of the pixel values in an image. However, unlike contrast stretching, no provision is made for interactivity, because applying a histogram equalization algorithm to an image with a fixed number of bins will always yield the same result. Let

$$p_r(r_j), j = 1,2,...,L \tag{2}$$

denote the histogram associated with the intensity levels of a given image, and recall that the values in a normalized histogram are approximations to the probability of occurrence of each intensity level in the image. For discrete quantities, the equalization transformation becomes

$$s_k = T(r_k) = \sum_{j=1}^{k} p_r(r_j) = \sum_{j=1}^{k} \frac{n_j}{n} \tag{3}$$

for $k = 1,2,...,L$, where s_k is the intensity value in the output (processed) image corresponding to value r_k in the input image, n is the total number of pixels, and n_j is the number of pixels in bin j.

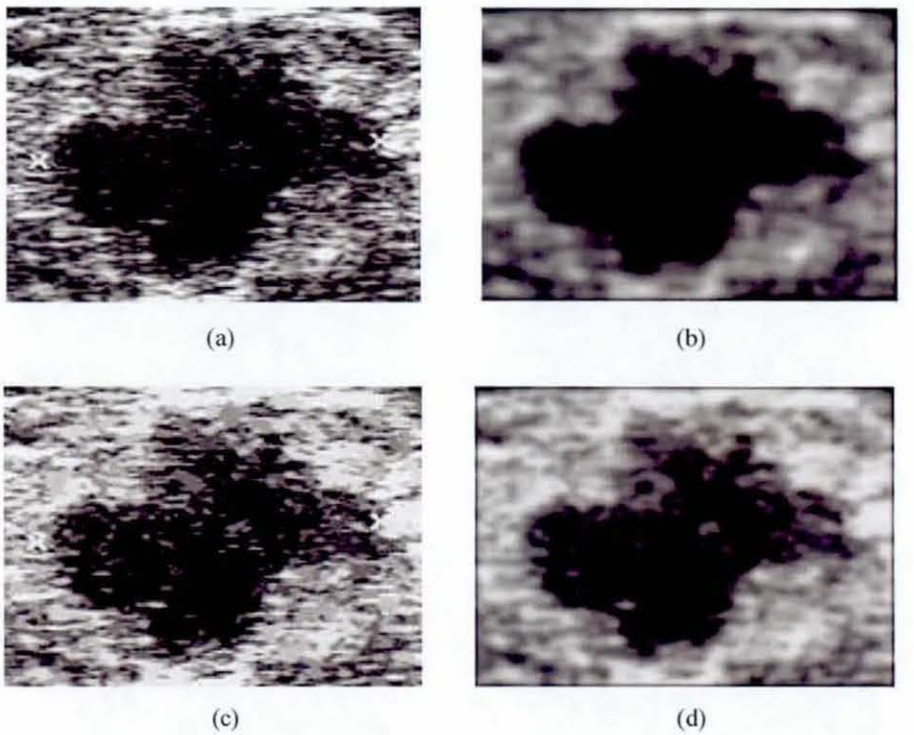
C. Hybrid filtering

The function of the filtering stage is to remove noise, which is a major obstacle to accurate segmentation of the images. Median filtering is a popular approach used in removing speckle noise in ultrasound images.[6,18,19] However, Yap et al.[17] proved that part of the reason for the inaccuracy of the boundary detection in Drukker et al.,[6] Joo et al.,[18] and Kupinski et al.[19] was that, although the median filter managed to filter out the speckle noise, it also removed the important edge information—in particular, edges that belonged to the lesion.

Further, Gaussian blur[4] is a linear filtering technique that has been widely used to reduce the oversegmentation problem in ultrasound images. Gaussian blur is very effective in removing speckle noise, but it blurs and dislocates edges,[20] which may negatively affect subsequent lesion segmentation. Perona and Malik[21] proposed a nonlinear partial differential equation approach for smoothing images on a continuous domain. Anisotropic diffusion was shown to perform well for images corrupted by additive noise. However, in cases where images contain speckle noise, anisotropic diffusion enhances that noise instead of eliminating it.[22] On a more positive note, nonlinear diffusion filtering,[23] as compared with linear diffusion, has deservedly attracted much attention in the field of image processing for its ability to reduce noise while preserving (or even enhancing) important features of the image such as edges or discontinuities.

Within the context of proposed research, we make use of a hybrid filtering approach that combines the strength of nonlinear diffusion filtering to produce edge-sensitive speckle reduction, with linear filtering (Gaussian blur) to smooth the edges and to eliminate oversegmentation. The result of hybrid filtering is visually compared in Fig. 3. Section IV sets out our experimental results and a detailed analysis to justify the use of hybrid filtering as compared with either nonlinear diffusion filtering or Gaussian blur alone.

Subsequent to hybrid filtering, we use multifractals[24] to further enhance the partially processed images. Section IV sets out experimental results to show that this stage allows for better segmentation of lesions than does the application of hybrid filtering only. Subsection D, next, provides a brief overview of multifractals and associated analysis methodology.



(a) (b)

(c) (d)

FIG. 3. Comparison of filtering approaches: (a) original image; (b) Gaussian blur; (c) nonlinear diffusion filtering; (d) hybrid filtering.

D. Multifractal dimensions

A fractal is generally "a rough or fragmented geometric shape that can be subdivided in parts, each of which is (at least approximately) a reduced-size copy of the whole,"[25] a property called self-similarity. In analyzing the fractal geometry of an image, an attempt is made to exploit the self-similarity present. In fractal geometry, the term "fractal dimension" refers to a statistical quantity that indicates how completely a fractal appears to fill space, as one zooms down to finer and finer scales.[25] Multifractal analysis refers to the analysis of an image using multiple fractals—that is, not just one fractal, as in fractal analysis.

The generalized formulation for multifractal dimensions D of order q[24] can be represented as

$$D_q = \begin{cases} \dfrac{1}{q-1}\lim\limits_{\varepsilon\to 0}\dfrac{\log(x_q(\varepsilon))}{\log(\varepsilon)} & \text{for } q \in \Re \text{ and } q \neq 1 \\ \lim\limits_{\varepsilon\to 0}\dfrac{\sum_i \mu_i \log \mu_i}{\log(\varepsilon)} & \text{for } q = 1 \end{cases}, \tag{4}$$

where ε is the linear size of the cells (in our case, because we use a 3×3-pixel mask, $\varepsilon = 3$), and q is the order for cell size ε. Note that when $q < 0$, D_q is sensitive to the parts where the measure is very dense. On the other hand, if $q > 0$, information on the sparse region can be obtained. In theory, q is in the range $-\infty$ to ∞, and D_q can have an infinite number of values. In practice, computing for all values of q is not possible. Hence, we empirically decided to use only four values of q in our experiments: -1, 0, 1, and 2.

The partition function χ_q is defined as

$$x_q(\varepsilon) = \sum_{i=1}^{N(\varepsilon)} \mu_i^q(\varepsilon), \tag{5}$$

where $N(\varepsilon)$ is the total number of cells of size ε, and $\mu_i^{\alpha}(\varepsilon)$ is the measure that is defined on a given set. In this case, the measures are defined as the probability of the grayscale level in the images, where all the gray levels fall in the range of $0 - 1$.

To investigate the effect of various q values on ultrasound images, we carried out further empirical experiments.

The graph in Fig. 4 shows that the multifractals with $-\infty < q < \infty$ yields four types of results. We can classify those results into $q = 0$, $q = 1$, $0 < q < 1$, and ($q < 0$ or $q > 1$). Note that $0 < q < 1$ is the inverted value of ($q < 0$ or $q > 1$). From the results illustrated in Fig. 5, it is seen that any value range $0 < q < 1$ or ($q < 0$ or $q > 1$) will help improve the segmentation results. The value $q = -1$ (that is, D_{-1}) is chosen in our work because, as compared with other values, it has the lowest associated computational complexity.

To our knowledge, hybrid filtering has not been used in previous research in breast ultrasound boundary detection or contrast enhancement, and hence its use is an additional novel aspect of our approach.

After application of the hybrid and multifractal filtering, the images are ready for lesion segmentation. Subsection E introduces the approach adopted.

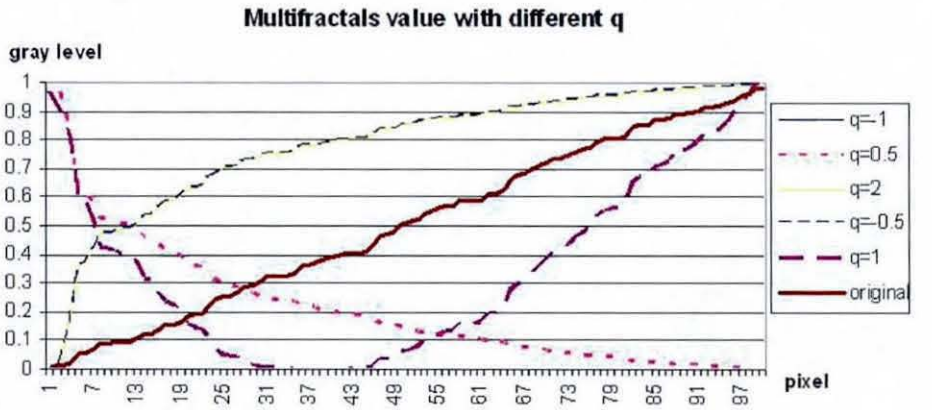


FIG. 4. Graph of multifractals dimension with various values of q.

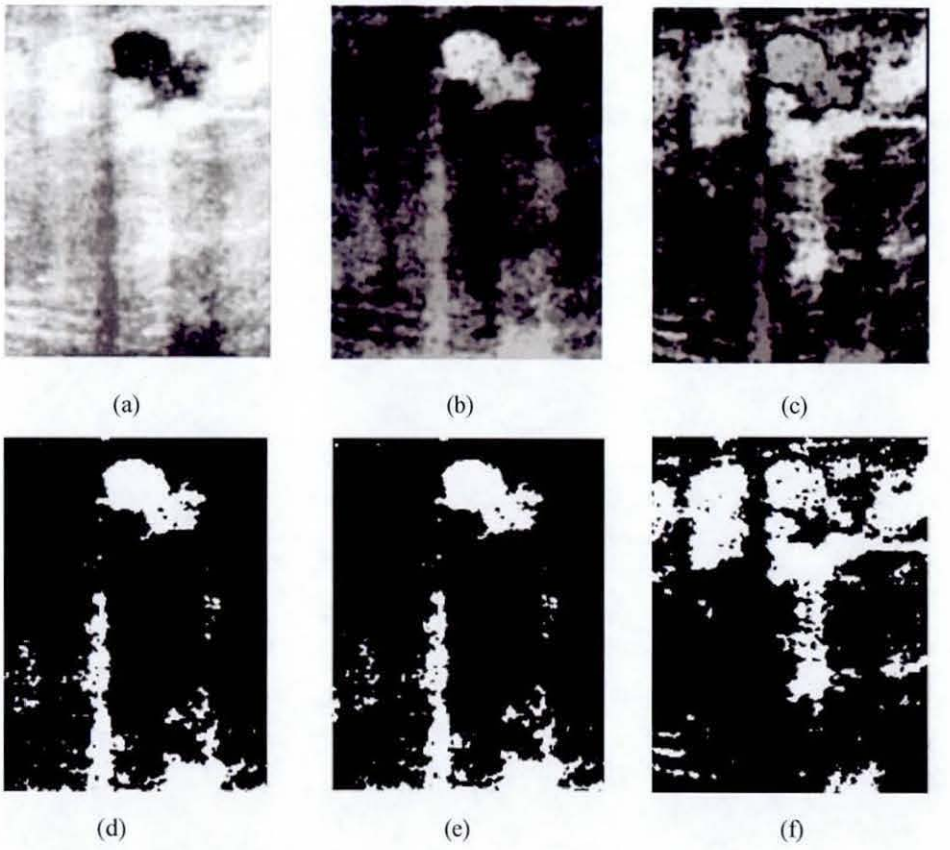


FIG. 5. Result of multifractal analysis with (a) $q < 0$ or $q > 1$; (b) $0 < q < 1$; (c) $q = 1$; (d) threshold of (a); (e) threshold of (b); and (f) threshold of (c).

E. Thresholding segmentation

In general, segmentation is a process used to divide an image into its constituent parts. Thresholding segmentation[26] is the most basic, the simplest, and the fastest algorithm in segmentation. A thresholding procedure attempts to determine an intensity value, called the "threshold," that separates pixels into desirable classes. A parameter θ, called the "brightness threshold" is chosen and applied to the image $a[m, n]$ as follows:

$$\begin{aligned} &\text{If } a[m, n] \,^{3}\, q \\ &\text{then } a[m.n] := 1 \text{ (object)} \\ &\text{else } a[m.n] := 0 \text{ (background).} \end{aligned} \tag{6}$$

Within our present research context, we used a fixed threshold for segmentation—that is, a threshold chosen independently of the image data. If it is known that very high-contrast images (where the objects are very dark and the background is homogeneous and very light) are being dealt with, then a constant threshold of 128 on a scale of 0 to 255 was experimentally found to be sufficiently accurate. That is, the number of falsely-classified pixels is kept to a minimum. The sensitivity of the threshold selection on segmentation accuracy was further found to be low, justifying the use of a image-independent fixed threshold.

F. Initial lesion detection

The thresholding segmentation of Subsection III.E often leads to the identification of multiple ROIs, of which generally only one or two would be of diagnostic importance (that is, belonging to abnormal lesions). Further, the location of the abnormal lesions requires the specification of both position and orientation (assuming two-dimensional ultrasound images). We therefore propose the use of a rule-based approach to identify these important ROIs.

The first criterion for the identification of lesions is the size of the segments. The suspect lesions are identified as the largest segments among the likely multiple segments that result from applying the single-threshold segmentation. In addition, based on the additional guidance provided in the ultrasound image CD from which we obtained our test data,[15] we observed that 95% of tumors are located at the upper regions of the images. Hence, a reference point at (x, y), where

$$x = \frac{image\ height}{3},\ and\ y = \frac{image\ width}{2} \tag{7}$$

is chosen as the center of attention. The lesion that is located closest to that point and that satisfies the above mentioned size-related criterion is selected as the final detected lesion. The size-related criterion can be appropriately slightly relaxed if more than one lesion must be detected.

IV. RESULTS AND ANALYSIS

Fig. 6 compares the effects of using the various filtering techniques discussed in Section III with multifractal processing. When Gaussian filtering is used [see Fig. 6(b)], the smoothing effects introduced at the edges result in noisy regions that ideally should be disconnected remaining connected. This effect is a problem in the subsequent lesion detection stage, because the largest connected region may not now refer to the true lesion that should be detected. The use of nonlinear filtering results in oversegmentation and causes many problems, as illustrated by Fig. 6(c). However, with the use of hybrid filtering and multifractal processing, the single largest segmented area detected is identified as the lesion.

Fig. 7 compares final lesion boundary detection accuracies for the use of hybrid filtering alone and for hybrid filtering followed by multifractal processing. The boundary detection accuracy in Fig. 7(e) is better than that in Fig. 7(c), justifying the positive contribution of multifractal processing to subsequent boundary detection. The boundary in Fig. 7(c) is closer to the boundary that a typical radiologist might intend to identify, because the boundary is smoother and excludes effects of noise and other artifacts more effectively.

Figs. 8, 9, and 10 illustrate the stage-by-stage operation of the algorithms from Drukker et al.,[6] Yap et al.,[7] and the present proposal. These three algorithms use different steps in identifying lesions. The intentions at each stage within the three algorithms are significantly different, and hence, the intermediate results are not comparable. A more comprehensive performance comparison based on the accuracy of the final detected lesion regions is given later in the present paper.

Our detailed experiments, performed using the 360 test ultrasound images, revealed that the proposed method performs exceptionally well in identifying ROIs for most cyst lesions and malignant lesions, and for some fibroadenoma lesions. Because of the high degree of similarity in texture between normal and fibroadenoma regions, accurate identification of such regions is always a challenge. Fig. 11 provides a visual comparison of the results of using the proposed algorithms on various types of abnormalities. Fig. 12 illustrates two examples of fibroadenoma ROIs that are not detected accurately.

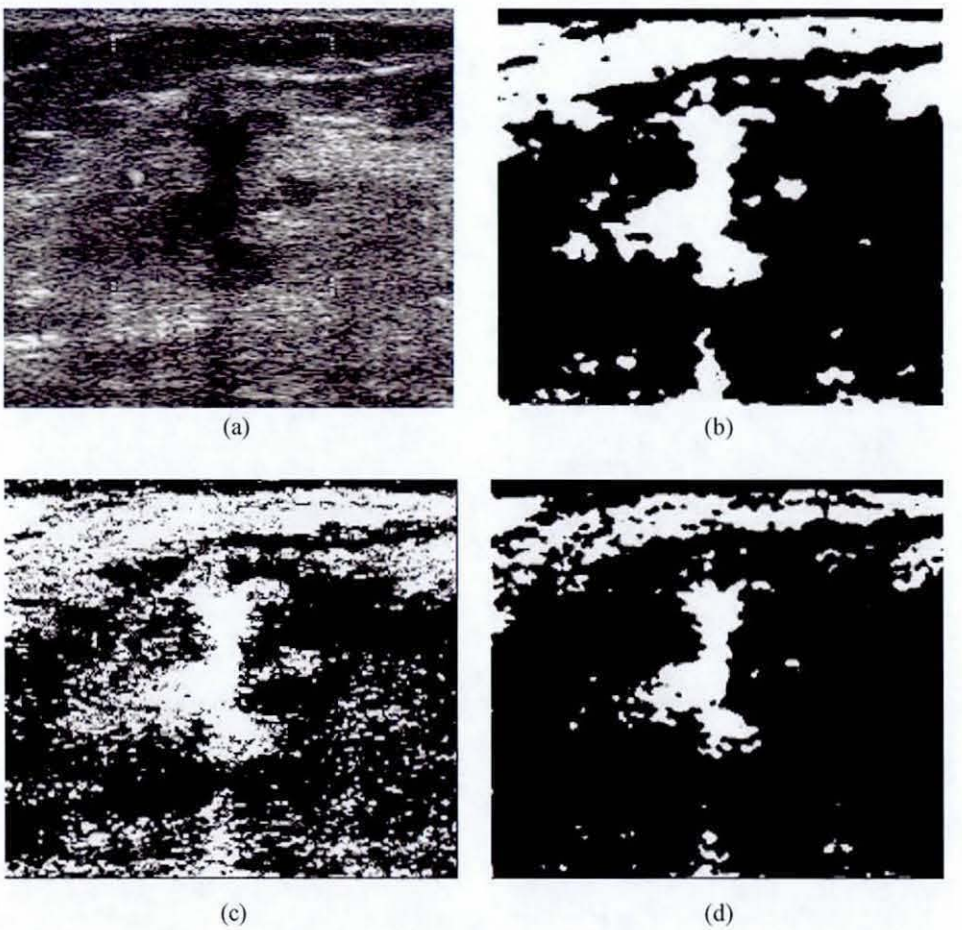


FIG. 6. Comparison of segmentation approaches: (a) original image, and use of (b) Gaussian plus multifractal filtering, (c) nonlinear plus multifractal filtering, (d) hybrid plus multifractal filtering.

Based on diagnoses by radiologists as provided on the test ultrasound breast imaging CD used in our experiments,[16] Table 1 summarizes the accuracy figures obtained for each type of abnormality for the three benchmark algorithms and the proposed algorithm. Fig. 13 graphically presents ROI detection accuracy for each of the four algorithms and each type of abnormality. The results clearly show the improvements obtainable with the proposed improved approach to ROI lesion detection accuracy. The detection accuracy for fibroadenoma-type lesions has generally been the lowest for all methods, but in this category, the proposed algorithm shows a 15% accuracy improvement as compared with the methods of Drukker et al.[6] and Yap et al.[7] Further, when using the proposed algorithm, 90% accuracy in detecting malignant-type lesions is indicated.

Fig. 14 visually compares the performance of the proposed approach with that of the three benchmark algorithms. This comparison clearly illustrates the improved accuracy of lesion identification demonstrated by the proposed approach. The three benchmark algorithms can be seen to be more likely to identify non-lesion regions—or only parts of lesions—as ROIs.

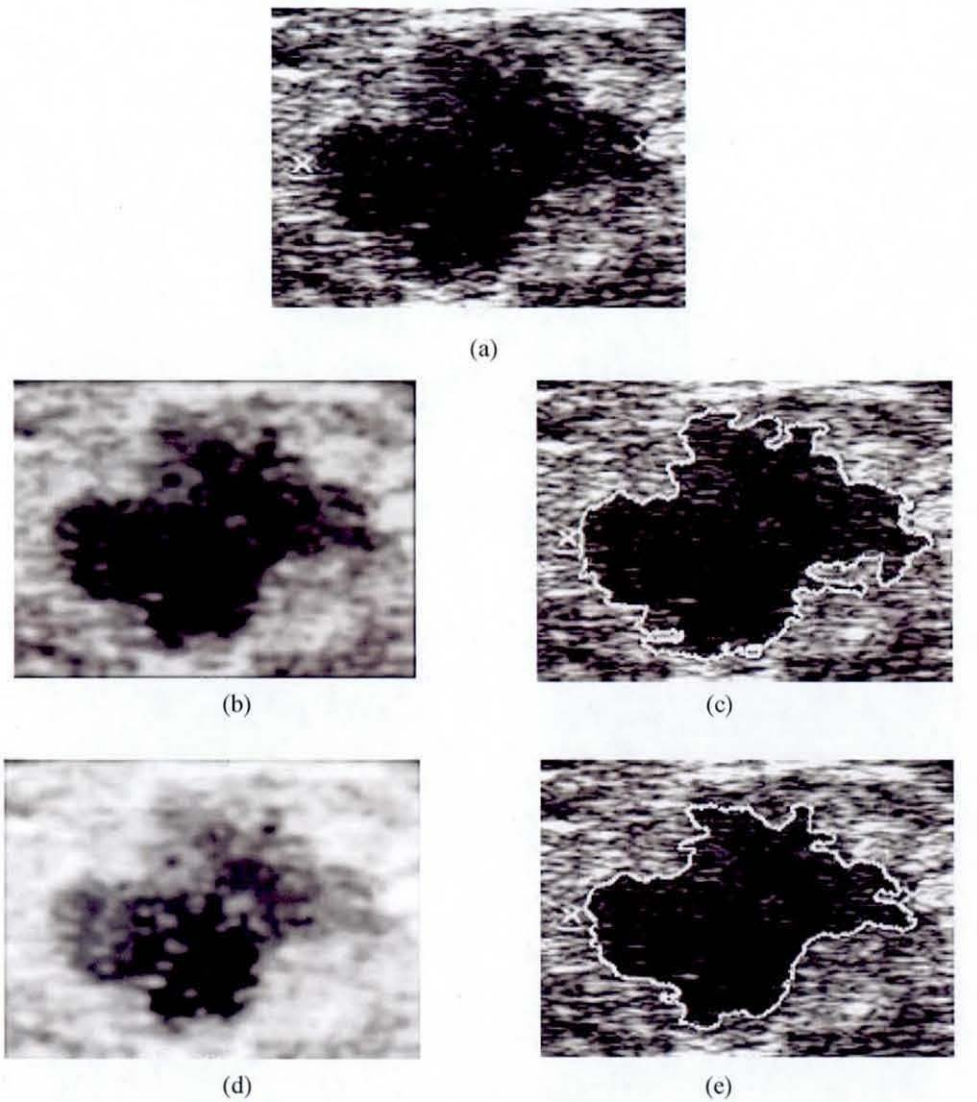


FIG. 7. Justification of boundary enhancement with multifractal processing. (a) Original image; (b) hybrid filtered image; (c) boundary detected on hybrid filtered image; (d) multifractal processed image; (e) boundary detected on multifractal processed image.

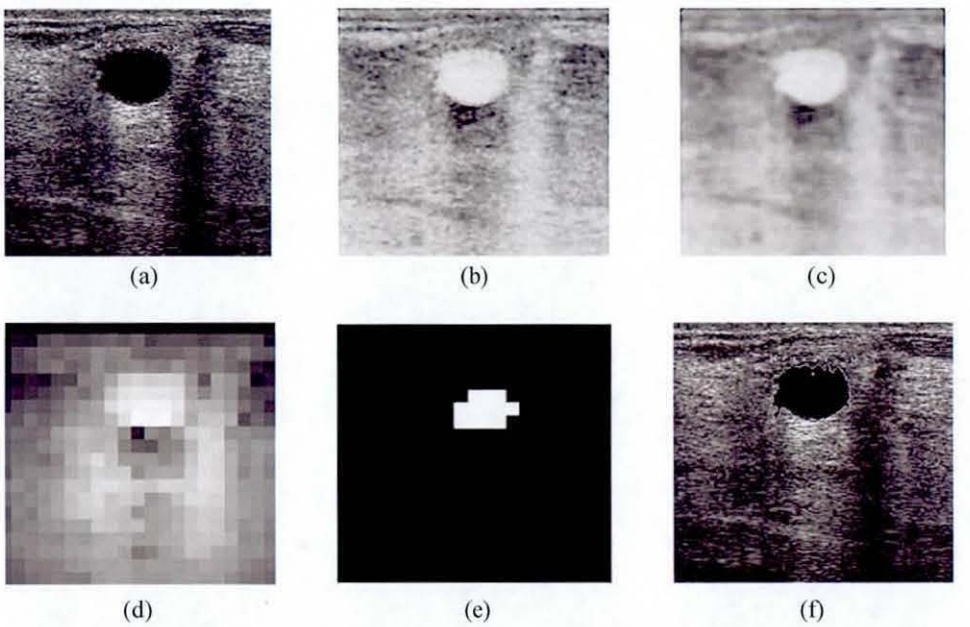


FIG. 8. Illustration of the operation of the intermediate stages of the Drukker et al.[6] algorithm. (a) Original image; (b) grayscale inverted image; (c) median filtered image; (d) radial gradient index (RGI) filtered image; (e) thresholded RGI image; (f) final detection.

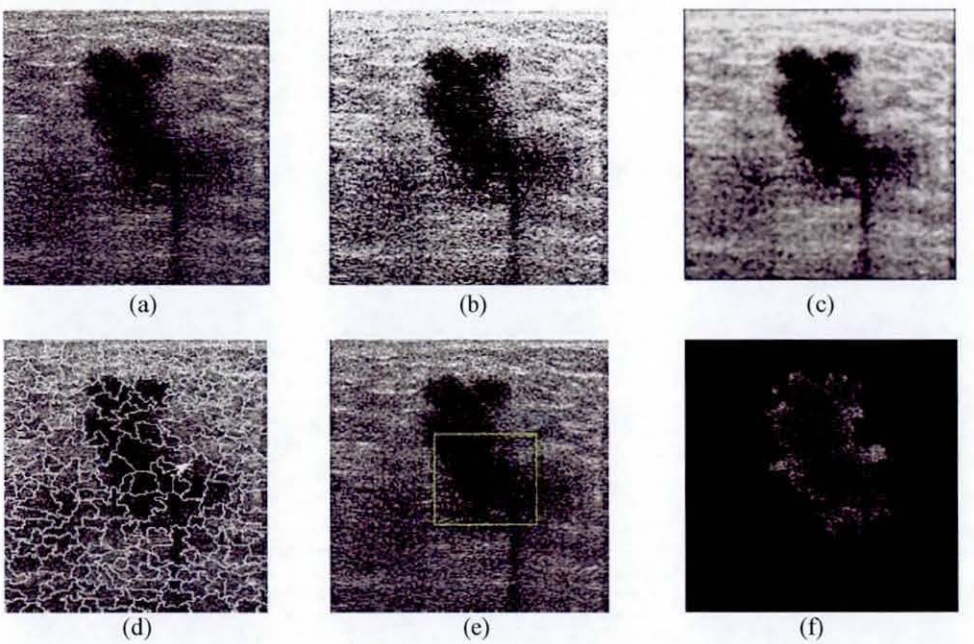


FIG. 9. Illustration of the operation of the intermediate stages of the Yap et al.[7] algorithm (local mean). (a) Original image; (b) image after preprocessing (histogram equalization); (c) image after hybrid filtering; (d) watershed segmentation mapped onto the original image; (e) location of the initial lesion; (f) combination of the neighborhood segments with the initial lesion.

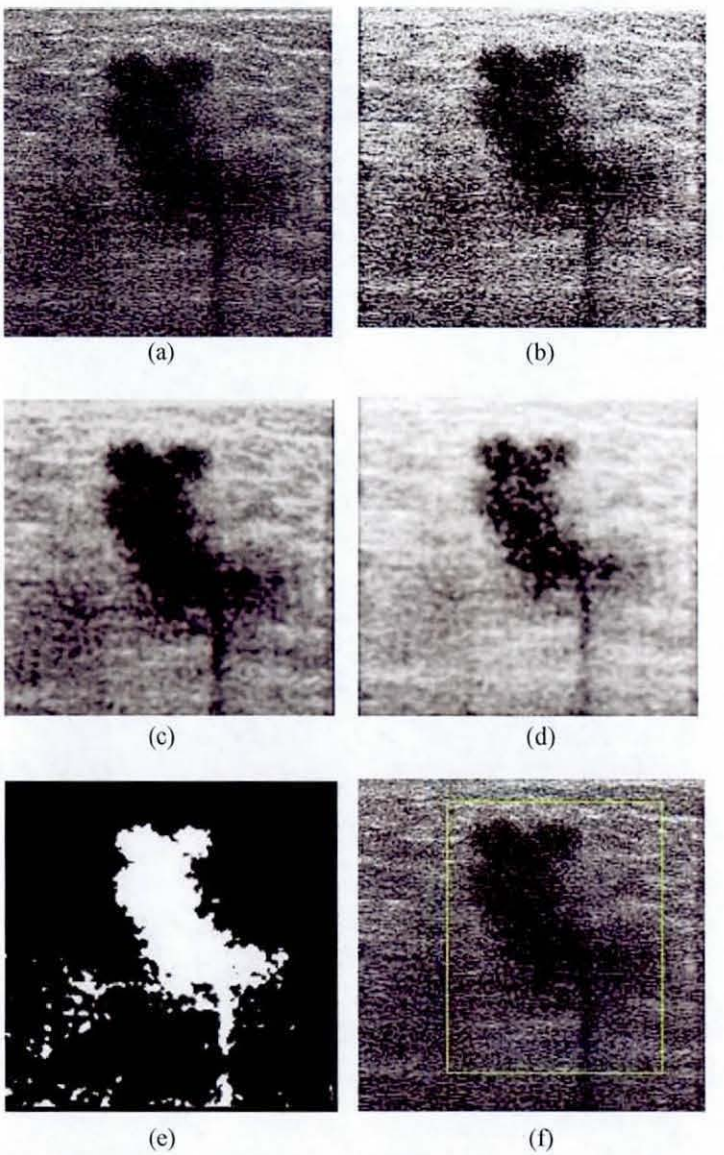


FIG. 10. Illustration of the operation of the intermediate stages of the proposed algorithm. (a) Original image; (b) image after pre-processing (histogram equalization); (c) image after hybrid filtering; (d) image after multifractal processing; (e) image after thresholding segmentation; (f) labeling of region of interest.

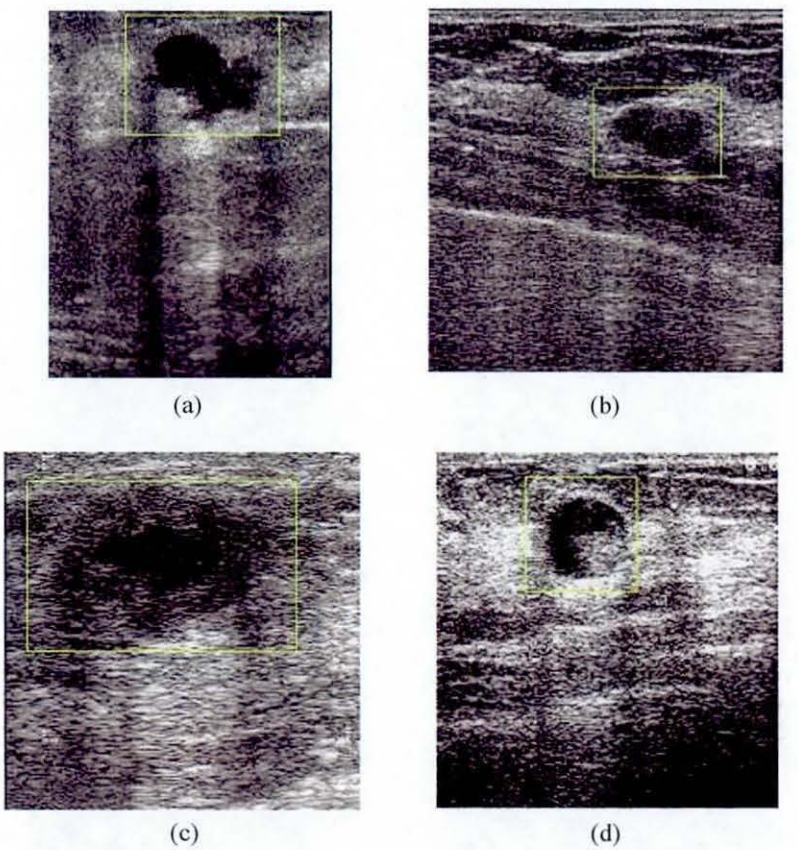


FIG. 11. Results of automated region of interest lesion labeling using the proposed method. (a) Malignant tumour; (b) simple cyst; (c) fibroadenoma; (d) complex cyst.

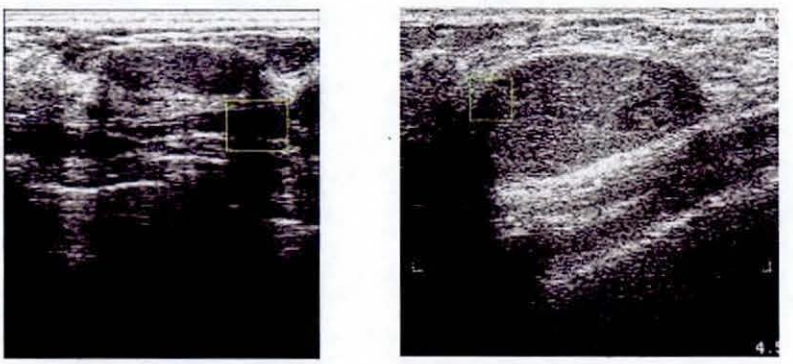


FIG. 12. Examples of unsuccessful lesion identification for two cases of fibroadenoma.

TABLE 1. Summary of region of interest detection accuracy for each type of abnormality under four different algorithms

Diagnosis	*Total images (n)*	*Accuracy (%)*			
		Drukker et al.[6]	*Yap et al. (local mean)*[7]	*Yap et al. (fractal)*[7]	*Proposed method*
Malignant	20	80.00	65.00	45.00	90.00
Simple cysts	76	67.11	60.53	55.26	86.84
Complex cysts	76	72.37	71.05	59.21	89.47
Fibroadenoma	58	62.07	63.79	46.55	77.59
Carcinoma	38	57.89	78.95	52.63	78.95
Occult lesions	18	94.44	88.89	66.67	88.89
Adenosis	15	80.00	73.33	46.67	93.33
Others	59	59.32	64.41	54.24	89.83
Total	360	67.78	68.05	53.89	86.11

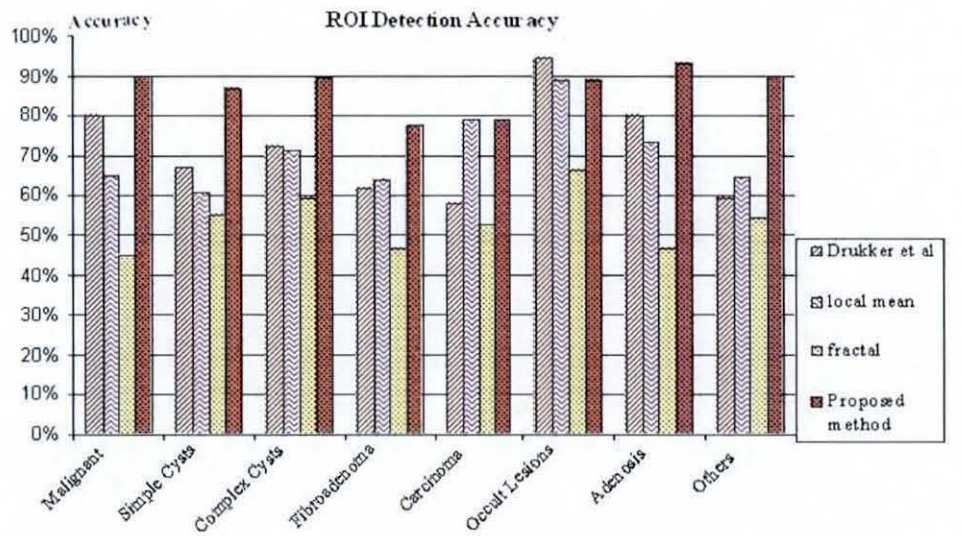


FIG. 13. Graphical presentation of the region of interest (ROI) detection accuracy for each of the four algorithms and each type of abnormality.

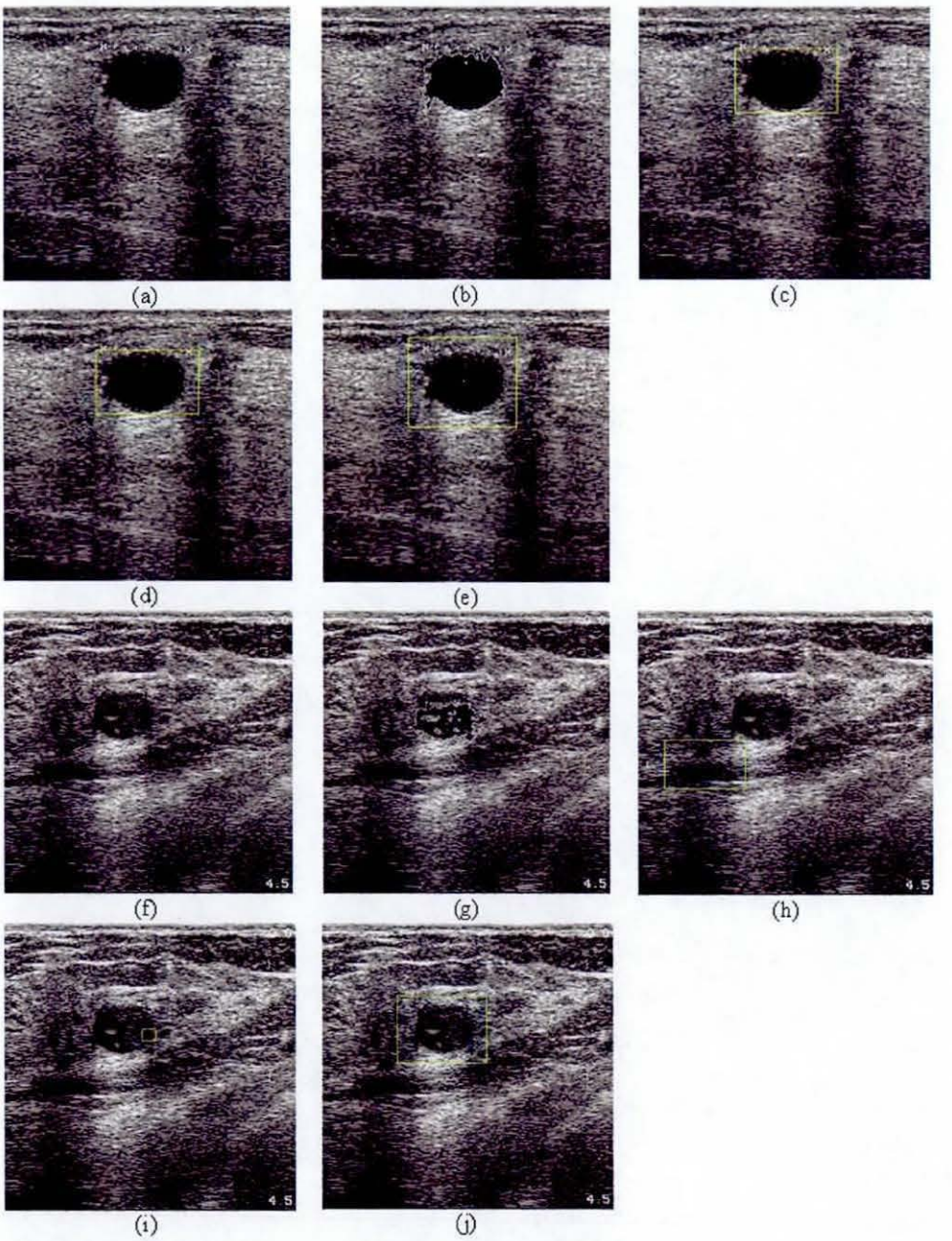


FIG. 14. Visual performance comparison with the benchmark algorithms. (a,f) Original images; (b,g) results of the Drukker et al.[6] algorithm; (c,h) results of the Yap et al.[7] algorithm (local mean); (d,i) results of the Yap et al.[7] algorithm (fractal dimension); (e,j) results of the proposed method.

V. SUMMARY AND CONCLUSIONS

We propose a method that is able to fully automate ultrasound CAD. We have successfully proved that the proposed algorithm achieves an improvement as compared with benchmark algorithms. The proposed method is able to very accurately label most lesions, with its best performance being the identification of malignant lesions (90%) and its worst being the identification of fibroadenomas (77.59%). We are currently considering the use of shape information and frequency domain analysis, among other techniques, to further improve the performance of the approach presented here.

A complete computer-aided ultrasound diagnostic system (ultrasound CAD) must be able to classify the diagnosis of each ROI. In future, we will investigate the use of classification techniques such as neural networks and support vector machines, among others, to form a fully automated breast cancer detection system.

REFERENCES

1. Lawson A, Lawson J. Breast cancer—can you prevent it? Sydney (Australia): McGraw–Hill; 1999. 224 p.
2. BBC News. Breast cancer. BBC News Channel [Electronic version]. 30 January 2004. Retrieved 6 August 2008 from http://news.bbc.co.uk/go/pr/fr/-/1/hi/health/medical_notes/3244315.stm.
3. Freedman DA, Petiti DB, Robin JM. On the efficacy of screening for breast cancer. Int J Epidemiol. 2004;33(1):43–55.
4. Chen DR, Chang RF, Wu WJ, Moon WK, Wu WL. 3-D Breast ultrasound segmentation using active contour model. Ultrasound Med Biol. 2003;29(7):1017–1026.
5. Chen D, Chang RF, Huang YL. Breast cancer diagnosis using self-organizing map for sonography. Ultrasound Med Biol. 2000;26(3):405–411.
6. Drukker K, Giger ML, Horsch K, Kupinski MA, Vyborny CJ, Mendelson EB. Computerized lesion detection on breast ultrasound. Med Phys. 2002;29(7):1438–1446.
7. Yap MH, Ewe HT. Region of interest (ROI) detection in ultrasound breast images. TS11-2. In: Proceedings of MMU International Symposium on Information and Communications Technologies (M2USIC). Cyberjaya, Malaysia: Multimedia University; 2005: 5–8.
8. Boone JM. Neural networks at the crossroads. Radiology. 1993;189(2):357–359.
9. Gurney JW. Neural networks at the crossroads: caution ahead. Radiology. 1994;193(1):27–28.
10. Boukerroui D, Basset O, Guérin N, Baskurt A. Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation. Eur J Ultrasound. 1998;8(2):135–144.
11. Sehgal CM, Cary TW, Kangas SA. Computer-based margin analysis of breast sonography for differentiating malignant and benign masses. J Ultrasound Med. 2004;23(9):1201–1209.
12. Wintz P, Gonzalez RC. Digital image processing. 2nd ed. Boston (MA): Addison–Wesley; 1987: 146–152.
13. Gonzalez RC, Woods RE, Eddins SL. Digital image processing using MATLAB. Upper Saddle River (NJ): Prentice-Hall; 2003. 609 p.
14. SourceForge.net. LTI-LIB. Introduction [Electronic resource]. 24 November 2005. Retrieved 6 August 2008 from http://Ltilib.sourceforge.net/doc/homepage/index.shtml.
15. Friedland N, Adam D. Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated annealing. IEEE Trans Med Imaging. 1989;8(4):344–353.
16. Prapavesis ST, Fornage BD, Weismann CF, et al. Breast ultrasound and US-guided interventional techniques [Multimedia teaching file]. Thessaloniki (Greece): authors; 2001.
17. Yap MH, Edirisinghe EA, Bez HE. Object boundary detection in ultrasound images. In: Ley M. Proceedings of Third Canadian Robotics and Machine Vision Conference (CRV 2006). Quebec (QC): IEEE Computer Society; 2006: 53. Abstract retrieved 7 November 2006 from http://doi.ieeecomputersociety.org/10.1109/CRV.2006.51.
18. Joo S, Yang YS, Moon WK, Kim HC. Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features. IEEE Trans Med Imaging. 2004;23(10):1292–1300.
19. Kupinski M, Giger ML, Lu P, Huo ZM. Computerized detection of mammographic lesions: performance of artificial neural network with enhanced feature extraction. Proc SPIE. 1995;2434:598–605.
20. Mrazek P. Nonlinear diffusion for image filtering and monotonicity enhancement [PhD thesis]. Prague (Czech Republic): Czech Technical University; 2001. 110 p. Document retrieved 6 August 2008 from ftp://cmp.felk.cvut.cz/pub/cmp/articles/mrazek/Mrazek-phd01.pdf.
21. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell. 1990;12(7):629–639.
22. Yu Y, Acton ST. Speckle reducing anisotropic diffusion. IEEE Trans Image Process. 2002;11(11):1260–1270.
23. Weickert J. Nonlinear diffusion filtering. In: Jähne B, Haussecker H, Geissler P, editors. Handbook on computer vision and applications. Vol. 2. San Diego (CA): Academic Press; 1999: 423–450.

24. Yeo TS, Du G. A multifractal approach for auto-segmentation of SAR images. In: Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium; Sydney, Australia; 9 – 13 July 2001. Los Alamitos (CA): IEEE; 2001: 5:2301–2303.
25. Mandelbrot BB. The fractal geometry of nature. Rev. ed. New York (NY): W.H. Freeman and Company; 1982. 468 p.
26. Sahoo PK, Soltani S, Wong AKC, Chen YC. A survey of thresholding techniques. Comput Vis Graph Image Process. 1988;41(2):233–260.

European Journal of Radiology xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

European Journal of Radiology

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

Processed images in human perception: A case study in ultrasound breast imaging

Moi Hoon Yap[a,*], Eran Edirisinghe[b], Alastair Gale[c], Helmut Bez[d]

[a] Department of Computer Science, Loughborough University, FH09, Ergonomics & Safety Research Institute, Holywell Park, UK
[b] Department of Computer Science, Loughborough University, FJ.05, Garendon Wing, Holywell Park, Loughborough LE11 3TU, UK
[c] Department of Computer Science, Loughborough University, FJ, Garendon Wing, Holywell Park, Loughborough LE11 3TU, UK
[d] Department of Computer Science, Loughborough University, Room N.2.26, Haslegrave Building, Loughborough University, Loughborough LE11 3TU, UK

ARTICLE INFO

Article history:
Received 18 August 2008
Accepted 10 November 2008

Keywords:
Ultrasound
Breast cancer
Human perception
Receiver Operating Curve
Computer-Aided Diagnosis

ABSTRACT

Two main research efforts in early detection of breast cancer include the development of software tools to assist radiologists in identifying abnormalities and the development of training tools to enhance their skills. Medical image analysis systems, widely known as Computer-Aided Diagnosis (CADx) systems, play an important role in this respect. Often it is important to determine whether there is a benefit in including computer-processed images in the development of such software tools. In this paper, we investigate the effects of computer-processed images in improving human performance in ultrasound breast cancer detection (a perceptual task) and classification (a cognitive task). A survey was conducted on a group of expert radiologists and a group of non-radiologists. In our experiments, random test images from a large database of ultrasound images were presented to subjects. In order to gather appropriate formal feedback, questionnaires were prepared to comment on random selections of original images only, and on image pairs consisting of original images displayed alongside computer-processed images. We critically compare and contrast the performance of the two groups according to perceptual and cognitive tasks. From a Receiver Operating Curve (ROC) analysis, we conclude that the provision of computer-processed images alongside the original ultrasound images, significantly improve the perceptual tasks of non-radiologists but only marginal improvements are shown in the perceptual and cognitive tasks of the group of expert radiologists.

© 2008 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Breast cancer is the leading cause of death of women in developed countries [1]. It is now the most common cancer in the UK [2]. According to recent statistics [2], more than 44,000 women are diagnosed with breast cancer in the UK each year and worldwide, more than a million women are diagnosed with breast cancer every year. These alarming statistics have motivated research towards developing Computer-Aided Detection (CAD) tools that can effectively be used in the early detection of breast cancer, which is the key to reducing mortality.

Medical image analysis is an important element in the development of CAD and Computer-Aided Diagnosis (CADx) systems. Over the past two decades, a number of CAD tools have been developed to aid radiologists in detecting likely cases of breast cancer. Further CAD tools have been developed for the automated differentiation between benign and malignant lesions. In some tools the technology provides functionality beyond straightforward computer-aided detection by providing a measure of the cancer likelihood for a detected lesion, given the image and/or patient characteristics [3].

However, it is widely accepted that CAD is not intended to completely replace the role of a radiologist. Though computer-processed images have the potential to improve the functionality of a CAD system, a medical expert or a specially trained radiologist will take the final decision in the detection and interpretation of any lesion. However, in the recent past there has been a serious worldwide shortage of consultant radiologists which has resulted in a significant growth in the use of specially trained radiographers and other non-radiologists [4] in the above decision-making process. This trend is expected to continue in the future. Therefore, the development of reliable CAD systems to aid the decision-making process of such individuals, is timely.

In the UK, currently there are some 109 cancer screening centres and a growing number of individuals (circa 650) nationally performing approximately 1.5 million examinations per year [5]. As a result there has increasing interest in Intelligent Tutoring Systems (ITS) in the area of medical training [6,7]. In general the rationale of ITS is based on the assumption that the learner's cognitive processes can be modelled, traced, and corrected in the context of problem-solving [7]. To this effect, Crowley and Medvedeva [8] have

* Corresponding author. Tel.: +44 1509 635739; fax: +44 1509 635709.
E-mail addresses: M.H.Yap@lboro.ac.uk, moihoon@gmail.com (M.H. Yap).

0720-048X/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ejrad.2008.11.007

vorked on the adaptation of ITS and knowledge-based systems KBS) for creating intelligent educational systems. However, though TS have provided a basic pedagogical approach with proven effiacy in domains outside the field of medicine, it was not designed to e used in large, frequently changing, or existing knowledge bases 8] within the area of medical diagnosis. The aim of the proposed esearch is to investigate the effects of using computer-processed mages alongside non-processed images traditionally used in ITS, articularly within knowledge bases with wide variations. In paricular we analyse the effects of computer-processed images in he judgement of individuals with backgrounds in radiology, comuter science, engineering, mathematics and arts. The results of his study provide valuable input and justifications to the use of AD/ITS in medical diagnosis, in particular screening of breast ancer.

For clarity of presentation, the paper is divided into five sections. Section 2.1 details the background and the motivation behind he proposed research. Section 2.2 provides an explanation of the xperimental methodology adopted, including information on test mages and a summary of the adopted computer processing techique. The experimental results are presented and discussed in letail in Section 3. Finally Section 4 concludes with an insight into uture directions of research.

. Materials and methods

.1. Background and motivation

Mammography is well accepted as the golden standard in breast ancer detection. It is a special type of X-ray imaging that can be sed to create detailed images of a breast [9]. However, mammogaphy alone cannot for certain differentiate between benign and nalignant tumours [10]. Hence, there is a need to look into differnt modalities to increase the sensitivity and specificity of breast ancer detection. Sonography, i.e. ultrasound images, is an imporant modality in the evaluation and treatment of breast masses. In general its use is often tailored to the case of a particular patient in nswering a specific question/doubt that may have been raised in a re-examination involving mammography or physical examination 11].

For ultrasound images, based on the perspective of human visual ystem, the abnormalities are hardly detected if they are above sycho-visual thresholds. If below threshold, perceptual skills of attern recognition or matching are needed to detect the abnornality. Once the suspicious region is detected, the cognitive skill of lecision-making (classification) is needed to group the feature to nalignant or benign categories [4].

To become an expert general radiologist, one needs to equip with adiological expertise, thorough acquiring knowledge of relevant natomy, physiology, pathology, physiopathology, projective geomtry of radiography, and the essentials of medicine and surgery [7]. By providing additional processed images, we aim to assist radiologists in detection and interpretation in ultrasound examinations. We believe that the additional information will boost the confidence level of the human psycho-visual system in decision-making. Within the context of our present research we are particularly interested in the effects of providing such additional information to nonexpert radiologists and non-radiologists. In the following section, ve provide details of the subjective testing methodology adopted by us within our present research context to prove the contribution f processed images in ultrasound medical training systems.

.2. Methodology

Two experiments were conducted to examine the effect of processed images in

- the detection (perceptual tasks) of abnormalities by a group of non-radiologists and
- the detection and interpretation (perceptual and cognitive tasks) of abnormalities by a group of expert radiologists.

2.2.1. Experimental design and rationale

The experiments were designed to subjectively measure the level of human perception (and cognition for the group of radiologists) in ultrasound breast imaging. Initially, the experiments were performed on the group of non-radiologists (perceptual tasks) and were subsequently extended to cover a larger sample of people, with diverse levels of exposure to ultrasound imaging, particularly to those groups who have had experience as radiographers and radiologists (perceptual and cognitive tasks). In particular, we are interested in further analysing the possibility of performance improvement of the latter groups of professionals by the effective use of computer-processed images.

The survey involved the following steps:

- Randomly select five original ultrasound images. Then select a further five images with similar diagnostic difficulty and obtain their corresponding computer-processed (using Yap et al.'s work of [12], see Section 3.3) images as sample test images.
- Design questionnaires to impartially investigate subjective judgement.
- Conduct the survey on two distinct subject groups: expert radiologists and non-radiologists.
- Analyse the results using statistical and ROC-based methods.
- Draw suitable conclusions.

Each of the steps is described in detail in the following subsections.

2.2.2. Test image dataset

The test images used in our work are obtained from a professionally compiled Breast Ultrasound CD [13], in which each image is accompanied by an accurate explanation of its diagnosis from an expert Radiologist. A total of 51 clinical ultrasound cases and more than 200 instructional videos are included in the CD ROM. It is particularly noted that the CD only contained images/videos of abnormal breast ultrasound images. In order not to overload the subject [14], the length of the questionnaires was limited to 10 questions. The subjects were required to study and comment on five randomly selected original ultrasound breast images of different diagnostic, and five further specially selected original ultrasound breast images an their corresponding processed images. All subjects were presented with the same set of randomly selected images. However, to avoid subjectivity due to performance variations as a result of to human fatigue, the five images were presented to each subject in a random order within the questionnaires.

In order to maintain the consistency of the survey, original ultrasound breast image pairs with similar level of diagnostic difficulty were chosen from the test ultrasound image database. These images are illustrated in Fig. 1.

2.2.3. Processed images

In medical image analysis, a number of algorithms are available to pre-process ultrasound images to improve chances of subsequent correct diagnosis either by manual, semi-automatic or automatic means. These algorithms use different filtering methods to produce an enhanced image for object segmentation. The degree of enhancement (or the level of visual aid) offered by alternative image processing algorithms may vary. However, human vision has a lot of constraints. Hence, not all the processing/filtering algorithms are suitable for human vision. Amongst the state-of-the-art preprocessing algorithms, Isogauss processing [12] has been proven

Please cite this article in press as: Yap MH, et al. Processed images in human perception: A case study in ultrasound breast imaging. Eur J Radiol (2009), doi:10.1016/j.ejrad.2008.11.007

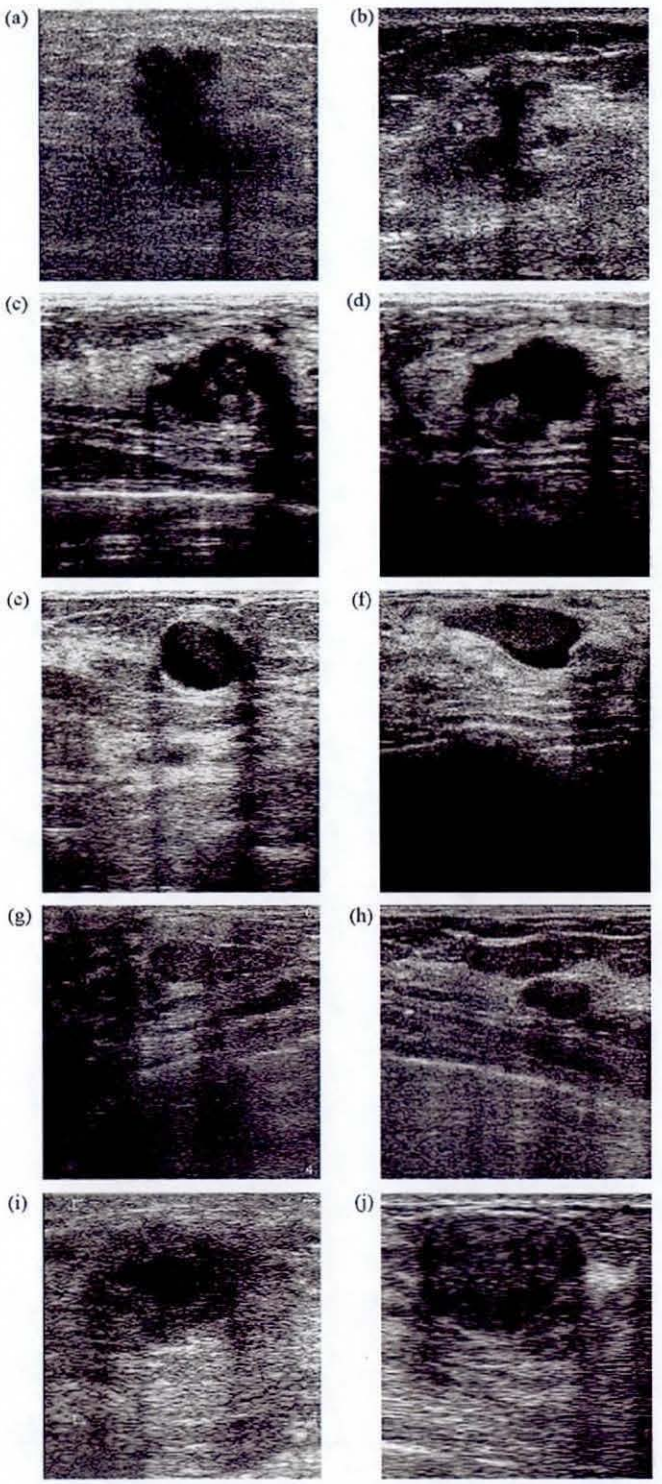


Fig. 1. (a and b), (c and d), (e and f), (g and h), (i and j) are five pairs of original ultrasound images selected for the survey. (a, c, e, g and i) are selected randomly, while (b, d, f, h and j) are purposely selected to be of same level of diagnostic difficulty.

to perform the best (see **Fig. 2** for a comparison of performance). Essentially, Isogauss processing involves the multiplication of the inverse of the image by an Isotropic Gaussian Function (with a fixed variance as the constraint function), centred at a seed point location. Readers interested in more technical details are referred to **[12,15]**.

2.2.4. Subjects

The experiments were conducted on two groups of subjects, non-radiologists and radiologists. The non-radiologist group was represented by, 40 subjects, from different academic backgrounds, i.e. in computer science, engineering, mathematics and arts. One major difficulty in setting up the experiments was the search for participants to represent the group of expert radiologists. For this reason only 10 expert radiologists participated in the survey. The convincing, conclusive evidence gathered from the subjective tests confirm this number of subjects as sufficient.

2.2.5. Design of the questionnaire

Two sets of questionnaires were designed, one for analysing and commenting on a random set of original images and the second for similar on the set of original-processed image pairs. Due to the differences between the two sets of subject groups (e.g. familiarity with medical terms, abilities, etc.), the questionnaires had to be appropriately re-worded for each group. Readers interested in the details of the questionnaires are referred to **[16]**. In the above process, careful consideration has been given to the re-wording to maintain impartiality of responses from the group of subjects. We use the popular Absolute Category Rating (ACR) **[17]** method to measure the level of perception of subjects. ACR involves a five-scale evaluation, in which the subjects are asked to observe the images and determine whether an identifiable object is present in terms

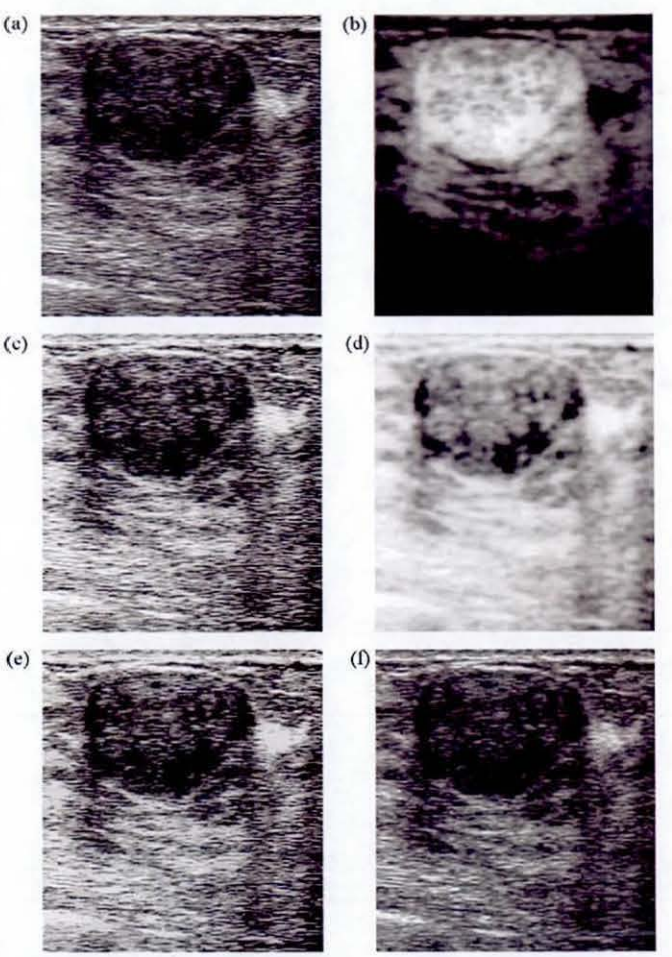


Fig. 2. (a) Original image (b–f). Processed images by using different algorithms: (b) isogauss processing, (c) histogram equalization, (d) multifractal filtering (e) nonlinear filtering and (f) median filtering.

Please cite this article in press as: Yap MH, et al. Processed images in human perception: A case study in ultrasound breast imaging. Eur J Radiol (2009), doi:10.1016/j.ejrad.2008.11.007

Table 1
Detection accuracies of the group of non-radiologists.

Sub-group/background	Images presented to the subject		
	Original images	Processed images	Both images
Computer science	0.64	0.66	0.76
Mathematics	0.5556	0.4222	0.8667
Engineering	0.56	0.52	0.82
Arts	0.44	0.54	0.76

of, certainly not exist, probably not exist, not sure, probably exist and certainly exist.

In the case of the volunteer expert radiologists, the questions were directed in medical terms. First they were asked to decide the suspected region. If either probably exist or certainly exist was selected, the subjects were asked to draw the boundary of the suspected region on the images, and provide the diagnosis of the suspected region with five-scale evaluation: benign, probably benign, unknown, probably malignant and malignant. In the case of non-radiologist, some appropriately guided questions, without medical terms, were presented in the questionnaire. First they were asked to decide the existence of objects in the images. If an object was detected the subjects were further asked to provide the location/size of the suspected region in terms of (x, y) coordinate values and to determine how the width compares with the height. Subjects were further asked to elaborate on the shape of the object by classifying it to be of regular, smooth or unknown shape. The detection was considered to be successful if the subjects managed to accurately detect the presence of an object, while the interpretation was considered to be successful if the subjects managed to provide accurate width, height, and boundary information.

Further subjective experimentation considerations such as, quality of the images presented, their ordering, human fatigue, lighting conditions, time-of-day, age group, etc., were also carefully considered.

2.2.6. *Receiver Operating Characteristics (ROC) analysis*

As a means of illustrating the experimental results, two ROC curves are plotted for each group of subjects with the help of the software package ROCKIT [18]. ROC analysis is a popular tool used to assess, define and compare classification results in medical diagnostics. In medical research, decision threshold represents how often on is right or wrong in a diagnostic test. ROC analysis shows the trade off between the sensitivity and specificity as a decision threshold is varied and uses maximum likelihood estimation to fit a binormial ROC curve [19] to the test data. As the non-radiologists group will not be sufficiently experienced to decide whether an abnormality represents a malignant or non-malignant lesion, this decision is deduced by the authors using the answers given by the group members to the additional questions asked in the relevant questionnaire (see Section 3).

Section 3 presents the experimental results and a detailed analysis, leading to conclusions that are presented in Section 4.

3. Results and discussion

Table 1 illustrates the detection accuracies of the perceptual tasks of the group of non-radiologists. The experimental results are classified according to the subjects' professional background. The detection accuracies are presented in a scale of 0–1, where 1 refers to 100% detection.

Fig. 3 provides a graphical presentation of the accuracy results obtained for the different sub-groups of subjects (of the group of non-radiologists). It is noted that subjects having a computer science background, most of whom having expertise in computer vision and image analysis, managed to obtain the highest accuracy rates when individual images, i.e. the original and processed images, were separately shown. When only the original images were shown, subjects with an arts background performed worse and when only the processed images were shown subjects with a mathematical background performed worse. When further queried it was revealed that these results were mostly due to the amount of previous exposure each group has had with ultrasound images and forms of computer-processed images. Therefore, it can be concluded that the experience one has had with images play an important role in their ability to use them in perceptual tasks.

The most important observation from Fig. 3 is the ability of all groups of subjects to improve their detection accuracy (perceptual task) significantly when both images are provided simultaneously, as compared to the accuracy shown by them separately. For example the overall average detection accuracy when using original images is 54.89%, which increases to 80.17% when both images were provided simultaneously for inspection. Interestingly the best percentage improvement has also been illustrated by the group of people, mathematicians, who had the least performance accuracy when the images were shown separately.

Fig. 4 illustrates the results obtained from the ROC tool for the group of non-radiologists. It shows that providing the original and processed images side-by-side allows the non-radiologists to per-

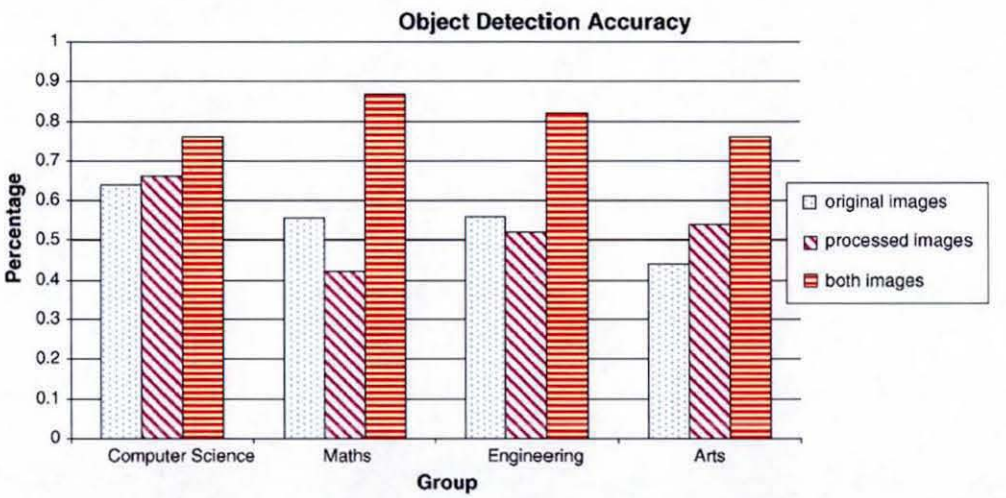


Fig. 3. Comparison of the detection accuracy between subjects from different professional backgrounds.

Please cite this article in press as: Yap MH, et al. Processed images in human perception: A case study in ultrasound breast imaging. Eur J Radiol (2009), doi:10.1016/j.ejrad.2008.11.007

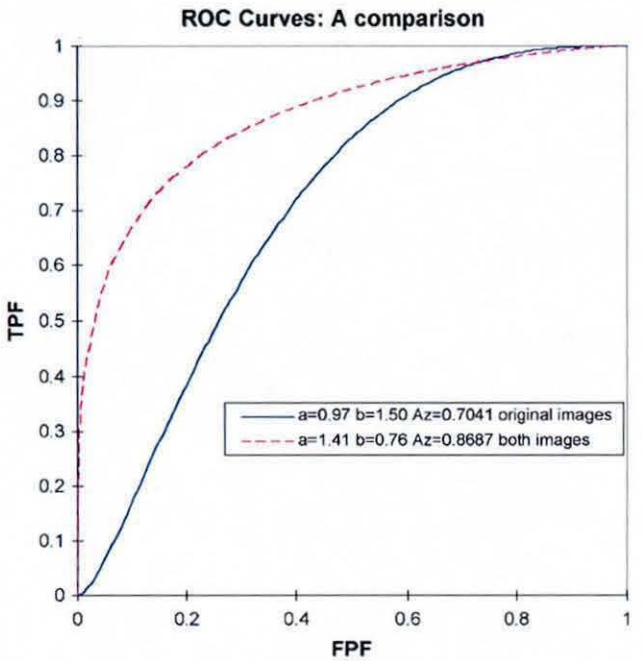


Fig. 4. Comparison of ROC curves (non-radiologists).

form substantially better, compared to providing them with only the original images. A lack of experience in identifying abnormalities in ultrasound images results in either: (i) wrong judgements or (ii) more uncertainties in the decision-making process, when only original images are examined by this group of people. The processed images provide a clearer segmentation of abnormal regions which increases the accuracy of judgement substantially. Summarising the above observations, it can be concluded that computer-processed images, when provided as an aid, alongside original images, increases the perceptual capability of most non-radiologists.

Fig. 5 illustrates that providing the original and processed images, side-by-side, only allows the group of expert radiologists

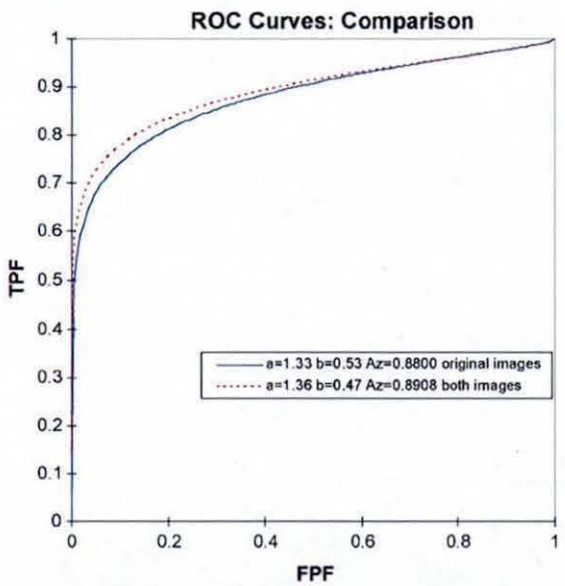


Fig. 5. Comparison of ROC curves (radiologists).

to perform marginally better. It is noted that subjects in this group, who have expertise in reading ultrasound images in their original format, managed to obtain the high sensitivity and specificity when only the original images were presented.

Given the above observations, it can be concluded that radiological experience with ultrasound images plays an important role in accurate diagnostics, which involved the cognitive process. It can be concluded that the provision of processed images significantly aids the perceptual performance of non-radiologists and marginally improves the performance of experienced radiologists. Given the fact that a trainee radiologists at the start of training, may not be significantly different experience wise as compared to a subject from our group of non-radiologists, it can be concluded that computer-processed images may be a useful way to boost their confidence in diagnostics and improve speed of training. Further we believe that despite the marginal improvement in performance accuracy shown by the expert radiologists group, computer-processed images can help in decision-making, particularly when diagnosing images with noise and artefacts.

4. Conclusions

We have carried out subjective experiments to analyse the improvement of human performance, in perceptual and cognitive aspects, that can be obtained by providing computer-processed images alongside original images when using ultrasound images in breast cancer screening/detection. We have shown that the sensitivity and specificity are improved when both images are made available for judgement. It can be concluded that computer-processed images, when provided as an aid, alongside original images, increases the perceptual task capability of most individuals, particularly those who are inexperienced radiologists or trainees. These observations clearly prove the effectiveness of using imaging technologies/algorithms in CAD systems. Therefore, in an era where there is a shortage of experienced radiologists, CAD systems that provide computer-processed images as an aid for diagnostics are highly beneficial. We suggested that processed images could be adapted in the radiologist's training system.

At present we are investigating the relationship between level of training, training methods, age/background of trainees and the improvement of diagnostic accuracy that can be provided by CAD systems. We are also investigating the effects of different state-of-the-art segmentation algorithms in subjective performance.

Acknowledgement

This research is supported by Department of Computer Science, Loughborough University.

References

[1] Chen DR, Chang RF, Huang YL. Breast cancer diagnosis using self-organizing map for sonography. Ultrasound Med Biol 2000;26(3):405–11.
[2] Cancer Research UK–Breast Cancer. **http://info.cancerresearchuk.org/cancerstats/types/breast/** (accessed on January 30, 2008).
[3] Kallergi M. Computer-aided diagnosis of mammographic microcalcification clusters. Med Phys 2004;31:314–26.
[4] Gale AG. PERFORMS—a self-asssessment scheme for radiologists in breast screening. Seminars in Breast Disease: Improving and Monitoring Mammographic Interpretative Skills 2003;6:148–52.
[5] National Health Service Breast Screening Programme. **http://www.cancerscreening.nhs.uk/breastscreen/index.html** (accessed on January 15, 2008).
[6] Crowley RS, Medvedeva O. A general architecture for intelligent tutoring of diagnostic classification problem solving. AMIA Annu Symp Proc 2003:185–9.
[7] Azevedo R, Lajoie SP. The cognitive basis for the design of a mammography interpretation tutor. Int J Artif Intell Educ 1998;9:32–44.
[8] Crowley RS, Medvedeva O. An intelligent tutoring system for visual classification problem solving. Artif Intell Med 2006;36:85–117.

Please cite this article in press as: Yap MH, et al. Processed images in human perception: A case study in ultrasound breast imaging. Eur J Radiol (2009), doi:10.1016/j.ejrad.2008.11.007

[9] Imaginis—General Information on Mammography. http://www.imaginis.com/breasthealth/mammography.asp (accessed on December 15, 2007).
[10] Cowley HC. Assessment and training in breast cancer detection, PhD thesis, AVRC, ESRI, Loughborough University; 1999.
[11] Venta LA, Dudiak CM, Salomon CG, Flisak ME. Sonographic evaluation of the breast. RadioGraphics 1994;14:29–50.
[12] Yap MH, Edirisinghe EA, Bez HE. Fully automatic lesion boundary detection in ultrasound breast images, 651265123I. San Diego: SPIE Medical Imaging; 2007.
[13] Prapavesis ST, Fornage BD, Weismann CF, Palko A, Zoumpoulis P. Breast ultrasound and US-guided interventional techniques—a multimedia teaching file, 1st ed., Thessaloniki Greece; 2001.
[14] Market Research Society (MRS). Questionnaire design guidelines, May 2006.
[15] Yap MH, Edirisinghe EA, Bez HE. Object boundary detection in ultrasound breast images, Third Canadian Robotics and Machine Vision Conference (CRV 2006).
[16] Sample Questionnaires. http://www-staff.lboro.ac.uk/~ehmhy/samples.html (accessed on January 31, 2008).
[17] Moller S. Assessment and prediction of speech quality in telecommunication, chapter 3. Springer; 2000. ISBN 0792378946.
[18] Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC curve estimates obtained from partially-paired datasets. Med Decis Making 1998;18:110–21.
[19] Carsten S, Sebastian W, Tania S, Klaus J. Comparison of eight computer programs for receiver-operating characteristic analysis. Clin Chem 2003;49(3):433–9.

Please cite this article in press as: Yap MH, et al. Processed images in human perception: A case study in ultrasound breast imaging. Eur J Radiol (2009), doi:10.1016/j.ejrad.2008.11.007