

A Novel Approach for Classification and Early Detection of Breast Cancer using Image Enhancement, Image Segmentation and Feature Extraction on Mammograms

THESIS

Submitted in partial fulfilment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by

PARAMKUSHAM SPANDANA

IDNO: 2011PHXF0013H

Under the Supervision of
Prof. B.V.V.S.N.Prabhakar Rao

&

Under the Co-supervision of
Dr.K.M.M.Rao



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2018

CERTIFICATE

This is to certify that the thesis entitled A Novel Approach for Classification and Early Detection of Breast Cancer using Image Enhancement, Image Segmentation and Feature Extraction on Mammograms and submitted by Paramkusham Spandana ID No 2011PHXF0013H for award of Ph.D. Degree of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor

Prof. B.V.V.S.N.PRABHAKAR RAO,
Associate Professor,
BITS-Pilani, Hyderabad

Date:

Signature of the Co-supervisor

Dr.KM.M.Rao,
Deputy Director (Rtd), NRSC, ISRO
Former Adjunct Faculty,
BITS-Pilani, Hyderabad

Date:

Acknowledgements

First I would like to thank almighty for giving me strength throughout my research journey and thank everyone who helped me to accomplish my research work.

I would like to express my sincere gratitude to my supervisor Prof. B.V.V.S.N Prabhakar Rao. I am extremely grateful to my co-supervisor Dr. K.M.M. Rao, former Deputy Director (DPA), NRSC, ISRO without him this work would not have been completed. I am indebted to both supervisor and co-supervisor for their intense guidance and encouragement. They were always available for me to give suggestions and constructive feedback. My sincere thanks to Prof.N.Moorthy Muthukrishnan and Prof.Yoganandam, the previous HODs, EEE department, supported me to carry out my research work without any hindrances

I am thankful to Prof.V.S.Rao, former Director BITS-Pilani, Hyderabad Campus, Prof.G. Sundar, Director BITS-Pilani, Hyderabad Campus and Prof.Souvik Bhattacharyya, Vice Chancellor BITS-Pilani, for giving me opportunity to pursue my research work. Further, I would like to thank Prof.Vidya Rajesh, former Dean ARD and office of ARD and Dr.Venkata Vamshi Krishna, Associate Dean, Academic-Graduate Studies and Research for providing immense support.

I would also like to thank my Doctoral Advisory Committee (DAC) members, Dr.Venkateswaran Rajgopalan and Dr.S.K.Chatterjee for their valuable comments and suggestions. I would like to express my sincere thanks to Prof.Sanket Goel, Head of the Department of Electrical Engineering, for his constant support and encouragement. I am thankful to Dr.Prasant Kumar Pattnaik, DRC Convenor, Department of Electrical Engineering, for overseeing all the research procedures to carry out work in the department. I am also very

thankful to Prof. S.K.Sahoo for his support through the course of my research work. I am very grateful to all the faculty of Dept EEE for their encouragement and support to carry out my research work. I would like to express my sincere thanks to Dr. Jwala srikala, Radiologist KIMS Hyderabad for spending time, discussing my research work and providing me mammography Images and suggestions.

A special thanks to all my colleagues who helped me during different stages of my research work. I want to thank Mr. Gautam Makkena for his help in writing Journal. I thank Mr. M.K.Kaushik and Ms U.Poorna Laxmi for their valuable suggestions and support during my difficult times. I thank Dr. R.Manjula sri for her constant support and constructive feedback on my research work and thesis.

My heartfelt thanks to my mother in law and father in law, Smt. S.Bhagya laxmi and Sh. S.Venkata swamy for their blessings and encouragement. I especially thank my brothers in law S.Raghavendra swamy for his encouragement to enroll into PhD program. I appreciate all my family members for their love and affection.

Words are not enough to thank my husband Madhav for his continuous encouragement, sacrifices, love, patience and support during my research work. I appreciate my son and daughter S.Akshith and S. Sriya for their compromises in my absence.

Last but not the least, I would like to thank my parents, Sh P.Chakradhara swamy and Smt P.Padma sree for educating me and their support throughout my research work. Their blessings and love helped me to reach goal.

Abstract

Breast cancer is the most common cancer in women in India and world over the last decade. It is the second most leading cause of cancers death among women. Breast cancer accounts for 25% to 31% of all cancers in women in India. Incidence rate has been increased from 22.2% in 2008 to 27.05% in 2012 in India. According to WHO for the year 2012, an estimated 70,218 women died in India due to breast cancer, more than any other country in the world. According to the reports of Indian Council for Medical Research in 2016 the breast cancer cases expected to rise by about 14.5 lakhs and this may increase to 17.3 lakhs in 2020. Digital mammography plays a vital role in the early detection of breast cancer. It helps to avoid wrong diagnosis, unwanted examinations, and inadequate surgeries, which directly affects the outcome and longevity of the patient. However, due to low contrast and high noise in the mammograms, it is challenging task to detect the abnormalities. Thus, for early detection and diagnosis of breast cancer in mammograms, there is a significant need of computer aided system (CAD). CAD system uses image processing and computerized techniques to pinpoint the abnormalities and classify mammogram into several classes namely normal or abnormal, benign or malignant etc.

In this thesis, the contributions include the development of new algorithms for image enhancement, segmentation and feature extraction for early detection of breast cancer using mammograms. In addition to that, an automatic CAD system has been developed for the detection and analysis of abnormalities without user intervention. In this context, we have developed algorithms for image enhancement, image segmentation and different feature extraction schemes based on both shape and gray level characteristics of abnormality in mammograms. These relevant features extracted are submitted to classifier for further analysis.

The analysis includes classification of breast regions into normal-abnormal and benign-malignant. These schemes developed for CAD system has been validated using Digital Database for Screening Mammography (DDSM) and Image Retrieval in Medical Applications (IRMA) databases. The results obtained by the algorithms are very promising in terms of accuracy and computation time when compared to the existing systems in the literature and have been detailed in chapters.

In chapter 1, an introduction of different medical imaging modalities was given and discussed about the types of breast cancer followed by survey on the different methods developed by researchers for early detection of breast cancer.

In chapter 2, the novel method for the mammogram enhancement is developed using two dimensional discrete wavelet transform and Gaussian pyramid techniques. Further our method was validated using EBCM (Edge based contrast measure) and entropy values. We also developed an algorithm for image segmentation in this chapter and validated using dice coefficient.

In chapter 3, features have been extracted using Rotation Invariant Local Frequency (RILF), Local binary pattern (LBP) and Segmentation based Fractal Texture analysis (SFTA) techniques for the classification of breast regions into normal-abnormal . For evaluation of these feature extraction techniques, we have used linear SVM classifier via 10 fold cross validation. On application of all the above three methods, it is observed that RILF attained better accuracy.

In chapter 4, features are generated from abnormalities (masses) to classify them into benign-malignant. Geometric features, polynomial regression, triangular area representation (TAR) and beam angle statistics (BAS) techniques are applied on the mass contours to generate significant

shape based features. Among all these methods, TAR method gave superior classification accuracy.

In chapter 5, we have discussed the simulation results for image enhancement, segmentation and feature extraction techniques and these are applied on mammograms, breast tissue regions called as regions of interest (ROIs) and mass contours respectively. We also implemented all the above methods by considering complete mammogram for detection and analysis of abnormal regions. All the simulations are carried out using MATLAB software.

Chapter 6 describes a hybrid method that automatically detects and analyzes the abnormality in mammogram without user intervention. Hybrid method includes combination of Dunn index based k-means algorithm for mass detection and modified SFTA method to generate features for normal, benign and malignant regions in mammogram. This feature vector is submitted to SVM classifier for further analysis.

In chapter 7, the summary of the entire work has been furnished. Also, we have discussed about the scope of future work on different types of breast abnormalities and multiclass classification.

Contents

Certificate.....	i
Acknowledgements	iii
Abstract.....	v
Contents.....	viii
List of Figures.....	xii
List of Tables.....	xvii
Abbreviations.....	xxii
Chapter 1: Introduction.....	1
1.1 Different types of imaging modalities.....	1
1.2 Breast cancer.....	4
1.3 CAD systems.....	6
1.4 Performance measures.....	8
1.5 Databases.....	10
1.6 Literature Survey.....	12

1.7 Motivation.....	17
1.8 Classifier.....	17
1.8.1 SVM classifier.....	17
1.8.2 KNN classifier.....	18
1.8.3 Decision Tree Classifier.....	19
1.8.4 Ensemble methods.....	20
1.9 Research Objectives.....	22
1.10 Organization of thesis.....	23
Chapter 2: Image enhancement and segmentation of mammograms.....	24
2.1 Database.....	25
2.2 Multiresolution analysis using 2D wavelets and wavelet fusion.....	25
2.3 Gaussian Pyramid.....	27
2.4 Contrast Limit Adaptive Histogram Equalization (CLAHE).....	27
2.5 Algorithm for Image enhancement.....	29
2.6 Image Segmentation.....	29
2.6.1 Algorithm for segmentation.....	31
2.7 Summary.....	33
Chapter 3: Mammogram classification to normal and abnormal	
 using texture features.....	34
3.1 Extraction of ROI.....	35
3.2 Local Binary Pattern and its variants.....	35
3.2.1 Local binary pattern.....	35
3.2.2 Local ternary pattern.....	37

3.2.3 Local quinary pattern.....	38
3.3 Segmental Fractal Texture Analysis (SFTA).....	39
3.4 Rotation Invariant Local Frequency (RILF).....	41
3.5 Implementation of Feature Extraction techniques.....	42
3.5.1 Method 1.....	43
3.5.2 Method 2.....	44
3.5.3 Method 3.....	44
3.6 Classification.....	45
3.7 Summary.....	46

Chapter 4: Mammogram classification to benign and malignant using

contour based features.....	47
4.1 Extraction of mass contours.....	48
4.2 Geometric features.....	49
4.3 Polynomial regression.....	49
4.3.1 Polynomial regression fit by gradient descent algorithm.....	50
4.3.2 Extraction of features.....	51
4.4 Beam Angle Statistics.....	53
4.4.1 Conversion of 2D contour to 1D signature using Beam angle statistics (BAS).....	54
4.4.2 Extraction of features from BAS moments.....	56
4.5 Triangular area Representation.....	58
4.5.1 Triangular area representation (TAR) signatures.....	58
4.5.2 Conversion of 2D contour into 1Dsignal and feature extraction.....	59

4.6 Classification.....	60
4.7 Summary.....	61
Chapter 5: Experimental Results and analysis.....	62
5.1 Image Enhancement.....	62
5.2 Image Segmentation.....	66
5.3 Analysis of mammograms into normal/abnormal.....	68
5.3.1 Database.....	68
5.3.2 Results of Method 1.....	69
5.3.3 Results of Method 2.....	74
5.3.4 Results of Method 3.....	78
5.3.5 Summary.....	82
5.4 Analysis of mammograms into benign-malignant.....	82
5.4.1 Geometric features.....	82
5.4.2 Polynomial Regression.....	82
5.4.3 Triangular area representation (TAR).....	85
5.4.4 Beam angle Statistics (BAS).....	88
5.4.5 Comparisons between all the developed methods.....	92
5.4.6 Comparisons of our developed methods and other existing methods...	92
5.5 Implementation of above algorithms on complete mammograms.....	93
5.6 Conclusion.....	97
Chapter 6: Automatic detection and analysis of abnormality in Mammogram.....	98
6.1 Preprocessing.....	99
6.1.1 Breast area region extraction and label removal.....	100

6.1.2 Pectoral muscle removal.....	100
6.1.3 Noise removal.....	100
6.2 Segmentation of suspicious areas.....	100
6.3 Feature extraction.....	104
6.3.1 MSFTA.....	104
6.3.2 Feature extraction for mas-non-mass classification.....	105
6.3.3 Feature Extraction from breast masses for malignant-benign Classification.....	106
6.4 Classification.....	108
6.5 Experimental analysis	108
6.5.1 Data set.....	108
6.5.2 Evaluation for mass-non-mass classification.....	109
6.5.3 Evaluation for benign-malignant classification.....	112
6.5.4 Discussion.....	119
6.6 Conclusion.....	122
Chapter 7: Conclusions and Future Work.....	123
7.1 Introduction.....	123
7.2 Implementations of the proposed methods.....	123
Bibliography.....	126

List of Figures

Figure.1.1: Block diagram of CAD system	8
Figure 1.2: Typical ROC curve for non-mass-benign classification	10
Figure 1.3 : Benign and malignant mammograms	11
Figure.1.4: ROIs of IRMA Database	12
Figure 2.1: Block diagram of image enhancement	30
Figure 2.2: a) Original mammogram image	30
Figure 2.2: b) proposed method enhanced image	30
Figure 2.3: Block diagram of segmentation process	32
Figure 2.4: a) Original mammogram	33
Figure 2.4: b) Output after thresholding,	33
Figure 2.4: c) Segmented image	33
Figure 3.1: a) Circular symmetric neighbor for $P=8$ and $R=1$	36
Figure 3.1: b) Illustration of LBP operator	36
Figure 3.2: Splitting of LTP pattern into two different LBP codes	38
Figure 3.3: Example for splitting of LQP into four LBP patterns	39
Figure 3.4: Block diagram for the classification of ROIs using method 1	43
Figure 3.5: Block diagram for the classification of ROIs using method 2	44
Figure 3.6: Block diagram for the classification of ROIs using method 3	45
Figure.4.1: a) Benign	50
Figure.4.1: b) Malignant	50
Figure.4.1: c) Signature of benign	50

Figure.4.1: d) Signature of malignant	50
Figure.4.2: a) Malignant signature and its 15 degree polynomial hypothesis	52
Figure.4.2: b) Benign Signature and its 15 degree polynomial hypothesis	52
Figure 4.3: Block diagram for the extraction of features and classification using polynomial regression	52
Figure 4.4: Block diagram for the extraction of features and classification using BAS	53
Figure 4.5: a) Malignant	54
Figure 4.5: b) Benign	54
Figure.4.6: a) The beams of point $p(i)$	55
Figure.4.6: b) The 3-curvature and the beam angle at the neighborhood system 3 for the boundary point $p(i)$	55
Figure 4.7: a) First and second order statistics of malignant contour	56
Figure 4.7: b) First and second order statistics of benign contour	56
Figure 4.8: Block diagram for the extraction of features and classification using TAR	59
Figure. 4.9: TAR signatures for benign contour and malignant contour	60
Figure 5.1: a) Original mammogram image	63
Figure 5.1: b) proposed method image	63
Figure 5.1: c) unsharpened image	63
Figure 5.1: d) CLAHE image	63
Figure 5.2: a) Original images selected from DDSM	67
Figure 5.2: b) Ground truth segmented mass	67

Figure 5.2: c) Output Segmented mass	67
Figure 5.3: Results of LBP (mass-normal)	71
Figure 5.4: Results of LTP (mass-normal)	71
Figure 5.5: Results of LQP (mass-normal)	71
Figure 5.6: Results of LBP (microcalcification-normal)	72
Figure 5.7: Results of LTP (microcalcification-normal)	72
Figure 5.8: Results of LQP (microcalcification-normal)	72
Figure 5.9: Results of SFTA (mass-normal)	73
Figure 5.10: Results of SFTA(microcalcification-normal)	73
Figure 5.11: Results of RILF (mass-normal)	73
Figure 5.12: Results of RILF (microcalcification-normal)	74
Figure 5.13: Results of BLBP (mass-normal)	76
Figure 5.14: Results of BLBP (microcalcification-normal)	76
Figure 5.15: Results of BSFTA (mass-normal)	76
Figure 5.16: Results of BSFTA (microcalcification -normal)	77
Figure 5.17: Results of BRILF(mass-normal)	77
Figure 5.18: Results of BRILF(microcalcification-normal)	77
Figure 5.19: Results of Pyramidal Local binary Pattern (mass-normal)	79
Figure 5.20: Results of Pyramidal Local binary Pattern (microcalcification-normal)	80
Figure 5.21: Results of Pyramidal SFTA (mass-normal)	80
Figure 5.22: Results of Pyramidal SFTA (microcalcification-normal)	80
Figure 5.23: Results of Pyramidal RILF(mass-normal)	81
Figure 5.24: Results of Pyramidal RILF (microcalcification-normal)	81

Figure 5.25: Plot of performance evaluation parameters with different classifiers	84
Figure.5.26: ROC analysis with SVM classifier	87
Figure 5.27: Confusion matrix of our SVM(Quadratic) classifier	88
Figure 5.28: Block diagram for identification and analysis of mass in Mammograms	95
Figure 5.29: a) Enhanced Mammograms	96
Figure 5.29: b) Segmented Masses for benign-malignant classification	96
Figure 5.29: c) ROIs for normal-abnormal classification	96
Figure 6.1: Block diagram of the proposed model for automatic detection and Analysis	99
Figure 6.2: a) Original mammogram	102
Figure 6.2: b)Pre-processed mammogram	102
Figure 6.2: c) Output after DAK algorithm	102
Figure 6.2: d) Dense cluster with mass	102
Figure 6.3: Plot of Dunn index values vs number of clusters	103
Figure 6.4: Flow chart for the selection of Dense cluster	103
Figure.6.5: Original mammogram, dense cluster, suspicious areas selection from dense cluster	104
Figure.6.6: Plot of sensitivity and specificity for different feature sets	110
Figure 6.7: ROC analysis for mass-non-mass	111
Figure 6.8: Plot of sensitivity and specificity for different feature sets	116
Figure 6.9: ROC analysis for benign-malignant	116
Figure 6.10: ROC analysis for non-mass-malignant classification	117
Figure 6.11: ROC analysis for non-mass-benign classification	117

List of Tables

1.1	Confusion Matrix for binary classification system.....	9
5.1	EBCM values of mammographic images	64
5.2	Entropy values of KIMS mammographic images (Images from SNO 1 to 14 are from KIMS and images from SNO 15 to 23 are from KIMS).....	65
5.3	Dice Coefficients of benign and malignant mass.....	68
5.4	Evaluation Parameters With SVM (%).....	83
5.5	Evaluation Parameters With LDA (%).....	83
5.6	Evaluation Parameters With Bayes Linear Classifier.....	84
5.7	Accuracies in (%) with different classifiers.....	86
5.8	Results achieved with different testing/training configuration using 148 mass contours	87
5.9	Different Classifiers and its input parameters.....	91
5.10	Accuracies of 1D BAS functions with FD, DWC and RMS.....	91
5.11	Comparison of accuracies with 1D BAS and CDF functions with FD, DWC and RMS features.....	91
5.12	Classification accuracies with different classifiers.....	92
5.13	Comparison of benign-malignant feature extraction techniques.....	94
5.14	Comparison of benign-malignant feature extraction techniques with state of the art methods	95
5.15	Accuracy for the normal-abnormal classification using complete mammogram with the proposed methods.....	97

5.16	Accuracy for the benign-malignant classification using complete mammogram with the proposed methods.....	97
6.1	Features for mass-non-mass.....	108
6.2	Average classification accuracy (%) with different features sets using tenfold cross validation with SVM for mass-non-mass.....	110
6.3	Average classification accuracy, sensitivity and specificity with MSFTA-PGF feature set in (%).....	110
6.4	Comparison for benign-malignant classification in (%) using SVM (linear) kernel with different methods.....	112
6.5	AUC with different methods for mass-non-mass.....	112
6.6	The computational time evaluation (mass-non-mass).....	112
6.7	Features for benign-malignant.....	115
6.8	Average classification accuracy with different features using tenfold cross validation with SVM for benign-malignant (%).....	115
6.9	Average classification accuracy, sensitivity and specificity with MSFTA-PGF feature set (%).....	117
6.10	Mean evaluation parameters for non-mass-malignant.....	118
6.11	Mean evaluation parameters for non-mass-benign.....	118
6.12	Average accuracies for three class study using ten fold cross validation with different classifier in (%).....	118
6.13	Comparison for benign-malignant classification in (%) with different methods	118
6.14	Comparison of AUC with different methods (benign-malignant).....	119

6.15	Computational time Evaluation (benign-malignant).....	119
6.16	Comparison of feature extraction methods for the classification of mass-non- mass.....	121
6.17	Comparison of benign-malignant classificatin feature extraction techniques.....	122

Abbreviations

Artificial Neural Network	ANN
Beam Angle Statistics	BAS
Block Wise-Local Binary Pattern	BLBP
Block Wise-Local Quinary Pattern	BLQP
Block Wise-Local Terenary Pattern	BLTP
Block Wise-Rotational Invariant Local Frequency	BRILF
Block Wise-Segmentation Based Fractal Analysis	BSFTA
Breast Imaging Reporting And Data System	(BI-RADS)
Computed Tomography	CT
Computer Aided Diagnosis	CAD
Contrast Limit Adaptive Histogram Equalization	CLAHE
Cumulative Distribution Function	CDF
Detail Wavelet Coefficients (DWC)	DWC
Digital Database For Screening Mammography	DDSM
Discrete Fourier Transform	DFT
Discrete Wavelet Transform	DWT
Ductal Carcinoma In Situ	DCIS
Dunn Index based Automated K-Means Algorithm	DAK
Elliptical Binary Pattern	EBP
Edge Based Contrast Measure	EBCM
False Positive Rate	FPR

Fourier descriptors	FD
Gray Level Co-Occurrence Matrix	GLCM
Gray Level Run Length Matrix	GLRLM
Ground Truth	GT
Image Retrieval In Medical Applications	IRMA
Infiltrating Ductal Carcinoma	IDC
Infiltrating Lobular Carcinoma	ILC
K-Nearest-Neighbor	KNN
Krishna Institute Of Medical Sciences	KIMS
Local Binary Pattern	LBP
Local Circular Function	LCF
Local Quinary Pattern	LQP
Magnetic Resonance Imaging	MRI
Maximal Radial Length	MRL
Modified Segmented Fractal Texture Analysis	MSFTA
Negative Predictive Value	NPV
Optical Density Cooccurrence Matrix	ODCM
Principal Component Analysis	PCA
Positive Predictive Value	PPV
Positron Emission Tomography	PET
Pyramidal-LBP	PLBP
Pyramidal-LQP	PLQP
Pyramidal-LTP	PLTP

Pyramidal-RILF	PRILF
Pyramidal-SFTA	PSFTA
Radial Basis Function	RBF
Receiver Operating Characteristic	ROC
Regions Of Interest	ROI
Root Mean Square	RMS
Rotation Invariant Local Frequency	RILF
Rotation Invariant Local Frequency Magnitude Descriptors	RILFMD
Segmented Image	SI
Segmentation based Fractal Texture Analysis	SFTA
Single-Photon Emission Computed Tomography	SPECT
Support Vector Machine	SVM
Triangular Area Representation	TAR
True Negative Rate	TNR
True Positive Rate	TPR
Two Dimensional Discrete Wavelet Transform	2D-DWT
Two-Threshold Binary Decomposition	TTBD
Ultrasound	US

Chapter 1

Introduction

Biomedical image processing techniques plays a significant role in today's health care due to increased utilization of medical images in clinical investigation and disease diagnosis. It has become interdisciplinary research field for experts in mathematics, engineering, biology, physics, statistics and medicine. Medical images visually represent the internal organs or tissues of a body for clinical analysis. Generally, medical images are produced using different imaging techniques namely, magnetic resonance imaging (MRI), ultrasound imaging, X-rays, computed tomography (CT) and molecular imaging etc. An organ or tissue in digital medical images is composed of individual pixels to which discrete bright or color value is assigned. The internal physiological structures hidden inside the skin in medical images can be manipulated to produce high quality information for the disease diagnoses and treatment. The reconstruction and processing of medical images depends on the type of medical imaging modality that is used to acquire images. Medical imaging modality means the mode by which the digital image is acquired by an equipment to get visual information on structural or functional parts of the body. such as radiography, ultrasound, nuclear medicine, computed tomography, magnetic resonance imaging and visible light. These imaging modalities uses the basics of transmission, reflection or refraction of light, radiation, temperature, sound, or spin [1]. Different imaging modalities that are typically used by doctors are described below.

1.1 Different types of imaging modalities

X-rays: In this imaging modality, X-rays are penetrated into the human body and absorbed by a film. The amount of X-rays penetrated into the human body depends on the type of material.

Human body consists of soft tissues, bones and air. A film exposed to the X-rays that come from human body will have light or very bright areas (less exposure), gray areas (high exposure), and black areas (heavy exposure) [2] depending on the amount of X-rays penetrated into different parts of body. For example, bones are hard structures and let very little amount of X-rays to pass through. So, they will appear bright. Air in the body allows more amounts of X-rays to pass through. So, it will appear dark. This modality is used to diagnose heart and chest diseases, breast cancer (mammography), and dental problems etc.

Ultrasonography: In ultrasonography, sound waves of frequencies 2 to 20 MHz are used [3]. This medical imaging modality is predominately used to examine soft tissues of human body. Ultrasound (US) consist of transducers which have a specific frequency. The sound waves that are transmitted into the body are reflected back to the surface (at the receiver). The time taken for the sound wave to go through soft issues and reflected back at the receiver is used by a computer to form the image. The main limitation of ultrasonography is that, the sound waves cannot penetrate into air or bone like structures. So, it cannot be used for examinations of skeleton and brain which are surrounded by bones. The sound wave frequency also plays a vital role to generate in the forming image with different spatial resolution. Low frequencies generate image with less spatial resolution whereas high frequency sounds waves have less wavelengths and thus have the ability of reflecting from small structures resulting in high resolution images.

Computer Tomography: CT refers to computerized x-ray imaging procedure [4]. In computer tomography the images are formed from X-rays. The measurements are taken based on the amount of X-rays transmitted to the body. Multidirectional scanning is performed to collect multiple data from X-rays transmitted through the body. These contain detailed information on the different areas inside the body.

In this CT system, an x-ray source was placed inside the circular opening called gantry. A stationary patient on the table is moved inside the gantry to irradiate the patient's body part through x-ray source and the detector on the other side of the patient collect exiting x-rays from patient's body at some angle. In this way, x-rays are collected at different angles and sent to computer. These rays are reconstructed into individual slice for one rotation of gantry with x-ray source. For typical examination, there can be 10 to 50 rotations of the x-ray tube around the patient. As x-ray sources cannot pass through soft tissues, CT doesn't show soft tissues inside the body. This modality is used for the diagnosis of brain tumors, kidney, liver diseases etc.

Magnetic Resonance imaging (MRI): MRI is a non-invasive imaging technology that uses magnetic field, radio waves and computer to produce anatomical pictures of human body [5]. MRI machine consists of magnets which generates strong magnetic field inside patient's body. The body contains water molecules. When a person goes inside the powerful magnetic field of MRI machine. This magnetic field results in the alignment of photons towards field direction. When the radio frequency current is send through patient's body, protons inside the body move against the field. However, when the radio frequency current is turned off, protons try to realign in the direction of magnetic field. The radio frequency of the signal depends on the strength of the magnetic fields. In this process protons release energy. This energy is collected by MRI sensors to form the image. MRI scanners are suited to image tissues of the body. It is used to diagnose brain tumors, breast cancer, spinal stenosis etc. The tumors can be detected easily as the protons return to the equilibrium state with different rate compare to healthy tissues. The contrast agents have been injected intravenously to enhance the visibility of blood vessels.

Molecular imaging: Molecular imaging is a procedure that gives images with detailed formation related to molecular and cellular level changes inside the human body. It consists of imaging

device and radioactive atom or isotope. The isotope is injected into human body and the distribution of isotope is observed by imaging device. The isotope attaches to the abnormal cells and this help the physicians to detect the severity of disease. Positron emission tomography (PET), and single-photon emission computed tomography (SPECT) are examples of molecular imaging. The main advantage of molecular imaging is that it gives functional information of body parts. This imaging procedure helps to diagnose alzheimer's disease, epileptic seizures, heart diseases etc.

In this thesis, we have investigated the analysis of abnormalities in mammograms to detect breast cancer in early stages. In the consequent part of the thesis, we are confined to the topic of this research.

1.2 Breast Cancer

Breast cancer is a major health problem among women throughout the globe. The mortality rate due to breast cancer has increased both in developed and developing countries. According to global estimates WHO report 2013, it is estimated that 5,08000 women died due to breast cancer in 2011 [6]. It is estimated that nearly 1.7 million cases were diagnosed with breast cancer in 2012. Breast cancer represented about 12% of all new cancer cases and 25 % of all cancers in women. This shows that there is a sharp rise in breast cancer cases in all types of countries. In India, 1,55,000 new cases of breast cancer has been diagnosed in the year 2015 and 76000 women were expected to die due to breast cancer [7].

Breast cancer occurs due to uncontrolled growth of cells. The breast consists of fat, connective tissue, glands and ducts. Most of the abnormal cells grow in ducts to form ductal carcinoma. The cells that grow in lobules is called lobular carcinoma [8]. The malignant tumors can be classified into two types: invasive and non-invasive. In invasive, cancerous tumor cells spread from the

duct or lobule to all other organs of a body. This type of breast cancer is called as an infiltrating ductal carcinoma (IDC), when it starts in ducts and infiltrating lobular carcinoma (ILC), when it initiates in lobules. Non-invasive cancerous tumor occurs when cancer cells are confined to ducts or lobule and occurs during the early stages. These types of breast cancers are called as Ductal carcinoma in situ (DCIS) or Lobular carcinoma in situ (LCIS) [8]. There are several abnormalities that affect the breast tissues. These abnormalities include masses, microcalcification and architectural distortion. When abnormal cells grow at one particular place, they form a mass. A mass can be defined by its shape margin and density [9]. Shape determines the malignancy of mass. Masses with stellate or nodular shape are malignant. Microcalcifications are tiny calcium deposits that occur with different patterns. A group of tiny calcifications with in a small area is an early indicator of breast cancer. Detection of calcifications is very difficult due to its fuzzy appearance, especially in dense breasts. The shape, density, distribution, location determines whether the calcifications are benign or malignant. These calcifications usually occur in young people. Architectural distortion is disruption of the normal parenchymal pattern. The mass does not appear on breast x-ray but, distortion occurs at the edge of the parenchyma, which appears as a stellate shape.

Early detection of breast cancer helps in increasing survival rates of women suffering from breast cancers. Digital mammography plays vital role by detecting the abnormalities when they are still localized. Mammogram helps to detect non-palpable breast masses in asymptomatic women. But, radiologists interpret numerous mammograms for detecting small number of cancers on daily basis. Hence, this results in decrease in accuracy or increase in repeatability. In order to increase the accuracy of the screening two radiologists have to double read the mammograms to validate the abnormalities. But, it is very difficult to practice in countries, where there is

insufficiency of radiologists. To overcome this problem, new automatic recognition system using computer environments have to be developed for second opinion. This developed system is called as computer aided diagnosis system (CAD). The procedure of CAD framework to detect abnormalities on mammogram is discussed below.

1.3 CAD system

CAD system uses image processing and computerized techniques to pin point the abnormalities in mammograms [10]. It has two stages: one is detection of suspicious area locations in mammograms and another one is to analyze suspicious abnormality. General block diagram of CAD system is given in Figure1.1. The three main phases of the CAD system are: training, testing and evaluation. Training phase includes preprocessing and feature extraction. Testing phase includes classification step. Evaluation involves the validation of developed algorithm with several metrics that include accuracy, sensitivity, specificity, receiver operating characteristic (ROC) analysis etc. The dataset which contains mammograms used for CAD system is divided into two parts: one is training dataset and another one is testing dataset. Phases of CAD system are discussed below:

1.3.1 Training

The input of this phase is training dataset and the output is the features extracted from mammographic tissues.

- a) **Preprocessing:** Mammograms contains noise due to the acquisition process. Preprocessing of mammograms involves noise removal, segmentation of breast area and pectoral muscle removal. It requires imaging filter, transformation to processing, thresholding etc .
- b) **Feature Extraction:** In feature extraction stage, features that define characteristics of abnormal and normal tissue are extracted to classify them. Feature extraction can be

performed manually or with many algorithms like morphology, segmentation etc. This stage also involves feature selection process if the number of features is high.

1.3.2 Testing

The input of this phase is features extracted from training dataset and testing dataset and output is the identification of mammographic tissue as benign or malignant and normal or abnormal.

a) Classification: In classification, the model is learned with the extracted features of training dataset. This model classifies extracted features of testing dataset into benign and malignant.

Classification can be performed by machine learning algorithms, neural networks, deep learning models etc.

b) Evaluation: In this phase, the developed algorithm for training and testing is validated using different performance metrics like ROC, sensitivity, accuracy and specificity. The definitions of these performance metrics are discussed below in the next section.

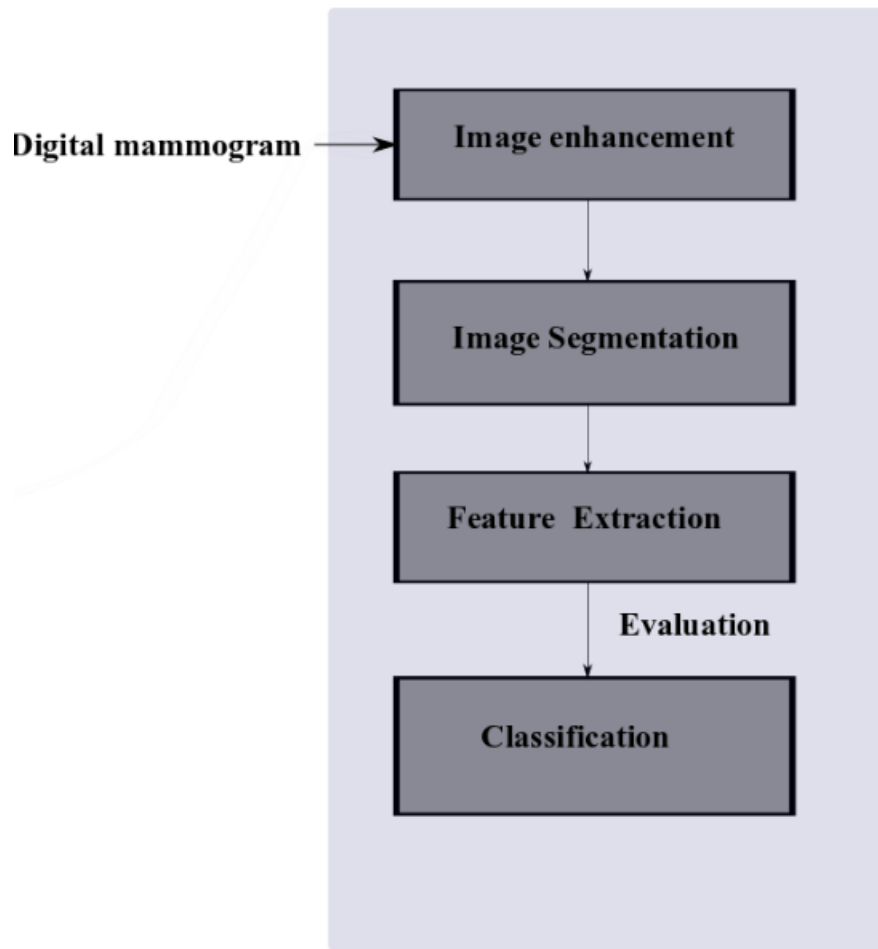


Figure.1.1:Block diagram of CAD system

1.4 Performance metrics

In two-class classification of normal-abnormal or benign-malignant, one class is considered as positive and other class as negative. Confusion matrix can be built based on classifier performance. Confusion matrix summarizes the number of cases which are correctly or incorrectly identified.in Table1.1.

Table 1.1: Confusion Matrix for binary classification system

Actual Class	Predicted Class	
	Positive	Negative
Positive	True positive(TP)	False Negative (FN)
Negative	False positive(FP)	True Negative(TN)

Following are the performance measures that are calculated from confusion matrix to evaluate the performance of classifier

a) **Accuracy (ACC):** It is the ratio of number of cases which are predicted correctly to total number of cases.

$$ACC(\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1.1)$$

b) **True positive rate (TPR) or Sensitivity (Sens):** Percentage of positive cases that are correctly predicted by the algorithm.

$$TPR(\%) = \frac{TP}{TP+FN} \times 100 \quad (1.2)$$

c) **False Positive rate (FPR):** It is the percentage of negative cases predicted as positive

$$FPR(\%) = \frac{FP}{TP+FN} \times 100 \quad (1.3)$$

d) **True Negative rate (TNR):** Percentage of negative cases predicted correctly

$$TNR(\%) = \frac{TN}{TN+FP} \times 100 \quad (1.4)$$

e) **False Negative rate (FNR):** Percentage of positive cases predicted as negative

$$FNR(\%) = \frac{FN}{TN+FP} \times 100 \quad (1.5)$$

f) **Positive predictive value (PPV):** It is defined as the percentage of cases that are actual positives to total number of positives predicted by classifier and is given by

$$PPV(\%) = \frac{TP}{TP+FP} \times 100 \quad (1.6)$$

g) **Negative predictive value (NPV):** It is defined as the percentage of cases that are actual negative to total number of negatives predicted by the classifier and is given by

$$NPV(\%) = \frac{TN}{TN+FN} \times 100 \quad (1.7)$$

h) **Receiver Operating Characteristics (ROC):** It is a two dimensional plot that shows tradeoff between true positive rate and false positive rate as shown in the Figure 1.2. TPR is represented on Y-axis and FPR is represented on X-axis. Area under this ROC curve (AUC) measures the performance of classifier. AUC of ideal classifier is one.

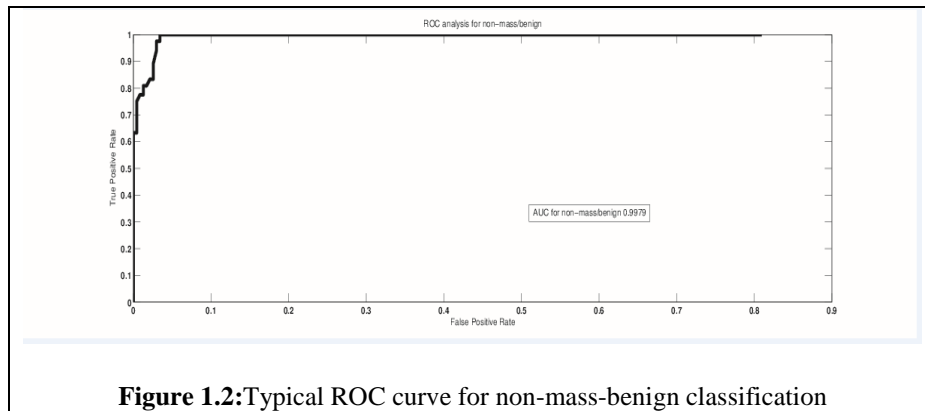


Figure 1.2: Typical ROC curve for non-mass-benign classification

1.5 Databases

To evaluate our developed methods mammographic images are retrieved from Digital database for screening mammography (DDSM) and Image Retrieval in Medical Applications (IRMA) database. For real time application,

we also tested some of our algorithms on mammographic images taken from Krishna institute of medical sciences (KIMS), Hyderabad. The DDSM images [11] are taken from Massachusetts General Hospital, Washington University School of Medicine, Wake. The primary purpose of the database is to facilitate sound research in the development of computer algorithms to aid in screening the images. They have 2620 cases in 43 volumes. The images are digitized by (HOWTEK-A, HOWTEK-D, LUMISYS, and DBA) scanners. Mammograms are taken from two

different views. They include Mediolateral (MLO) view and Craniocadal (CC) view along with patient age, American College of Radiology (ACR) breast density rating, subtlety rating for abnormalities and ACR keyword description of abnormalities. Each view has both left and right breast images. Figure 1.3 shows two mammograms having malignant and benign masses taken from DDSM database. Spatial resolution of images is 42.3 microns, 43microns and 50 microns for DBA, Howtek and Lumisys scanner respectively. IRMA is a standard mammographic database that comprises of regions of interest (ROIs) taken from digital mammogram images of selected patients from the Radiological Diagnosis Department of the University of Aachen [12]. The datasets in the IRMA project have three different types of breast regions or ROIs of mammograms, which include: Normal (no cancer), benign and malignant. The abnormal (benign and malignant) ROIs have three different types of abnormalities: Masses, microcalcifications and others (Architectural distortion) as shown in Figure 1.4. The images in the dataset are of 128 X 128 pixel size and at 256 gray levels.

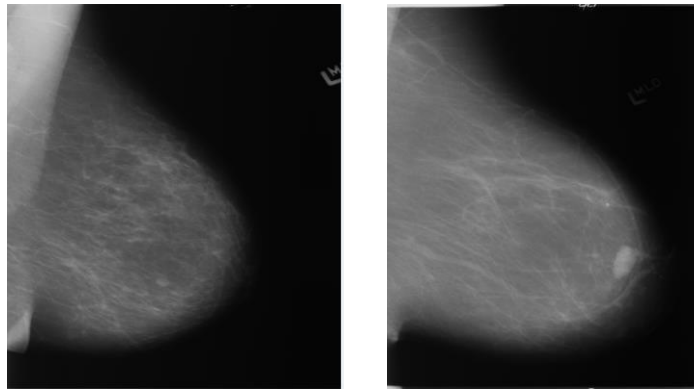


Figure 1.3 :Benign and malignant mammograms

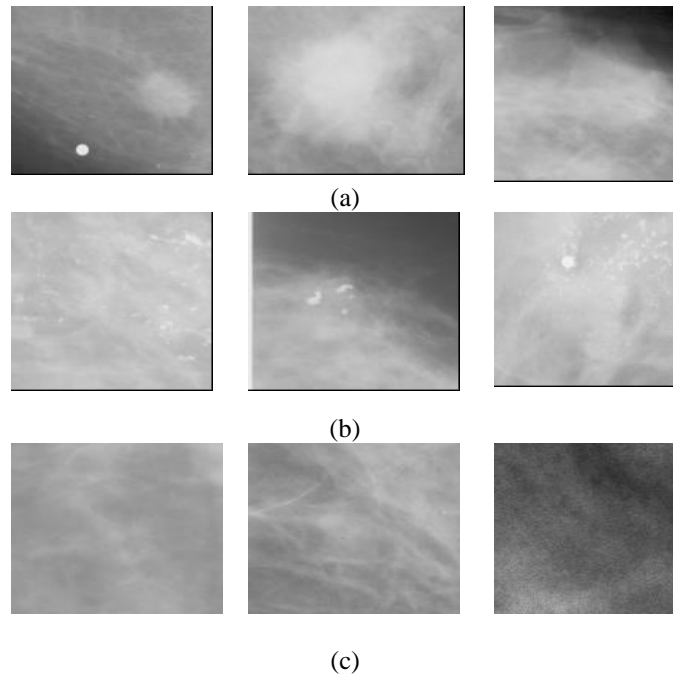


Figure.1.4: ROIs of IRMA Database a) Mass b) microcalcification c) Normal

1.6 Literature Survey

Digital mammography plays a vital role in the early detection of breast cancer. To detect breast cancer in early stages, radiologists look into masses, microcalcification, asymmetry and architectural distortion in mammograms. But, due to low contrast and high noise in the mammograms, it is very difficult task to detect the abnormalities. Thus, for early detection and diagnosis of breast cancer in mammograms there is a significant need of computer aided system (CAD) [13]. CAD system uses image processing and computerized techniques to pinpoint the abnormalities in mammograms. CAD systems in the field of medical image processing are classified into two types [14]. One type is CADe systems which can detect abnormalities from whole mammogram and these abnormalities are further analyzed by radiologists. Another type is CADx systems which are capable of analyzing suspicious masses into normal-abnormal, mass-non-mass, and benign-malignant from small tissue portion of mammograms.

CAD systems consist of four steps: preprocessing, segmentation, feature extraction and classification. Pre-processing involves removal of noise, unwanted areas like back ground, pectoral muscle etc. Among them, removal of pectoral muscle is very important for accurate diagnosis of mass or any other abnormality in mammogram and also decreases the processing time. Several algorithms were developed to detect pectoral muscle in MLO views. H.Abdellatif1 et al [15] proposed the detection of pectoral muscle using biased normalized graph cut segmentation and bezier curve. Li Liu et al [16] used statistical features for the detection of pectoral muscle. Mario Mustra et al [17] used geometry and morphological approximation for the automatic detection of nipple. Image enhancement is another stage in preprocessing which highlights the invisible abnormalities in dense breast. Zhe Wu et al in [18] applied modified unsharp masking method which is based on high pass filter for image enhancement. M. Sundarama et al [19] applied histogram modified local contrast enhancement technique which enhances the local contents in image with the increase in contrast. Image segmentation is another step which differentiates masses from the background and helps us to obtain relevant information of the abnormality. Peyman Rahmati et al in [20] applied segmentation methods based on a novel maximum likelihood active contour model using level sets. This algorithm delineates the mass from the background by modelling both regions (foreground and back ground) with Gamma distribution. Liu et al [21] segmented breast contour based on classical seed region growing algorithm to enhance contour of a mass from background by adjusting threshold value iteratively. Byung-Woo Hong et al [22] segmented the mass based on topographic representation called the isocontour map. Topological and geometrical structure of the image is analysed and features are extracted using multiscale isocontour maps for attaining good delineation. They are many spatial and frequency domain methods for extracting features from automatically detected

masses and breast tissues for classification. Nebi et al [23] presented a system by extracting curvelet coefficients from cropped ROIs from mammogram. They used linear discriminant analysis and principal component analysis (PCA) for feature reduction and support vector machine (SVM) for classification of ROIs into normal-abnormal and benign-malignant. Xiaoming et al [24] detected mass automatically from whole mammogram using adaptive region growing technique and narrow band based active contour method with tunable parameters. Gray level co-occurrence matrix (GLCM) and completed local binary pattern features are extracted from these regions and was used to classify them as normal and abnormal with SVM classifier. Joberth et al [25] developed automatic detection of abnormal regions using quality thresholding and Haralick correlogram. Further geometric features have been extracted for classification of abnormal regions into mass-non mass with SVM classifier. Shubhi Sharma et al in [26] extracted ROIs manually from mammograms for training and used k-means algorithm for extracting ROIs from high density breast cluster for testing. They extracted features from high density cluster using Zernike moments and classified them into malignant-non-malignant using SVM classifier. Llado et al [27] manually selected regions of interest (ROIs) from mammograms and divided the image into different squared regions and further applied local binary pattern for all regions with different configuration for extracting features to classify ROIs into mass-non-mass with SVM classifier. They also compared their method with two dimensional principal component analysis (2DPCA) method. Nanni et al [28] applied local ternary pattern and local phase quantization by varying their parameters on the ROIs for feature extraction. Classification is performed by fusing SVM classifiers. The features were randomly selected by PCA and Neighbourhood preserving Embedded (NPE) methods. Shen-Chuan Tai et al [29] applied template matching using sech templates to detect suspicious regions. Subsequently, complex feature extraction methods based

on GLCM and Optical Density Co-occurrence (ODCM) Matrix are developed for classification of suspicious regions into normal and abnormal. Ying et al [30] presented adaptive region growing algorithm with hybrid assessment function. This function is combined with maximum likelihood analysis and maximum gradient analysis for segmentation of contour accurately. Intensity, shape and edge sharpness features have been extracted with SVM for classifying segmented contours into benign-malignant. Min Dong et al [31] applied improved vector field convolution (VFC) snake for segmentation of the mass region from the ROIs. The shape features, gray level histogram, pixel value fluctuation features of mass and its surroundings were calculated and classified using random forest classifier. They compared accuracy with different classifiers like SVM, grid search SVM (GS-SVM), particle swarm optimization SVM (PSO-SVM) classifiers. Chun-Chu et al [32] used gamma correction equalization and morphology to detect abnormalities. The abnormal regions are classified using automatic detection classifier with normal texture features and principal component analysis (PCA). Shradhananda et al [33] implemented the segmentation-based fractal texture analysis (SFTA) method [34] to extract the textural features from the ROI patches of mammograms to classify them into normal-abnormal. They used fast correlation based filter for feature selection. Edward et al [35] applied 2D wavelets, at different scales and calculated texture features for each subbands. They analysed these wavelet features for different abnormalities of mammogram like microcalcifications and masses and applied naive Bayes classifier for classification. Mencattini et al [36] implemented phase portrait analysis for mass identification and segmented mass using fuzzy c means clustering. Fractal dimensions using box counting and ruler methods, compactness, fractional concavity, spiculation index features were extracted for classifying them into benign-malignant. Marcelo et al [37] extracted multiresolution analysis features from the breast regions or ROIs

using three different wavelet functions, Daubechies 8, Symlet 8 and bi-orthogonal 3.7. For classification, they used polynomial classifier to define ROIs into normal-abnormal. They compared developed classifier output with decision tree, SVM, k-nearest neighbourhood (K-NN). Marcin et al [38] generated optimal template using brightness distribution and Laplacian filter. This template is matched with all regions in mammograms. The regions with high correlation coefficient are detected as abnormality. Daniel et al [39] determined certain asymmetric regions between pairs of mammograms of the left and the right breast by cross variogram function and a variogram function to differentiate asymmetric region into mass or non-mass.

Rangayyan et al [40] calculated fractal features for 1D and 2D contours of breast masses for classifying them into benign-malignant. Pelin et al [41] implemented local seed region growing (LSRG) technique for the detection of abnormalities. Then, they applied spherical wavelet transform (SWT) for extracting features from abnormalities and classified them into mass-non-mass and benign-malignant. Danilo et al [42] developed a method for detection and segmentation of masses using multiple thresholding, wavelet transform and genetic algorithm in CC and MLO views. Aswini et al in [43] calculated gray level co-occurrence matrix (GLCM) and gray level run length matrix (GLRLM) features for classifying ROI patches into malignant-benign using decision tree classifier. Dubey et al [44] compared two semi-automated segmentation methods, level set and marker controlled watershed method for segmenting tumour in mammogram. Oliveira et al in [45] segmented abnormal regions in mammograms using the k-means algorithm and co-occurrence matrix. The segmented regions are further classified into mass-non-mass using SVM with Haralick and shape features.

1.7 Motivation

From the above survey it has been observed that they are two areas of research in the early detection of breast cancer. One area is automatic detection of abnormalities in mammogram using segmentation techniques and these abnormalities are further analyzed by radiologists. Another area is to crop the region of interest (ROIs) from mammograms using bounding box. Then, different feature extraction techniques have been applied on ROIs to classify them into normal-abnormal or mass-non-mass or benign-malignant. From the existing literature it can be observed that image enhancement, image segmentation and feature extraction methods play a crucial role in the existing algorithms. These algorithms have been mostly evaluated using DDSM database. In view of the existing literature, we observed that there is an ample scope to develop new texture and shape features and an algorithm that automatically detects and classify abnormal regions in mammogram without manual intervention.

1.8 Classifiers

To validate our developed feature extraction algorithms, the features extracted from mammograms are given to various classifiers. These classifiers include Support Vector Machine (SVM), K-Nearest Neighborhood (KNN), ensemble adaboost and simple tree classifiers.

1.8.1 Support vector machine (SVM) classifier

SVM (Support vector machine) is a supervised machine learning rule. Support Vector Machines [46] map the input vectors nonlinearly into high dimensional feature space. It constructs the optimal hyper planes that maximize the margin of separation between two classes of points.

Let X be a training set which has N features and has class labels $Y \in \{-1, +1\}$ and it is given as

$(y_1, x_1), \dots, (y_l, x_l)$ where $y_i \in \{-1, +1\}$

For a linearly separable dataset there exists a vector \mathbf{w} and scalar \mathbf{b} such that

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b &\geq 0 \text{ if } y_i = +1 \\ \mathbf{w} \cdot \mathbf{x}_i + b &\leq 0 \text{ if } y_i = -1 \end{aligned} \quad (1.8)$$

This above equation. (1.8) can be expressed in generalized form as

$$Y (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 0 \text{ for } i = 1, 2, \dots, N$$

The distance between two hyperplanes is $\frac{2}{\|\mathbf{w}\|}$. To maximize the distance between two classes, \mathbf{w} is minimized by Lagrangian function. If x_{test} is new object then it is classified as given below in equation (1.9)

$$f(x_{test}) = \sum_{i=1}^N \alpha_i w_i (x_i \cdot x) + b \quad (1.9)$$

where α_i is the Lagrange multiplier. Objects in $X \{x_1, x_2, \dots, x_l\}$ whose $\alpha_i \geq 0$ are called support vectors. These support vectors participate for getting optimal hyper plane.

The optimal hyper plane for non-separable data is given in equation (1.10)

$$f(x_{test}) = \sum_{i=1}^N \alpha_i w_i k(x_i \cdot x) + b \quad (1.10)$$

where $k(x_i \cdot x)$ in the above equation is a kernel function and it's equation is given below in equation (1.11)

$$k(x_i \cdot x) = e^{-\gamma \|x_i - x\|^2} \text{ for } \gamma > 0 \quad (1.11)$$

1.8.2 K-Nearest-neighbor (KNN) classifier:

KNN is instance based supervised learning method which is used in many applications such as data mining, medical image processing, face recognition etc. KNN is a simple algorithm that stores all cases and classifies new cases based on similarity measure.

Let the data sample consists of n attributes which combine to form an n -dimensional vector and it is given by equation (1.12)

$$X = (x_1, x_2, x_3, \dots, \dots, x_n) \quad (1.12)$$

There is another variable known as y which depends on n attributes. This y is categorical variable which defines class of vector \mathbf{x} .

Let us consider new sample defined as \mathbf{u} . The following are the steps to identify the class of \mathbf{u} using KNN

- Identify k samples in the training set whose independent variables such as \mathbf{x} are similar to \mathbf{u}
- Find the similarity between new sample \mathbf{u} and k training samples. To measure the similarity between two variables, various distance measures are used such as Euclidean distance, Manhattan distance, Minkowski distance etc. Most popular one is Euclidean distance and is given by equation (1.13)

$$D(\mathbf{x}, \mathbf{u}) = \sqrt{\sum_{i=1}^n (x_i - u_i)^2} \quad (1.13)$$

- Sort the distances and apply the simple majority rule to find the best sample suitable to \mathbf{u} in k samples

However, the classification with KNN depends on selection of k . Large value of k makes the performance of KNN less precise and highly biased. Small value of k makes the performance less stable and it is influenced by noise. The main advantage of KNN is that it is very simple to use and more robust to noise.

1.8.3 Decision Tree Classifier

Decision Tree learning utilizes inductive inference instead of supervised data. It does not require any knowledge of parameter setting and efficient for processing large amount of data. In a decision tree, there are three types of nodes

Root node: It does not have incoming edges and zero or more outgoing edges.

Internal nodes: It has one incoming edge and two or more outgoing edges.

Leaf or terminal nodes: It has exactly one incoming edge and no outgoing edge.

Following are the steps for inducing a decision tree from training samples

Let D be a set containing training samples with different attributes $x = \{x_1, x_2, x_3, \dots, x_c\}$ and class

labels $y = \{y_1, y_2, y_3, \dots, y_c\}$

- Create a root node for the tree
- If all the training sets are of same class y_c , return to the leaf node labeled with class y_c .
- If training set contains more than one class, calculate the entropy $H(s)$ with respect to the attribute and it is given by equation (1.14)

$$H(s) = \sum_{x \in X} p(x) \log_2 \frac{1}{p(x)} \quad (1.14)$$

- Choose the attribute that gives us highest possible information gain (IG) with x as attribute and IG is given in equation (1.15)

$$IG(S, x) = H(s) - \sum_{i=0}^n P(x) * H(x) \quad (1.15)$$

- The highest information gain attribute is removed. This step is repeated to all the attributes until decision tree is left out with leaf nodes.
- Then new sample is submitted to decision tree for further classification

1.8.4 Ensemble methods

Ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions. A necessary and sufficient three issues that are to be addressed by ensemble methods are the learning algorithms must not be statistical, representational and structural. Hence, to overcome these issues three methods can be implemented to construct ensembles

- Bayesian voting
- Bagging
- Boosting

Adaboost and decision tree classifier: Adaboost ensemble classifier has been implemented in this thesis for validation. Boosting is one of the methods used to construct ensemble classifiers. Adaboost algorithm has been utilized with decision tree as weak learner for classification of mammograms.

Boosting is an ensemble method that combines weak learners. Adaboost algorithm is developed by Freund and Schapire. It uses training data for the binary classification.

Let the training data set be $X \rightarrow \{x_i, y_j\}$, $i = 1, 2, \dots, K$ contains K number of instances. Each attribute of X have its class label y_j . The class labels for binary classification is given by $+1$ and -1 . Adaboost uses training data for generating hypothesis.

Following are the steps used by Adaboost algorithm

- The initial weight for all training instance is set to $W(x_i) = 1/K$, where K is total number of training instances
- In each iteration l , training dataset is invoked by decision tree to generate hypothesis h_l . The error rate at each iteration is given by equation (1.16)

$$e_l = \sum_{i: h_l(x_i) \neq y_j} W_l(i) \quad (1.16)$$

Where $h_l(x_i)$ is the hypothesis function and $W_l(i)$ is weight matrix for each instance i

- The weight on the training samples is updated based on the weighted error computed by hypothesis h_l . The weight of training samples is changed based on the misclassification rate, i.e., more weight is applied to misclassified samples and less weight is assigned to correctly classified samples.
- The final classification is obtained by the linear combination of all the hypotheses and its weights. It is given by equation (1.17)

$$h_f(x) = \sum w_l h_l(x) \quad (1.17)$$

1.9 Research Objectives

The prime objective of our research is to reduce interobserver and intraobserver variability among radiologists by increasing the accuracy of detection. Therefore, the objectives of our thesis include

1. Develop image enhancement and segmentation techniques using Gaussian pyramid, wavelet fusion, contrast limited adaptive histogram equalization (CLAHE), morphology, thresholding and arithmetic operations.
2. Develop feature extraction algorithms using Local binary pattern (LBP) and its variants, Segmentation based fractal texture analysis (SFTA) and Rotation invariant local frequency (RILF) techniques to classify ROIs into normal and abnormal.
3. Develop shape based features algorithms to classify mammograms into benign-malignant.
4. Develop an automatic method that detects suspicious areas in mammogram using Dunn index based automated k-means algorithm (DAK) without any user intervention. It also analyses the detected area as mass-non-mass and benign-malignant.

1.10 Organization of thesis

This thesis is organized as follows. Chapter 2 introduces a technique for image enhancement and image segmentation. In chapter 3, different feature extraction techniques were implemented on breast regions to classify them into normal-abnormal. In chapter 4, contour based features were extracted to classify mammogram masses into benign and malignant. Chapter 5 gives quantitative results of the developed algorithms. It also describes comparison of algorithms that have been developed with the already existing methods. In chapter 6, automatic method without manual intervention is developed for detection and two stage classification of abnormalities in mammograms. This method was developed without any manual extraction of ROIs. Chapter 7 presents the summary and conclusions of the proposed work and also gives the future scope of the work.

Chapter 2

Image enhancement and segmentation of mammograms

The leading cause of deaths among women is due to breast cancer. About 10% of women are affected with breast cancer in their lives. Early detection and diagnosis of breast cancer using digital mammography and image processing can increase survival rate and chances for complete recovery of patients. Several image processing techniques have been developed to improve the detection of four abnormal features in breast mammograms. The four abnormal features include:

1. Microcalcifications 2. Masses 3. Bilateral Asymmetry 4. Architectural distortion.

Microcalcifications are small calcium deposits on mammogram with dimensions of 0.2 to 0.3mm. The presence of aligned or clustered microcalcifications in mammograms indicates breast cancer [47-48]. Masses are the most common asymmetric signs of cancer and appear brighter than the surrounding tissue in mammograms [49]. Most benign masses possess well-defined sharp borders, while malignant tumors often have ill-defined, microlobulated, or spiculated borders. Bilateral asymmetry is an asymmetry of the breast parenchyma between left and right breast which may indicate breast cancer in its early stage. An architectural distortion on a mammogram is basically a disruption of the normal 'random' pattern of curvilinear and fine linear radiopaque structures. There is no visible mass, but the distortion often appears as a 'stellate' shape or with radiating speculation [50]. Detection of these abnormalities using CAD helps radiologists to diagnose breast cancer in early stages.

Among all image processing techniques of CAD, image enhancement plays vital role, since it increases contrast between malignant tissue and normal dense tissue which increases the

visibility for human perception. In this chapter, we developed a new image enhancement algorithm using Gaussian pyramid, wavelets and contrast limit adaptive histogram equalization. Further, we developed a new algorithm for segmentation along with image enhancement.

This chapter is organized as follows: The databases used for the evaluation is described in section 2.1. The theory of 2D wavelets, Gaussian pyramid and contrast limit adaptive histogram equalization (CLAHE) methods are explained in sections 2.2, 2.3 and 2.4 respectively. Mammogram enhancement using the developed algorithm is described in section 2.5 and in section 2.6, and the developed algorithm for image segmentation has been discussed. This section describes the process of segmentation of mass from mammogram using thresholding and morphological operations. Further, section 2.7 gives the summary of overall work proposed in this chapter.

2.1 Database

To validate the image enhancement algorithm, mammograms are taken from Krishna Institute of medical sciences (KIMS), Hyderabad and DDSM database. The algorithm was used to test on 23 mammograms out of which 14 were taken from KIMS and 9 mammograms from DDSM database. The image segmentation was carried out using 17 mammograms with the mass taken from DDSM database. In order to validate the segmentation methods, the database provided the ground truth information of mass, based on the location of abnormality in the mammogram.

2.2 Multiresolution analysis using 2D wavelets and wavelet fusion

The decomposition of mammographic images was done using two dimensional discrete wavelet transform (2D-DWT). The result gives two types of coefficients: approximation coefficients and detail coefficients [50]. Approximation coefficients are obtained by low pass filtering the image

followed by down sampling. Detail coefficients are obtained by high pass filtering the image followed by down sampling. The resolution of images varies with each decomposition level.

The resolution decreases with the increase in number of decomposition levels. DWT decomposes mammogram into LL, LH, HL and HH components. This property of wavelet helps to extract edges, noise and separate bright and low contrast regions from mammogram. These regions can individually be emphasized or removed from the mammogram by choosing level of decomposition and selected component

2D-DWT requires a two dimensional scaling function $\varphi(x, y)$ and three two-dimensional wavelets, $\phi^H(x, y)$, $\phi^V(x, y)$ and $\phi^D(x, y)$ which gives horizontal, vertical and detail coefficients of the image. These wavelet functions are separable. The discrete wavelet transform of image $f(x, y)$ of size $M \times N$ is given by the equations (2.1) and (2.2)

$$W_{\varphi}(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j_0, m, n}(x, y) \quad (2.1)$$

$$W_{\phi}^i(j, x, y) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \phi_{j, m, n}^i(x, y) \quad i = \{H, V, D\} \quad (2.2)$$

where j_0 is an arbitrary starting scale, $W_{\varphi}(j_0, m, n)$ coefficients define an approximation of $f(x, y)$ at scale j_0 , $W_{\phi}^i(j_0, x, y)$ coefficients define horizontal, vertical and diagonal details for scales $j > j_0$. The value of j_0 is selected as 0 and $N=M=2^J$ so that $j=0,1,2,3,\dots,J-1$ and $m=n=0,1,2,\dots,2^j - 1$ and inverse discrete wavelet transform is applied to get $f(x, y)$

A Two-Dimensional Discrete Wavelet Transform is implemented using the combination of digital filter banks and down-samplers. The digital filter banks consist of high-pass (g) and low-pass (h) filters. As the image is 2D signal, 1D wavelet transform is evaluated for both the rows and columns of the image with separable wavelet functions. Thus, the output of 2D-DWT of an

image results into four orthogonal components like Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH).

2.3 Gaussian Pyramid

Gaussian Pyramid is based on the principle of image pyramid proposed by Burt and Adelson in 1983 [51]. This method decomposes the original image into sub-images with different spatial resolutions through mathematical operations such as recursive low pass filtering and interpolation [52].

Let G_0 is a source image, and is on the bottom of the pyramid. Let l -th level Gaussian pyramid is denoted by G_l and it is given by the equation (2.3)

$$G_l(i, j) = \sum_{m=-2}^2 \sum_{n=-2}^2 \omega(m, n) G_{l-1}(2i + m, 2j + n) \quad 1 \leq l \leq N, 0 \leq i \leq R_b, 0 \leq j < C_l \quad (2.3)$$

where N is the maximal level of pyramid, C_l and R_b represent the column and row number of l -th level pyramid respectively. $\omega(m, n)$ is a Gaussian kernel of any window size G_{l-1} is obtained using *Reduce* operation and it is given by the equation (2.4) .

$$G_l = \text{Reduce}(G_{l-1}) \quad (2.4)$$

Where *Reduce* operation indicates down sampling the image. $G_0, G_1 \dots \dots G_N$ are different levels of Gaussian pyramid, where G_N is at the top of the pyramid. The total number of layers in pyramid is $N+1$.

2.4 Contrast Limit Adaptive Histogram Equalization (CLAHE)

Histogram of an image plays an important role in image enhancement techniques. The objective of histogram equalization (HE) is to enhance the global contrast of the original image with uniform histogram. In case of histogram equalization, the transformation function is applied to input image by mapping pixel values such that the enhanced image have uniform histogram.

Histogram of the digital image of size $M*N$ with intensity values in the range 0 to $L-1$ is given by the equation (2.5)

$$h(r_k) = n_k \quad (2.5)$$

Where r_k is the k^{th} intensity value and n_k is the number of pixels in the image with intensity r_k [52]. A normalized histogram is given by equation (2.6)

$$p_r(r_k) = \frac{n_k}{MN} \quad k = 0,1,2,3 \dots L - 1 \quad (2.6)$$

The histogram equalization is obtained by following equation (2.7)

$$s_k = (L - 1) \sum_{j=0}^k p_r(r_j) \quad k = 0,1,2, \dots \dots L - 1 \quad (2.7)$$

where s_k is the cumulative distribution function (CDF). Histogram equalization method cannot resolve the problem of global change in pixel values, when grayscale distribution is highly localized.

Hence, Adaptive histogram equalization method was derived by transforming each pixel with a transformation function derived from neighborhood. In this method, the image is divided into non overlapping block and CDF is evaluated for each block to increase the contrast of each region But, this method over amplifies the contrast in constant intensity and noise regions. Hence, the Contrast Limited Adaptive Histogram Equalization (CLAHE) was proposed to overcome the problem of noise amplification. In CLAHE, the histogram is clipped by a predefined factor known as clip limit before computing CDF in the neighborhood region of the pixel. This clip limit depends on the normalized histogram and size of the neighborhood region and it is given in equation (2.8).

$$\beta = \frac{MN}{L} \left(1 + \frac{\alpha}{100} (s_{max} - 1)\right) \quad (2.8)$$

Where α is a clip factor (ranges from 0 to 1) and s_{max} is the max slope

Finally, CDF of the resultant histograms of all blocks are obtained for mapping pixel values.

2.5 Algorithm for Image enhancement

Gaussian pyramid based wavelet fusion is applied to mammograms for image enhancement. First Gaussian pyramid is generated for two levels. Level-0 is an original image which is filtered using Gaussian filter and down sampled to the half of the original image resolution to obtain level-1 image. This process highlights larger objects present in the mammogram. Then, the level-1 image is filtered by Gaussian filter to remove noise and up scaled using cubic spline interpolation. After this, two-step process has been implemented: first one is, the original image (level 0) is fused with up scaled version of level 1 image using max-max wavelet fusion technique. This wavelet fusion helps to highlight larger and smaller objects as the images are fused in both spatial and frequency domain. The output image obtained by step 1 is further enhanced by adaptive histogram equalization technique to increase the contrast. The performance of this method is determined using two parameters like edge based enhancement measure (EBCM) and entropy. The block diagram of the scheme is shown in the Figure 2.1 and the output of image enhancement is shown in the Figure 2.2.

2.6 Image Segmentation

Image segmentation plays a vital role in extracting the abnormality in mammogram. After successful segmentation, features are extracted using different methods. Then, these features are given to classifier. In literature, many methods have been proposed and few of them are listed as follows: active contour models [53], watershed transform[21], Convolution neural networks (CNN) and geostatistics [54], support vector machine and scale-variant mean curvature measures [55], graph based methods[56], local correntropy, k-means clustering[57] and discrete wavelet transform [58].

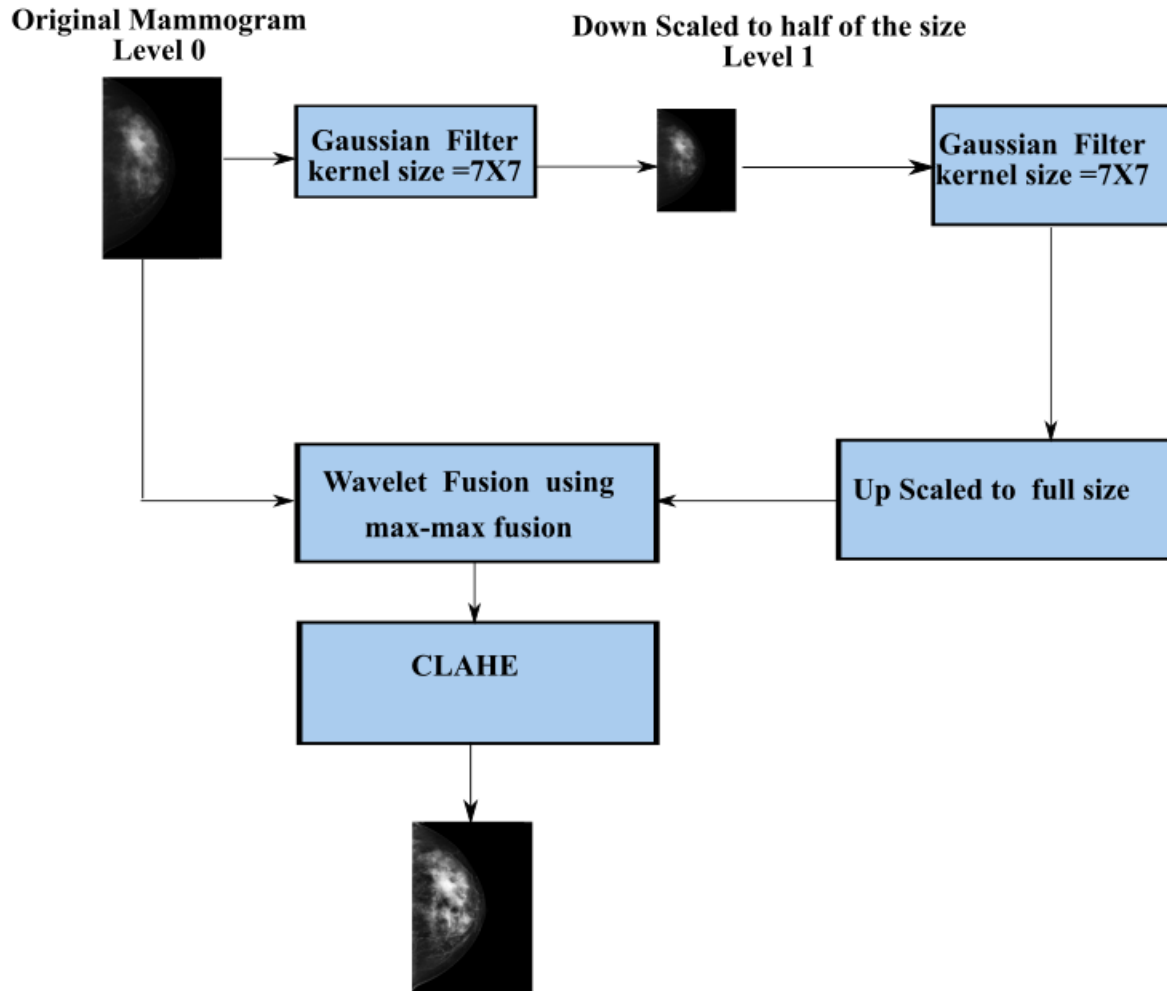


Figure 2.1: Block diagram of image enhancement

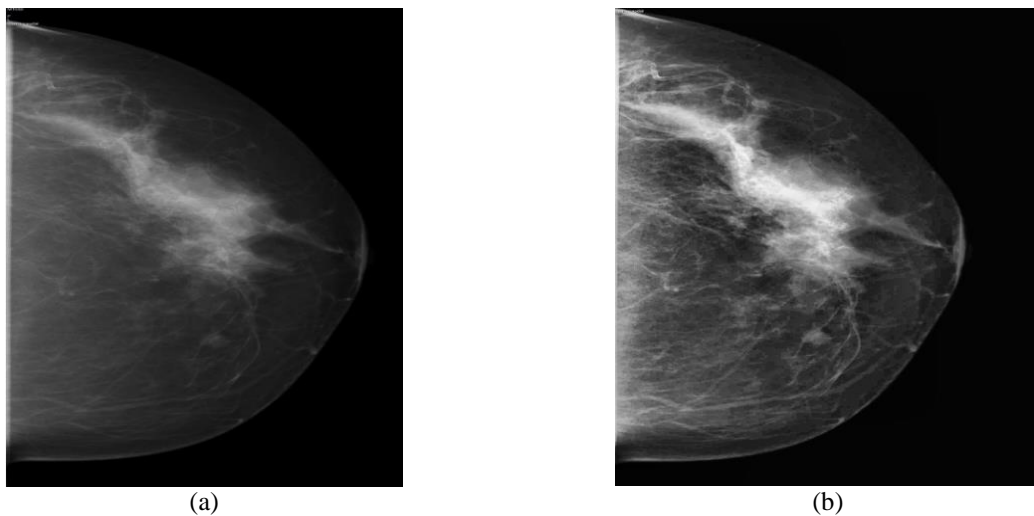


Figure 2.2: (a)Original mammogram image , (b) proposed method enhanced image

In all the above techniques, segmentation was carried out by considering a region of interest with mass in the mammogram, whereas in this present work, automatic segmentation has been carried out on the complete mammogram using histogram and advanced morphological operations. The process of segmentation method proposed in the present work is discussed below.

2.6.1 Algorithm for segmentation

We have used 17 mammograms with mass acquired from DDSM database for segmentation. Among them, six mammograms contain benign mass and eleven mammograms contain malignant mass. To validate our segmentation method, we have used the ground truth information of masses provided by radiologists. The simulation process for segmentation is carried out using NI vision assistant. Block diagram of segmentation process is shown in the Figure 2.3 and the following are the steps involved in the process of segmentation

Step 1: Mammogram is acquired using NI vision assistant. Then, the preprocessing of mammograms has been carried out for the extraction of breast area and to remove the noise and unwanted labels in mammogram. The original mammogram taken from DDSM is shown in Figure 2.4a

Step 2: In this step, bright and dark mammogram tissues are delineated using thresholding. The gray level value required for thresholding is obtained using histogram. The thresholding process converts gray scale image into binary image. The output obtained after thresholding is shown in Figure 2.4b. This image consists of false positive regions.

Step 3: To detect the abnormality from false positive regions of thresholded image, morphological operations have been applied.

Step 4: The masses in the DDSM database are with in radius of 3mm to 40mm [58]. Hence, morphological filters are applied to the binary image, based on the area and Heywood circularity

factor of regions (mammogram tissues) present in binary image .The mammogram tissue regions with radius less than 3mm and greater than 40mm are separated out using Top-Hat morphological transform. The mammogram tissue regions with circularity factor greater than two are filtered. The circular factor is taken into consideration because the mass is more circular than non-mass regions and the value of circularity factor is approximately one [59]. The output of segmentation is shown in Figure 2.4c.

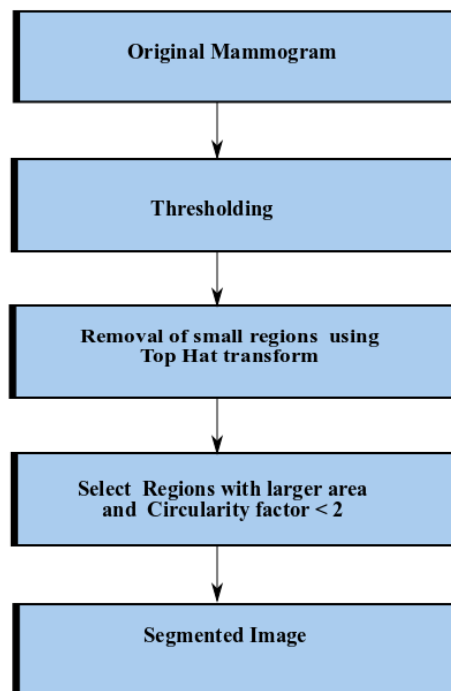


Figure 2.3: Block diagram of segmentation process

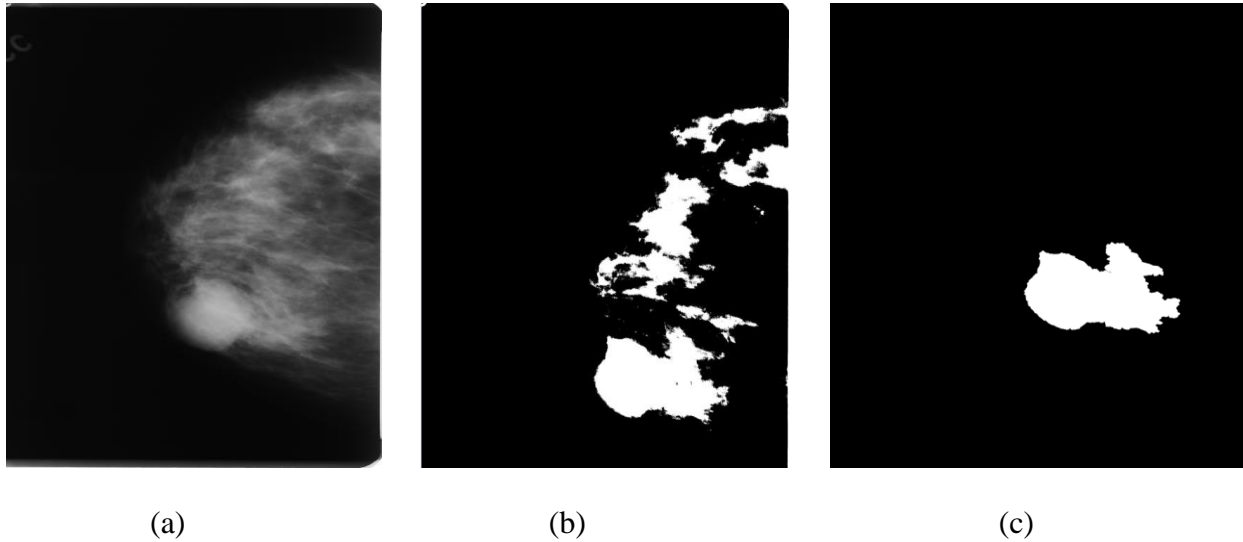


Figure 2.4: a) Original mammogram, (b) Output after thresholding, (c) Segmented image

2.7 Summary

In summary, algorithms for image enhancement and segmentation of mammograms have been developed. The enhancement algorithm uses wavelet fusion to highlight smaller and larger objects in mammograms and CLAHE is implemented to increase the contrast. This approach has been tested on 23 mammograms taken from two databases. Image segmentation was carried out using thresholding and morphological operations. Algorithm developed for segmentation is tested on 17 mammograms that are taken from DDSM database. The ground truth information of mass is given by expert radiologists in DDSM database for validation of image segmentation algorithm. The results of image enhancement and image segmentation algorithms are discussed in chapter 5.

Chapter 3

Mammogram classification to normal and abnormal using texture features

Among all the other image processing techniques for CAD, feature extraction is an important task. This process should properly delineate normal and abnormal areas in mammograms and the number of the extracted features should be less so that the CAD system is computationally efficient. In this chapter, we have extracted features from mammogram regions using Rotation Invariant Local Frequency (RILF), Local Binary Pattern (LBP) and Segmentation based fractal texture analysis (SFTA) techniques followed by classification. Further, these techniques have been predominantly used in real world texture analysis problems. In order to extract features the suspicious abnormality areas were cropped from mammograms and these areas are called region of Interest (ROIs). Three methods have been implemented for extraction of features. First method uses RILF, LBP and SFTA techniques on ROIs to extract features, RILF is more robust to noise and has relatively small number of features when compared to LBP and SFTA. Hence, we have made the first attempt to study the use of RILF technique in mammograms for abnormality detection. In second method, LBP, SFTA and RILF features are extracted from each block after dividing ROI into $N \times N$ blocks and in third method ROI is decomposed into two levels using Gaussian pyramid and further features extracted using LBP, RILF and SFTA from decomposed images. To evaluate the performance of these three methods, the extracted features from ROIs are given to SVM classifier to classify them into normal-abnormal.

The present chapter is organized as follows: The extraction of ROI from the mammographic image is described in section 3.1. The description of LBP and its variants, SFTA and RILF feature extraction techniques are given in sections 3.2, 3.3 and 3.4 respectively. Section 3.5 gives the information about implementation of extracted features using all the above three techniques and section 3.6 describes about the classification. The summary of our work is explained in section.3.7.

3.1 Extraction of ROI

The breast regions of mammograms contain unwanted labels and background which are not desirable for the analysis of texture. Hence, mammogram is cropped to extract Region of Interest (ROI) which has suspicious abnormality. However, IRMA database consists of normal and abnormal ROIs of resolution 128*128.

3.2 Local Binary Pattern (LBP) and its variants

3.2.1 LBP

Local binary pattern is an efficient rotation invariant technique for texture analysis. LBP has been presented by many authors in many applications because of its computational efficiency, high discriminative power and less vulnerability to illumination changes. This technique is first introduced by Ojala [59] considering statistics of gray level differences. In this technique, LBP operator assigns the binary code i.e., 0 or 1 to each pixel based on gray level difference of that particular pixel and its neighborhood. The neighboring pixel values which are not at the center are obtained by interpolation. For a center pixel g_c , LBP code is defined by the following equation (3.1) as given below

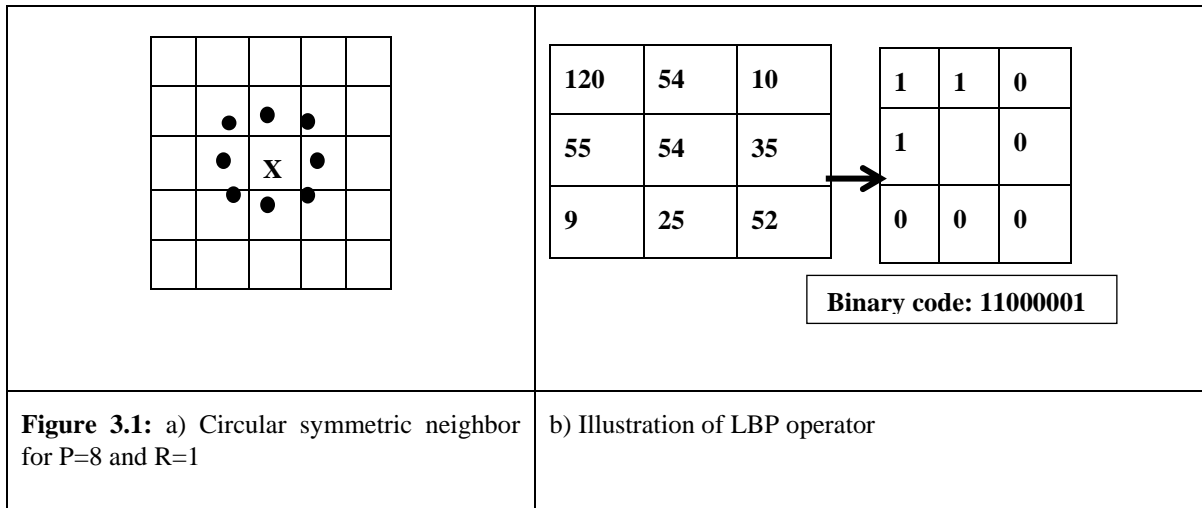
$$LBP_{(P,R)} = \sum_{p=0}^{p=P-1} s(g_p - g_c) 2^p \quad (3.1)$$

$$\begin{aligned} \text{where } s(x) &= 1 \text{ for, } x \geq 0 \\ &= 0 \text{ for, } x < 0 \end{aligned}$$

In equation 3.1, g_c represents the center pixel and g_0, g_1, \dots, g_{P-1} represents the neighboring pixels, where P corresponds to quantization of angular space (number of neighbors) and R (Radius of the neighborhood) corresponds to spatial resolution. If the center pixel g_c have coordinates $(0,0)$, then the exact coordinates of neighbor pixels are given by $(-R \sin(\frac{2\pi p}{P}), R \cos(\frac{2\pi p}{P}))$. Now the LBP descriptors are calculated from the histogram $H(k)$ of labeled image obtained by applying $LBP_{(P,R)}$ on all the pixels of image. The histogram $H(k)$ for $k=1,2,3 \dots K$ for the image of size $M \times N$ is given below in equation (3.2).

$$H(k) = \sum_{i=1}^M \sum_{j=1}^N I(LBP_{(P,R)}(i,j)) \quad (3.2)$$

where, K is the maximum label value in the $LBP_{(P,R)}$ image. Figure 3.1a shows the circularity symmetric neighbor sets for $P=8$ and $R=1$ and Figure 3.1b shows illustration of LBP operator.



There are many variants of LBP based on encodings, shape of neighborhood and transitions in binary code generated for each pixel. LBP variants which are based on transitions and rotation invariance include uniform LBP, rotation invariant LBP, rotation invariant uniform LBP. The LBP variants which are based on shape of neighborhood include EBP (Elliptical binary pattern) and HBP (Hyperbolic binary pattern), and these variants are considered to detect anisotropic features [60]. The LBP variants which are based on encoding include Local ternary pattern (LTP) and Local quinary pattern (LQP). The features are calculated using LBP variants based on encodings (LTP, LQP) and transitions in binary code (uniform, rotational invariant and uniform rotational invariant). LBP variants based on encodings are discussed below.

3.2.2 Local Ternary pattern

In LTP variant, the difference between the gray value of center pixel g_c and gray value of neighborhood pixel g_p is encoded into three values with the user threshold (τ) and they are $\{1, 0, -1\}$. LTP coding for a given pixel is given below in equation (3.3)

$$LTP_{(P, R)} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (3.3)$$

$$\begin{aligned} \text{Where } s(x) &= 1 \text{ for } x \geq \tau \\ &= 0 \text{ for } -\tau < x < \tau \\ &= -1 \text{ otherwise} \end{aligned}$$

3.2.3 Local Quinary pattern

The difference between the center pixel and the neighborhood pixel is encoded into five values $\{2, 1, 0, -1, -2\}$ with user thresholds (τ_1, τ_2). LQP coding for a pixel is given by equation (3.4)

$$\text{LQP}_{(P,R)} = \sum_{p=0}^{p=P-1} s(g_p - g_c) 2^p \quad (3.4)$$

$$\begin{aligned} s(x) &= 2 \text{ for } x \geq \tau \\ &= 1 \text{ for } \tau_1 < x < \tau_2 \\ &= 0 \text{ for } -\tau_1 \geq x < \tau_1 \\ &= -1 \text{ for } -\tau_2 \leq x < -\tau_1 \\ &= -2 \text{ for other wise} \end{aligned}$$

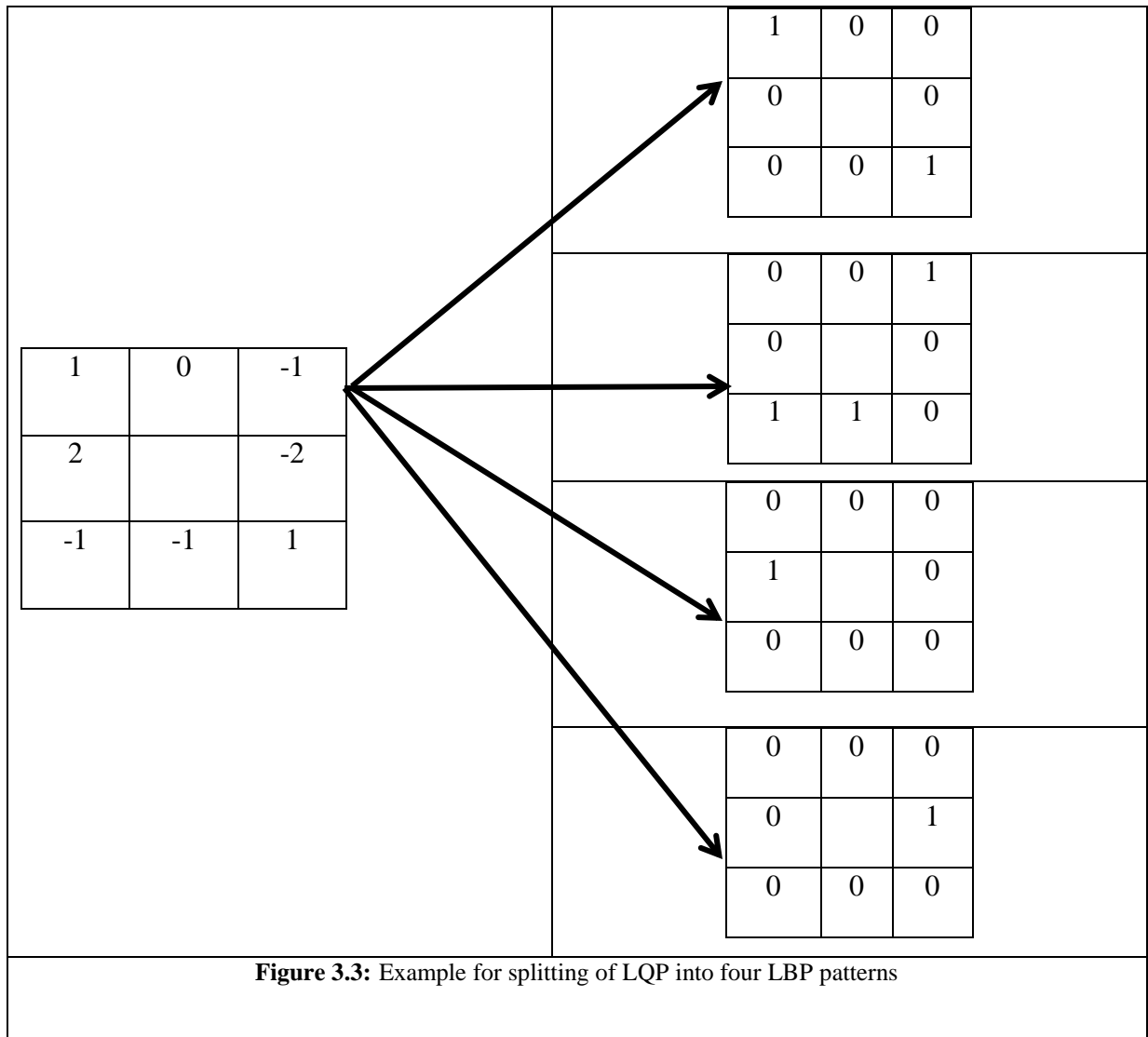
The local ternary pattern is split into two binary patterns by using three codes {1, 0,-1} and it is shown in Figure 3.2. First and second binary patterns are generated by considering the code =1 and -1 respectively. Further the histogram generated by this two LBP patterns are concatenated to form feature vector. In the same way, LQP is split into four LBP patterns and the histogram features of these four pattern forms feature vector and it is shown in Figure 3.3.

1	0	-1
1		-1
-1	-1	-1

1	0	0
1		0
0	0	0

0	0	1
0		1
1	1	1

Figure 3.2: Splitting of LTP pattern into two different LBP codes



3.3 Segmented Fractal Texture analysis (SFTA)

SFTA method [34] involves two steps: In first step, the input gray level images are decomposed into a set of binary images using Two-Threshold Binary Decomposition (TTBD). In second step, features like fractal dimensions, mean and size are computed for each binary image.

In step 1, TTBD takes input as gray level image $I(x, y)$ and multilevel Otsu's algorithm is implemented to obtain binary images. This algorithm calculates a set of threshold values (T)

based on minimization of intra-class variance and is applied to each image recursively until the desired number of threshold values (n_t) is reached. Initially, first set of binary images are obtained using the thresholds calculated from Otsu's algorithm. Hence, the total number of binary images is n_t in the first set. In the second set, the binary images are obtained by selecting pairs of thresholds from T . The gray level image $I(x, y)$ is then decomposed into binary images applying adjacent thresholds from $T = \{t_i, n_i\}$ where n_i is maximum gray level intensity in the image $I(x, y)$ and t_i is set of thresholds obtained from Otsu's algorithm and given by $t_1 = \{t_1, t_2, t_3, \dots, t_{n_t}\}$ First set of binary images $I_{b1}(x, y)$ is given by

$$I_{b1}(x, y) = \begin{cases} I & \text{if } I(x, y) > t_i \in T \\ 0 & \text{otherwise} \end{cases} \quad (3.5)$$

Second set of binary images are computed from pairs of thresholds, ($T = \{t_1, \dots, t_{n_t}, n_i\}$) and they are given as follows

$$I_{b2}(x, y) = \begin{cases} = 1 & \text{if } t_{lo} < I(x, y) \leq t_u \\ = 0 & \text{otherwise} \end{cases} \quad (3.6)$$

where t_{lo} and t_u are upper and lower thresholds.

In step 2, SFTA feature vector which includes mean of the gray level values, number of pixels and fractal dimension from the boundary are calculated from binary images obtained from step 1.

3.4 Rotational Invariant Local Frequency (RILF)

RILF method [61] is a rotation invariant texture classification method based on local frequency components. In this method, local circular function $LCF_{(N,R)}(x, y)$ is defined at each pixel to

obtain features. This local circular function considers N neighboring samples on a circle at each pixel (x, y) with radius R as in LBP. In LBP, thresholding the neighborhood pixels loses some textural information. To overcome this problem 1D Fourier transform is applied on $LCF_{(N,R)}(x, y) = (t_0, t_1, t_2, \dots, t_{N-1})$ and is given below in the equation (3.7)

$$f_n = \sum_{k=0}^{N-1} t_k e^{-\frac{2\pi i k(n-1)}{N}}, \quad (n = 1 \text{ and } 2) \quad (3.7)$$

The low frequency components f_1 and f_2 are considered to obtain features [62]. The reason for considering only the low frequency components f_1 and f_2 is that they have 90% of the texture energy and they do not contain noise information which appears in high frequencies [62]. Further, the magnitude based features are extracted from these two frequency channels (f_1 and f_2) by computing circularly shifted 2D Fourier transform. The 2D spectrum of each frequency channel CH_n , (where $n=1$ and 2) is given in the equation (3.8)

$$CH_n(k, l) = \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |f_n(x, y)| \cdot e^{-2\pi i (\frac{xk}{W} + \frac{yl}{H})} \quad (3.8)$$

where, W and H are width and height of the image. To have rotation invariance property for these features, circular band-pass disk shape filters are applied on the spectrum of frequency channels and these filters are defined below in equation (3.9)

$$D_{r_1, r_2}(x, y) \begin{cases} = 1 & \text{if } r_1 \leq \sqrt{x^2 + y^2} \leq r_2 \\ = 0 & \text{other wise} \end{cases} \quad (3.9)$$

where r_1 and r_2 are inner and outer radius of disk shape filters. The rotation invariant local frequency magnitude descriptors (RILFMD) are computed in equation (3.10) which is given below

$$RILFMD(r1, r2, n) = \frac{\sum_{k=-W/2}^{W/2-1} \sum_{l=-H/2}^{H/2-1} |CH_n(k,l) \cdot D_{r1,r2}(k,l)|}{\sum_{k=-W/2}^{W/2-1} \sum_{l=-H/2}^{H/2-1} D_{r1,r2}(k,l)} \quad (3.10)$$

The range of 2D spectrum changes from [0: W-1, 0: H-1] to [-W/2: W/2-1, -H/2: H/2-1] when the spectrum is circularly shifted.

3.5 Implementation of Feature Extraction techniques

The different feature extraction techniques discussed in the above sections 3.2, 3.3 and 3.4 are applied on mammogram ROIs using three different methods. The extracted features from these three methods are fed to a SVM classifier (linear kernel) via ten-fold cross-validation method to assess the significance of these feature extraction techniques.

3.5.1 Method 1

In this method, the features are extracted from ROIs using LBP, SFTA and RILF techniques and these features are fed to SVM classifier for further classification. The block diagram for this method is shown in Figure.3.4.

LBP: In LBP, different variants based on encodings (LTP and LQP) and transitions in binary code (uniform (u2), rotational invariant (ri) and uniform rotational invariant (riu2)) are considered for feature extraction. The histograms of all these variants are taken as feature vectors for further performance evaluation.

SFTA: In SFTA technique, texture features have been extracted from ROIs by considering different values of n_t (number of thresholds). These features are further fed to SVM classifier for performance evaluation.

RILF: In RILF technique, the magnitude descriptors (RILFMD) are calculated from the low frequency channels that are channel 1(ch1) and channel 2(ch2). These magnitude descriptors are considered as feature vectors and passed to SVM classifier for performance evaluation.

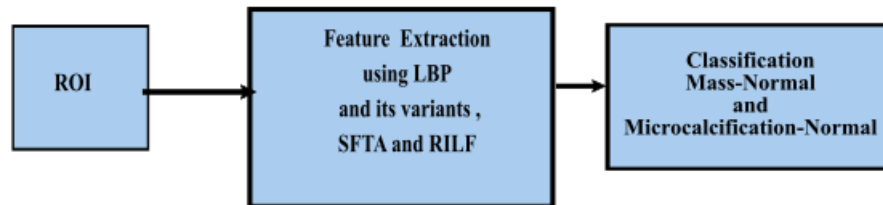


Figure 3.4: Block diagram for the classification of ROIs using method 1

3.5.2 Method 2

In this method, ROI is sub divided into NXN blocks. LBP and its variants (LQP and LTP), SFTA and RILF features are extracted for each block and concatenated to form feature vector and fed to SVM classifier for further validation. The block diagram of this method is shown in Figure 3.5. The feature extraction techniques used in this method are named as block wise-LBP (BLBP), block wise-LTP(BLTP), block wise-LQP (BLQP), block wise-SFTA(BSFTA) and block wise-RILF (BRILF) techniques.

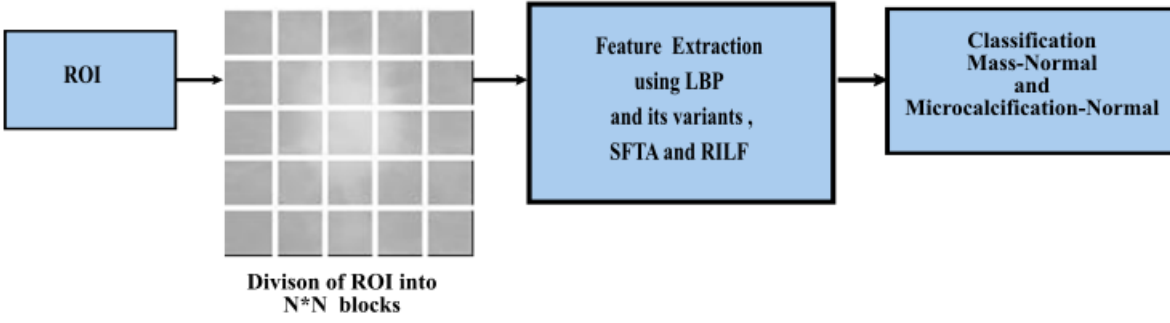


Figure 3.5: Block diagram for the classification of ROIs using method 2

3.5.3 Method 3

In this method, ROI is decomposed into image pyramid via Gaussian filter [62] which is also known as scale space representation of the image. The scale space equation of an image is given below in equation (3.11)

$$\begin{aligned}
 G_k(x, y) &= mnW(m, n)G_{k-1}(2x + m, 2y + n) \text{ for } k > 0 \\
 &= I \text{ for } k = 0
 \end{aligned} \quad (3.11)$$

where x and y are spatial co-ordinates of image I , $G_k(x, y)$ represents k^{th} level image pyramid and $W(m, n)$ is a pyramid filter (Gaussian) of size $m \times n$.

LBP and its variants, SFTA and RILF techniques are used to extract features from two levels of Gaussian pyramid and further concatenated to form feature vector for evaluation. The block diagram of method 3 is shown in Figure 3.6 and these techniques are named as pyramidal-LBP (PLBP), pyramidal-LTP (PLTP), pyramidal-LQP (PLQP), pyramidal-SFTA (PSFTA) and pyramidal-RILF (PRILF).

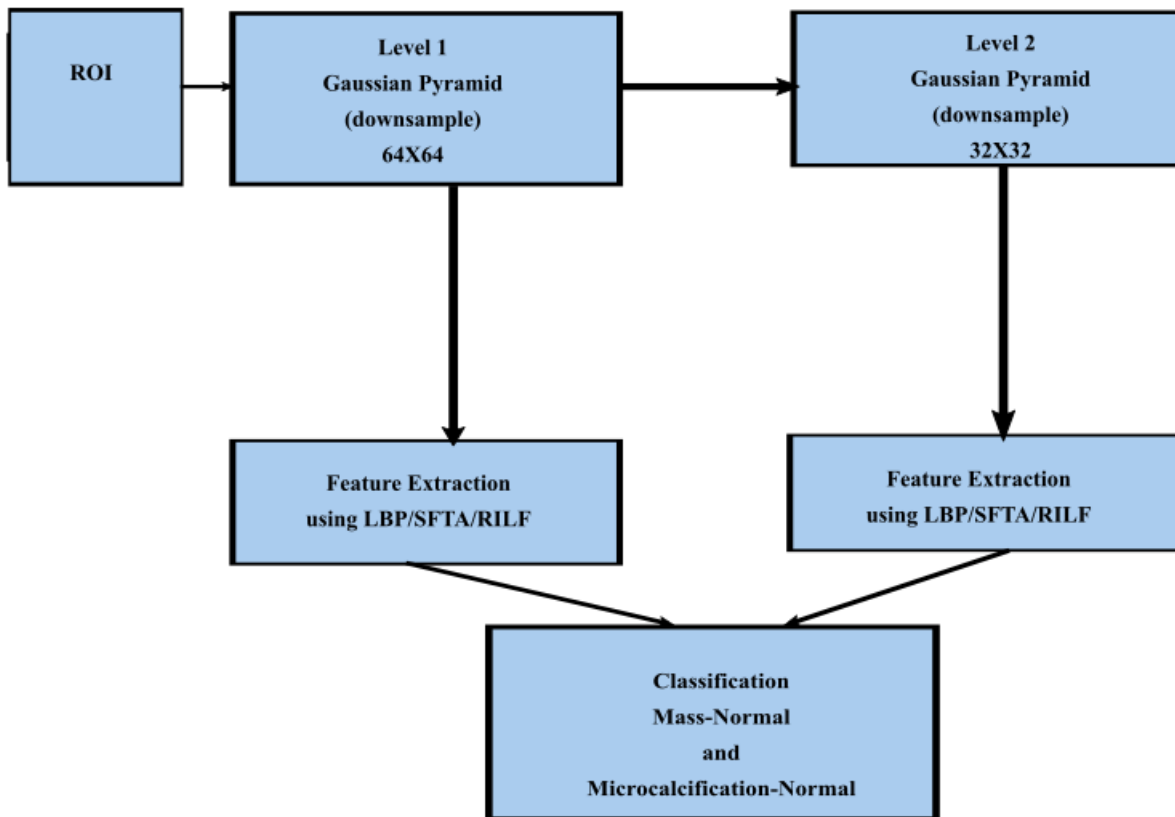


Figure 3.6: Block diagram for the classification of ROIs using method 3

3.6 Classification

Evaluation of our study is carried out using SVM classifier via ten-fold cross validation. In this validation method, feature data is divided into ten groups randomly. Among these groups, one group is selected for testing and other groups are for training. In this procedure each group will undergo testing and the mean of all evaluation parameters from ten groups is calculated to get final accuracy, sensitivity and specificity.

3.7 Summary

In this chapter, LBP, SFTA, and RILF feature extraction techniques are used to distinguish normal and abnormal (mass and microcalcifications) ROIs using three different methods. Among these feature extraction techniques, RILF works in frequency domain and it does not consider noise part of the image. This property aids RILF features to get high classification accuracy when compared to LBP and SFTA techniques. In summary, RILF has given best accuracy, sensitivity and specificity when compared to other techniques. The results obtained are discussed in chapter 5.

Chapter 4

Mammogram classification to benign and malignant using contour based features

Masses are the most common abnormality present in breast cancer patients. It is formed when healthy cells grow abnormally at one place. Masses can be characterized by various descriptors like shape, size and margins according to breast imaging reporting and data system (BI-RADS). Benign (non-cancerous) masses usually have oval, lobulated, and round shapes with smooth circumscribed margins but whereas malignant (cancerous) masses have irregular, ill-defined and microlobulated shapes with spiculated margins. We can classify breast masses as benign or malignant based on shape characteristics. The mass is differentiated by its gray scale, margin and its shape characteristics. Many texture based methods were also introduced to classify masses as benign and malignant. Gabor transforms, spherical wavelet transforms, contourlet transforms, wavelet transforms, local energy based descriptors, beamlet transforms and fractal dimensions are some proposed texture methods in the literature for classification of breast masses [63-71]. These texture methods generates high dimensional feature vector which increases the computational complexity of CAD (Computer aided diagnosis) system. In order to decrease the length of feature vector, many methods have been introduced for classifying breast masses using 2D contour or shape characteristics of masses. In this chapter we have studied contour based methods to classify masses into benign and malignant. The features are extracted from contours

using geometry properties, polynomial regression, beam angle statistics and triangular area representation. Subsequently, these features are given to different classifiers for validation.

The chapter is organized as follows: The extraction of mass contour from mammogram using ground truth information is described in section 4.1. Section 4.2 outlines about the features extracted from mass contours using the geometric 2D shapes. The procedure for feature extraction from mass contours using polynomial regression, beam angle statistics (BAS) and triangular area representation (TAR) are described in sections 4.3, 4.4 and 4.5. The classification of these mass contours using different classifier is outlined in section 4.6 followed by summary of the proposed work in section 4.7.

4.1 Extraction of mass contours

The mammography images used in this work are taken from “Digital Database for Screening Mammography” (DDSM), a publically available database. The database has overlay file for each patient. The information related to particular patient such as type of abnormality, number of abnormalities and its location is in overlay file. The outline of mass contour is given by chain code in overlay file. The database contains normal, benign and cancer volumes. For our research, we require benign and malignant mass contours.

4.2 Geometric features

Geometric features of mass contour can be used to characterize malignant or benign mass. Six morphologic features were extracted from each mass to describe features such as shape, contour, and size as in [72-74]. The following are the geometric features used on the mass contours.

Area (A): Number of pixels contained inside the mass contour.

Perimeter (P): Circumference of the mass contour.

P: A: It is the ratio of perimeter to the area of the mass.

Circularity Index: Circularity index is the ratio of the area of a mass to the area of a circle having the same perimeter. It is shown below

$$C = \frac{A_s}{A_c}$$

Where A_s is the area of the mass and A_c is the area of the circle.

L: S Ratio: It is the length ratio of the major (long) axis to the minor (short) axis of the equivalent ellipse of the mass. If L: S ratio is more, it is likely that the mass is malignant.

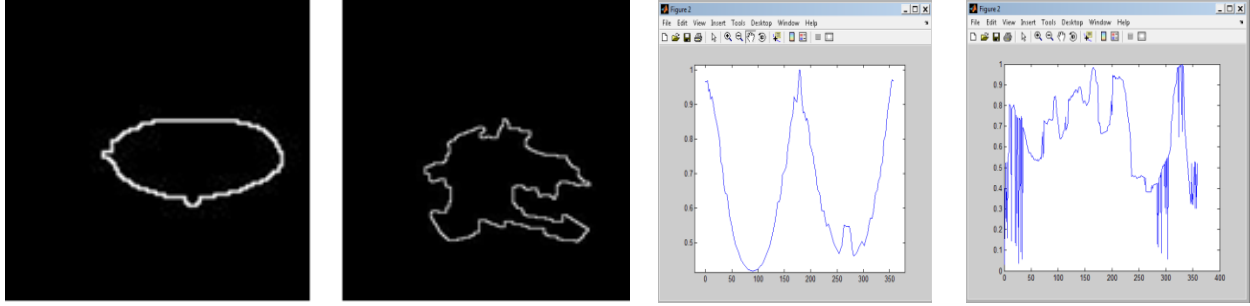
E: N (Elliptical normalized circumference): Anfractuosity is a common morphological feature for malignant contour. ENC is circumference ratio of the mass and its equivalent ellipse. Anfractuosity of a mass contour is characterized by ENC.

$$ENC = \frac{A_s}{A_E}$$

where A_s is area of the mass and A_E is the area of its ellipse .

4.3 Polynomial regression

A polynomial regression fit by gradient descent algorithm is used to extract features from 1D signal. Hence, mass contour is transformed into 1D representation of boundary known as signature. It is a plot of the distance from the centroid to the boundary as a function of angle. Figures 4.1a and 4.1b. shows the contours of benign and malignant mass respectively. Figures 4.1c and 4.1d show the signatures of these contours.



(a) Benign, (b) Malignant, (c) Signature of benign, (d) Signature of malignant

4.3.1 Polynomial regression fit using gradient descent algorithm

In this algorithm n^{th} degree polynomial that fits signature of mass contour is generated by considering polynomial hypothesis with n random coefficients. The polynomial hypothesis considered with n random coefficients is given by the below equation (4.1)

$$H(\theta) = a_0 + \sum_{n=1}^{\infty} (a_n \theta^n) \quad (4.1)$$

where θ is the angle measured from the vertical direction in the anti-clockwise direction and $a_0, a_1, a_2, a_3, \dots, a_n$ are the coefficients of the n^{th} order polynomial hypothesis. The degree of polynomial is to be chosen such that the signature of benign contour is close to polynomial hypothesis. A function is required in order to minimize the distance between the polynomial hypothesis and the signature of benign contour. The function which measures the difference between the generated $H(\theta)$ and signature $r(\theta)$ is called cost function (J) or mean square error. The equation of the cost function is shown below in equation (4.2)

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m (H(\theta^{(i)}) - r(\theta^{(i)}))^2 \quad (4.2)$$

where $r(\theta)$ is the actual radial distance of the boundary from the center at an angle θ and m is the number of samples. $\theta^{(i)}$ is a sample angle of the i^{th} pixel on the boundary which varies from $1 \dots m$ and θ varies from $-\pi$ to π .

All the polynomial coefficients $a_0, a_1, a_2, a_3, \dots, a_n$ are iteratively updated to optimize the hypothesis so that the cost function reaches the minimum value as shown in equation (4.3).

$$a_n = a_n - \alpha \frac{\partial}{\partial a_n} J(\theta) \quad (4.3)$$

Where α is known as the learning rate, n is the degree of the polynomial hypothesis. If α is too low then it takes a huge number of iterations for $J(\theta)$ to converge. If α is too high then $J(\theta)$ might not even converge, therefore a medium value must be chosen. This procedure is continued until $J(\theta)$ converges. The degree of polynomial is selected empirically such that polynomial hypothesis perfectly fits the signatures of benign and under-fit the signatures of malignant tumors.

4.3.2 Extraction of features

Mean absolute error and correlation between polynomial hypothesis $H(\theta)$ and signature $r(\theta)$ are the features extracted from benign and malignant contours to classify them. The correlation coefficient (CC) between two data sets $A = H(\theta)$ and $B = r(\theta)$ is given below in equation (4.4)

$$CC = \frac{\sum_{i=1}^m (A_m - A_{mean})(B_m - B_{mean})}{\sqrt{\sum_{i=1}^m (A_m - A_{mean})^2} \sqrt{\sum_{i=1}^n (B_m - B_{mean})^2}} \quad (4.4)$$

Where A_m and B_m is the individual sample point on signature and polynomial hypothesis and index m represents number of points on signature. A_{mean} and B_{mean} indicates mean of $H(\theta)$ and $r(\theta)$ respectively.

$$\text{mean absolute error} = \frac{1}{m} \sum_{i=1}^m |H(\theta^{(i)}) - r(\theta^{(i)})| \quad (4.5)$$

Equations 4.4 and 4.5 gives two features, that are correlation coefficient and mean absolute error of the signature $r(\theta)$ (green curve) shown in Figure 4.2a and its corresponding polynomial fit $H(\theta)$ (Red curve)shown in Figure.4.2b. We have applied polynomial fit with gradient descent algorithm with different degrees of polynomial to fit the signatures of benign masses. Correlation coefficient and mean absolute error were computed between signature of test image and polynomial hypothesis obtained by gradient descent algorithm for classification. Block diagram for the extraction of features is shown in Figure 4.3.

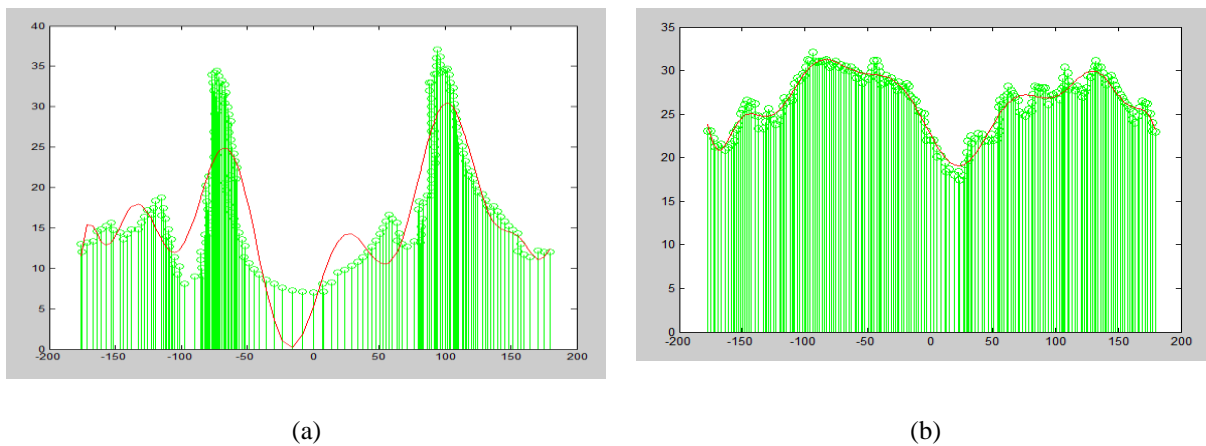


Figure.4.2: a) Malignant signature and its 15 degree polynomial hypothesis,b) Benign Signature and its 15 degree polynomial hypothesis

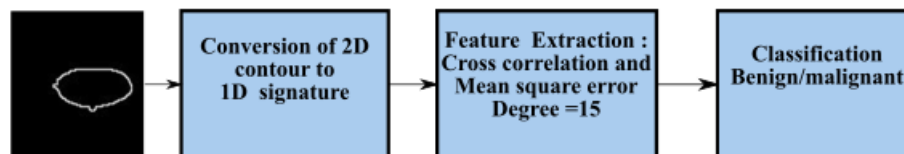


Figure 4.3: Block diagram for the extraction of features and classification using polynomial regression

4.4 Beam Angle Statistics

The 1D signature obtained by radial distance is sensitive to the noise in the contour. So, we proposed Beam angle statistics (BAS) [74] method to transform 2D contour of a mass into 1D signature and find the property of the regularity in benign and malignant masses. BAS method is insensitive to noise and it preserves concave and convex points of mass contours. Another advantage of this method is that its affine transforms and is invariant to size, orientation and position. At first, the contours have been taken from DDSM database using ground truth information. Second, we transform 2D contour into 1D signatures using BAS method. This signature gives spiculation degree and roughness of breast mass. Then, in the next step a compact feature vector is built, which retains the information in the BAS functions. The compact features have been extracted using different approaches, which are Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT). Along with these features, we also extracted root mean square (RMS) roughness, RMS slope, mean to standard deviation ratio of the 1D BAS functions. At the end, these feature vectors are fed to different classifiers. The block diagram of the proposed model is shown in Figure 4.4.

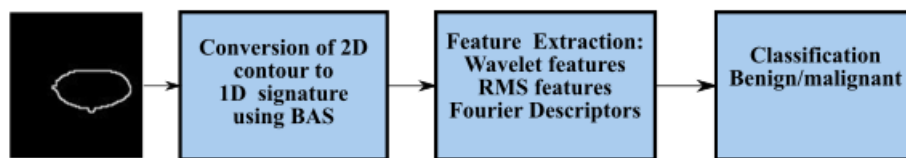


Figure 4.4: Block diagram for the extraction of features and classification using BAS

4.4.1 Conversion of 2D contour to 1D signature using (BAS)

Benign masses have smooth contours which gives simple signature. But, where as malignant contour have many spicules which results in rough signature. Figures.4.5a and.4.5b shows malignant and benign contours of breast masses.

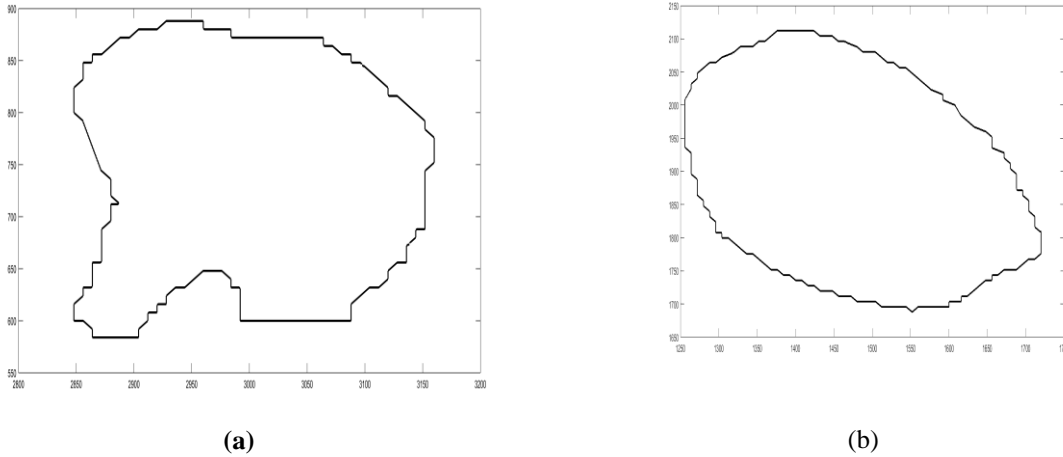


Figure 4.5: a) Malignant ,b) Benign

Any 2D contour is formed by orderly points and can be mathematically represented by $B = \{p(1), p(2), p(3) \dots p(i) \dots p(N)\}$, where N gives number of points on the contour and $p(i) = p(i + N)$. Each point on the contour have x and y coordinates which is given by $[p(i) = (x_i, y_i)]$. In BAS method, for each point $p(i)$, angle between two pair of beams are computed to obtain topological structure of contour. The beams of $p(i)$ are set of vectors, given by equation (4.6) and it is shown in Figure 4.6a

$$L(p(i)) = \{v(i + j), v(i - j)\} \quad (4.6)$$

In the above equation $v(i + j)$ and $v(i - j)$ are called forward and backward vectors connecting $p(i)$ with neighborhood points that are $p(i + j)$ and $p(i - j)$ on boundary with $j = 1, 2, 3, \dots, N/2$. K^{th} order neighborhood system is defined as

$p(i \pm k) \in n_k p(i)$, where $n_k p(i)$ is neighborhood of $p(i)$ and k varies from $1 \dots N/2$

In K^{th} neighborhood system, there exists one pair of beams that connects $p(i)$ to $p(i+k)$ and $p(i-k)$. Figure.4.6b shows pixel $p(i)$ connecting to 3 neighborhood systems.

The slope of each beam is given by equation (4.7)

$$\theta_{v_{i\pm l}} = \tan^{-1} \frac{\Delta y_{i\pm l}}{\Delta x_{i\pm l}} \quad \text{where } l = \pm k \quad (4.7)$$

For point, $p(i)$ the angle between two forward and backward beam vectors for K^{th} neighborhood system is shown below in equation (4.8)

$$C_k(i) = (\theta_{v_{i-k}} - \theta_{v_{i+k}}) \quad (4.8)$$

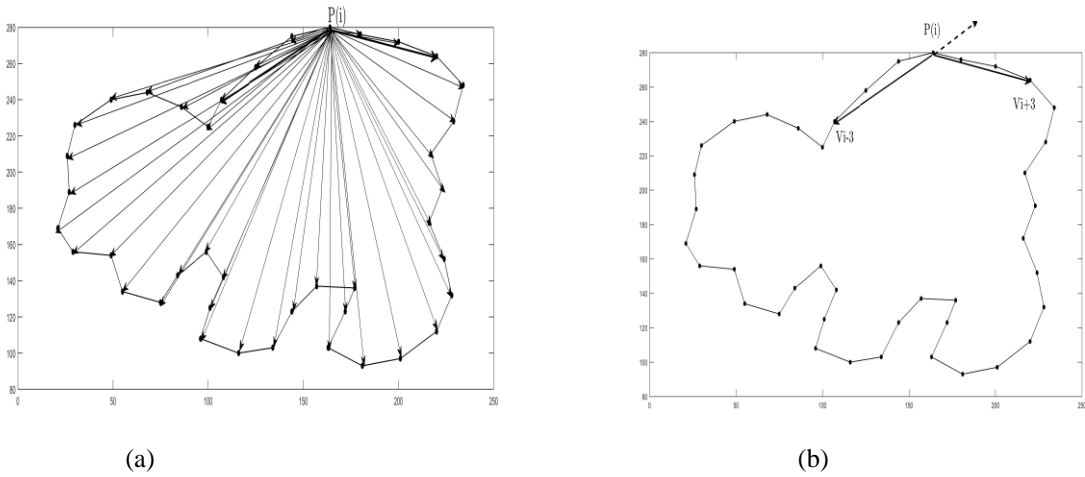
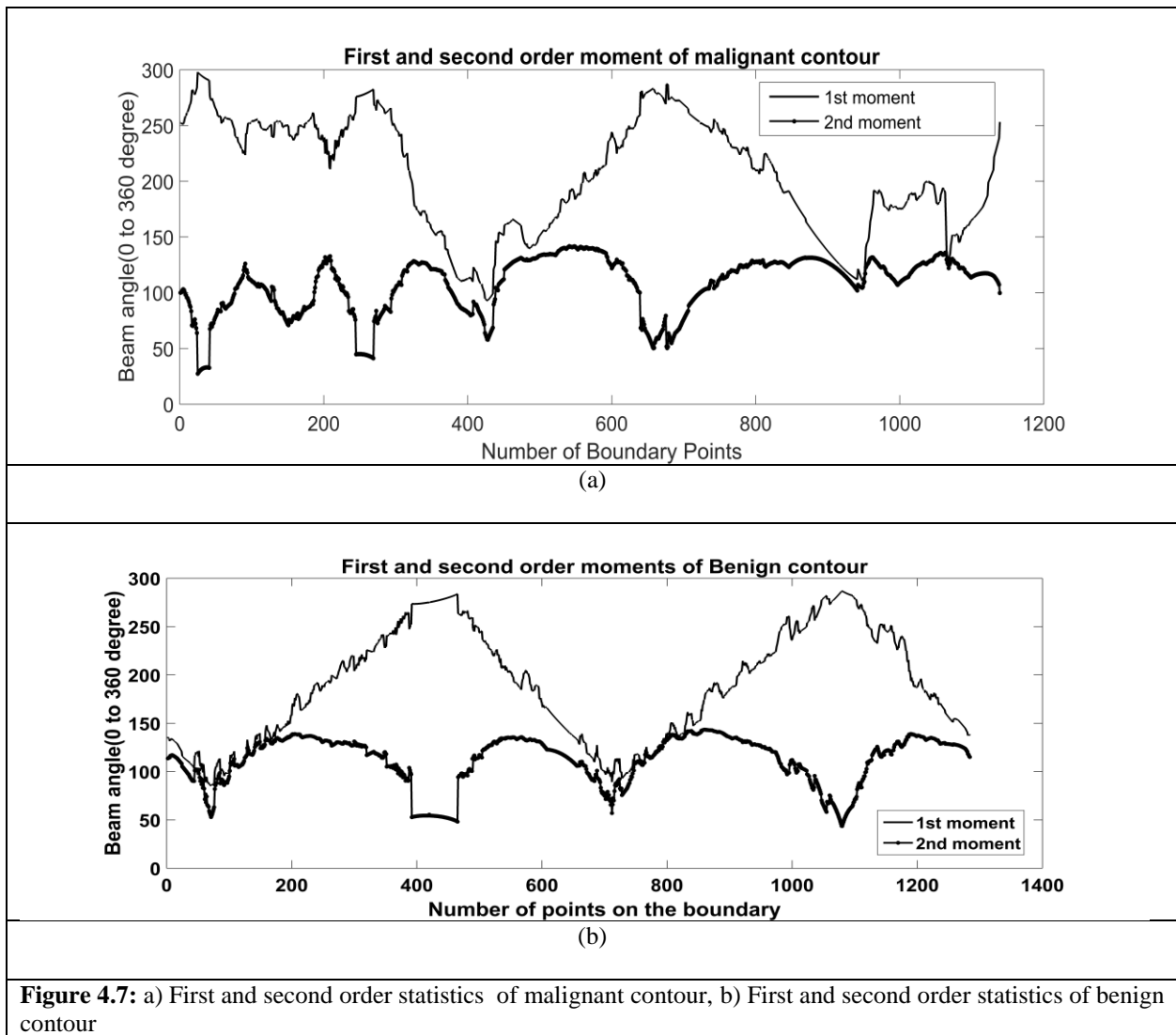


Figure.4.6: a) The beams of point $p(i)$, b) The 3-curvature and the beam angle at the neighborhood system 3 for the boundary point $p(i)$

Beam angle for any neighborhood system K varies from 0 to 2π . For each point $p(i)$, on the contour with the K^{th} neighborhood system $C_k(i)$ can be taken as random variable with the probability density function $P(C_k(i))$. The plot of beam angles $C_k(i)$ at any point $p(i)$ for any given neighborhood k vs i (number of points on contour) can result in any stochastic process which describes the shape at different scales. The beam angle statistics gives information on spiculation or fluctuations present in the contour at each point on the mass contour. The m^{th} moment of random variable $C_k(i)$ is given by

$$\tau(i) = [E^1(C_k(i)), E^1(C_k(i)), E^2(C_k(i)), E^3(C_k(i)) \dots \dots \dots]$$

The shape description of the mass contour is given by moments of $C_k(i)$ and is represented by $\tau^m(i)$. The first two moments of this random variable $C_k(i)$ for the contours in Figures.4.5a&4.5b is shown in Figures.4.7a&b.



4.4.2 Extraction of features from BAS moments

After transforming 2D contour into 1D BAS functions, we need to reduce the size of the feature vector, as BAS shape descriptor is computationally expensive. We need to retain the information of BAS function with a compact feature vector. Thus, we applied two different approaches, which include DFT and DWT. We also extracted root mean square roughness, RMS slope, mean to standard deviation ratio of the 1D BAS functions.

Discrete Fourier Transform (DFT): Fourier descriptors (FD) are extracted by applying DFT to 1D BAS functions. They describes shape boundary. FD for a given m^{th} moment 1D BAS function, $\tau^m(i)$ is defined by

$$a_n^{(m)} = 1/N \sum_{i=0}^{N-1} \tau^m(i) \exp\left(\frac{-j2\pi i n}{N}\right) \quad m= 1, 2, \dots, M$$

The magnitude of $a_n^{(m)}$ are considered as feature vectors to achieve invariance in the starting point of the boundary. Then first N low frequency FD are normalized by DC component and submitted as feature vector to classifier for further classification.

Discrete Wavelet Transform (DWT): 1D BAS functions $\tau^m(i)$ are decomposed into four levels using Daubechis wavelet transform. Wavelets Detail coefficients or Wavelet Descriptors (WD) are used as feature vector. The reason for choosing only detail coefficients is that they contain high frequency information [76]. Higher magnitude of detail coefficients indicates large fluctuations. Large fluctuations indicate malign and small indicates benign.

Root mean square (RMS) features

RMS roughness: 1D BAS function $\tau^m(i)$ is considered as 1D signal. Root mean square roughness [77] specifies the regularity degree of $\tau^m(i)$. RMS roughness is defined as $w = \sqrt{\langle h^2 \rangle - \langle h \rangle^2}$, where h is the magnitude in vertical direction of 1D BAS signal. The symbol $\langle a \rangle$

indicates the statistical average of any variable a . If, the value of w is low, shape of the contour is more regular and tends to be circle or ellipse. So, the mass can be more benign than malignant.

RMS slope: The slope at each point on the contour describes the smoothness. The slow variation of slope indicates contour's smoothness and large variation of slope implies that the contour has more spicules [80]. After transformation of 2D contour into 1D BAS function. The slope of 1D BAS function at each point is obtained in the vertical direction and RMS slope is evaluated. High value of slope in vertical direction indicates malignant contour and small value of slope value describes benign. RMS slope is defined by

$$S = \sqrt{\left\langle \frac{dh}{di} \right\rangle^2}$$

Where h indicates variation of 1D BAS function in vertical direction and i indicates points on contour and it varies from $i=1, 2, 3 \dots N$.

Mean to standard deviation ratio (M/sd): The ratio of mean value to standard deviation of the 1D BAS function in vertical direction is taken as one of the features. This ratio is small, if the standard deviation (**sd**) of 1D BAS function is high. Malignant masses have many spicules, which indicate that there is a major variation in beam angles at each boundary point. This indicates that **M/sd** ratio is small for a malignant and high for benign mass.

4.5 Triangular area Representation

4.5.1 Triangular area representation (TAR) signatures

Let three consecutive points on mass contour be $(x_n - t_s, y_n - t_s)$, (x_n, y_n) , and $(x_n + t_s, y_n + t_s)$, where $n \in (1, N)$ and $t_s \in (1, T_s)$ is the triangle side. The signed area of the triangle formed by these points is given below in equation (4.9)

$$\text{TAR}(n, t_s) = \frac{1}{2} \begin{vmatrix} x_{n-t_s} & y_{n-t_s} & 1 \\ x_n & y_n & 1 \\ x_{n+t_s} & y_{n+t_s} & 1 \end{vmatrix} \quad (4.9)$$

$t_s = 1, 2, 3, 4, \dots, T_s$, where the maximum T_s depends on periodicity of N .

When we traverse the contour in counter clock wise direction positive, negative and zero values of TAR signature represents convex, concave and straight line points on contour [78]. The triangle side length value (t_s) depends on the value of number of points (N) on contour. Equation (4.10) gives values of TAR with respect to N .

$$\text{TAR}(n, t_s) = \begin{cases} -\text{TAR}(n, N + 1 - t_s) & t_s = 1 \dots \left\lfloor \frac{N-1}{2} \right\rfloor \\ 0 & \text{at } t_s = \frac{N}{2} \text{ and } N \text{ is even} \\ \text{does not exist} & t_s = \frac{N}{2} \text{ and } N \text{ is even} \end{cases} \quad (4.10)$$

4.5.2 Conversion of 2D contour into 1D signal and feature extraction

We have applied triangular area function to study the complexity of mass contours. TAR converts 2D contour into 1D signal. It uses all boundary points in mass contour to measure the convexity and concavity of each point at different scales. This 1D TAR signature is invariant to translation, rotation, and scaling. The block diagram for classification using TAR is shown in Figure 4.8.

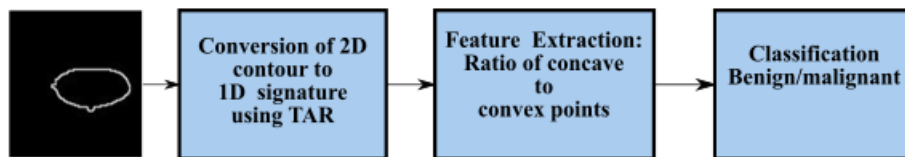
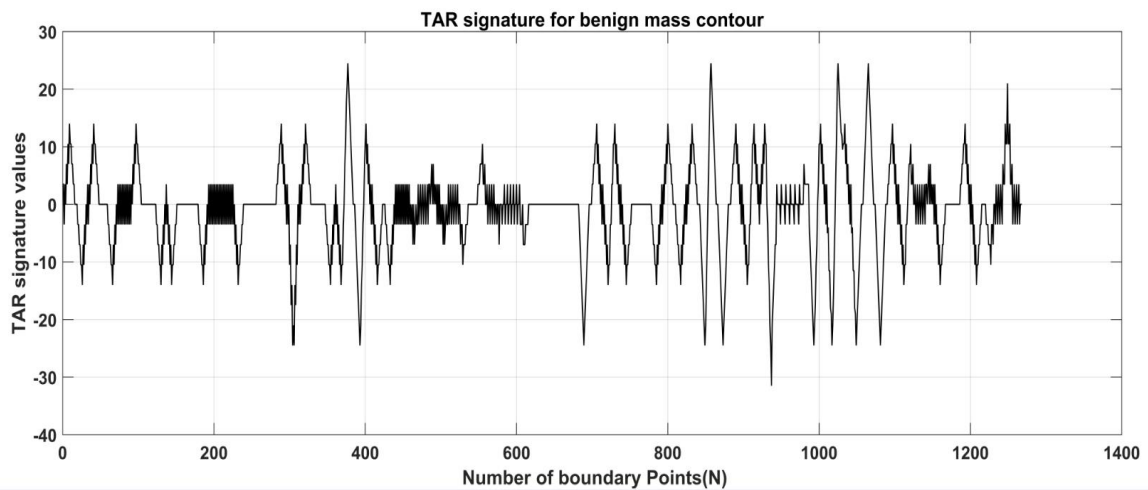


Figure 4.8: Block diagram for the extraction of features and classification using TAR

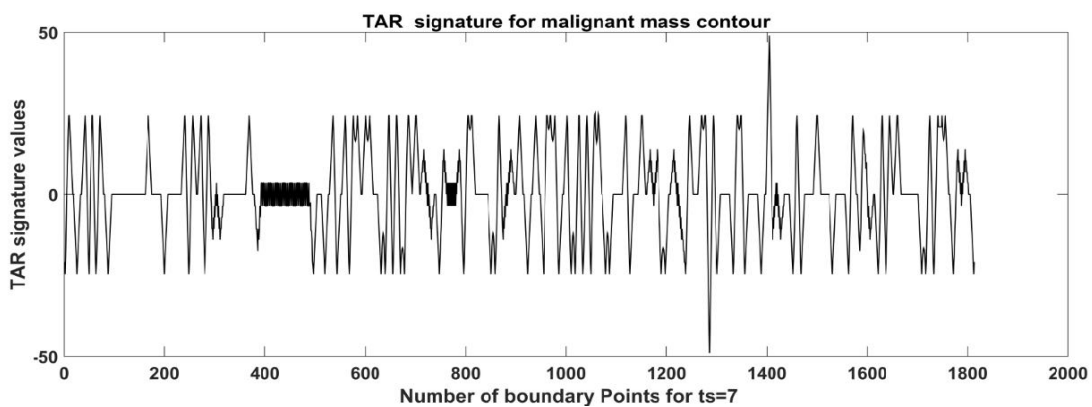
Feature Extraction from TAR Signature: As TAR signature values give number of concave and convex points. The ratio of number of concave points to number of convex points (R) is taken as a feature descriptor for different values of t_s and it is given below in the equation (4.11)

$$R = \frac{\text{Number of concave points}}{\text{Number of convex points}} \quad (4.11)$$

The malignant mass contour is very complex and benign contour is smooth. From the value of R we can discriminate mass contour or boundaries. To give more discrimination power to feature descriptors we calculated area, perimeter and eccentricity to mass contours. Figures 4.9a and b shows TAR signatures of malignant and benign mass contours. From the TAR signatures we can observe that they are many convex and concave regions in malignant TAR signature and many straight lines in benign TAR signature.



(a)



(b)

Figure. 4.9: TAR signatures for benign contour and malignant contour

4.6 Classification

K-Nearest-neighbor (KNN), support vector machine (SVM), adaboost decision tree classifier and artificial neural network (ANN) are used as classifiers in this work to delineate benign and malignant masses of breast.

4.7 Summary

In this chapter, we study the application of different shape descriptors for classification of masses into benign and malignant. For classification, two types of features are extracted from mass contours. One type is, features computed from 2D contours using their geometric properties. Second type is, two dimensional contours are transformed into one dimensional signatures using radial distance from the center, BAS and TAR to extract fluctuations in contour. Then, statistical methods, polynomial regression etc. are the methods used to compute features from these 1D signatures. Further, several classifiers are used to classify masses to benign/malignant. The results of all these methods are discussed in chapter 5

Chapter 5

Experimental Results and analysis

The algorithms developed for early detection of breast cancer in chapter 2, 3 and 4 are evaluated to depict the image processing techniques which help in the CAD system. Validation of the algorithms has been performed for image enhancement, segmentation, feature extraction and classification. In this chapter, simulation results of different algorithms and their analysis have been discussed.

5.1 Image Enhancement

The proposed image enhancement method is evaluated on 14 mammographic images taken from Krishna Institute of medical sciences (KIMS) and 9 images from DDSM database. The original and enhanced mammograms using the proposed and other methods such as unsharp masking [18] and contrast limit adaptive histogram equalization (CLAHE) [19] are shown in Figures 5.1 a-d. Figure 5.1a (column wise) shows images of four different patient out of which above two images are taken from KIMS database and the below two images are considered from DDSM database for image enhancement and Figures 5.1b to 5.1d (column wise) shows the enhanced images using the proposed and other methods. From the figures, we can observe that our proposed method mammogram (Figure 5.1b) shows better contrast compared to original mammogram (Figure 5.1a). The quality of the proposed method is evaluated using enhancement quantitative measures like edge based contrast measure (EBCM) [79] and entropy (E) [80]. The quantitative measures (EBCM and entropy) are tabulated in Tables 5.1 & 5.2 for mammograms taken from KIMS and DDSM database

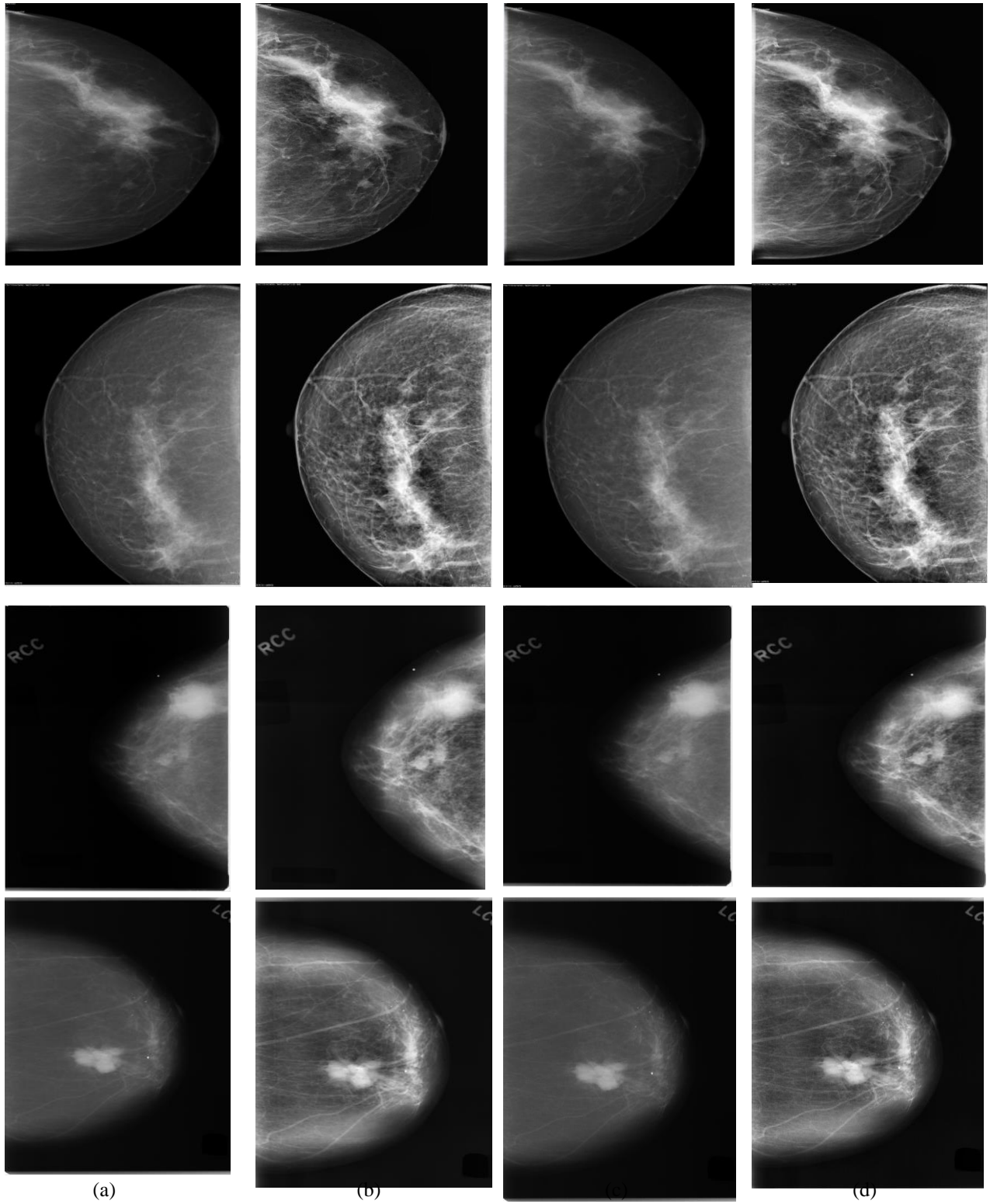


Figure 5.1: a) Original mammogram image, b) proposed method image, c) unsharpened image, d) CLAHE image

Table 5.1: EBCM values of mammographic images (Images from SNO 1 to 14 are from KIMS and images from SNO 15 to 23 are from KIMS)

SN O	Sample mammographic images	Original	Proposed method	Unsharp	CLAHE
1	Image 1	48.89	103.76	46.59	92.70
2	Image 2	46.50	92.55	43.97	85.00
3	Image 3	32.88	93.33	31.30	87.80
4	Image 4	56.61	101.33	53.92	89.34
5	Image 5	66.61	112.46	64.39	98.31
6	Image 6	70.21	114.48	67.85	101.00
7	Image 7	53.53	91.81	50.97	85.65
8	Image 8	39.38	98.57	37.99	92.17
9	Image 9	96.06	92.53	89.57	78.03
10	Image 10	43.43	102.45	42.18	95.17
11	Image 11	56.55	102.55	54.02	92.36
12	Image 12	37.80	97.32	36.20	90.82
13	Image 13	31.23	90.27	29.58	85.86
14	Image 14	33.93	97.89	32.68	91.37
15	Image 15	11.02	43.19	11.10	37.56
16	Image 16	17.05	48.87	17.01	44.91
17	Image 17	16.53	42.51	16.53	43.92
18	Image 18	25.99	46.60	27.01	55.12
19	Image 19	4.31	30.91	4.32	29.47
20	Image 20	11.56	19.96	11.44	20.30
21	Image 21	2.33	36.04	2.34	35.35
22	Image 22	2.10	22.03	2.10	16.85
23	Image 23	9.50	35.65	9.57	34.97

Table 5.2: Entropy values of KIMS mammographic images (Images from SNO 1 to 14 are from KIMS and images from SNO 15 to 23 are from KIMS)

SNO	Sample mammographic images	Original	Proposed Method	Unsharp	CLAHE
1	Image 1	3.09	3.39	3.08	3.45
2	Image 2	3.23	3.57	3.22	3.61
3	Image 3	2.47	2.79	2.45	2.83
4	Image 4	3.54	3.88	3.53	3.93
5	Image 5	3.84	4.11	3.83	4.22
6	Image 6	4.10	4.33	4.09	4.45
7	Image 7	3.75	4.07	3.74	4.11
8	Image 8	2.69	2.91	2.68	2.96
9	Image 9	5.83	6.16	5.83	6.20
10	Image 10	2.86	3.05	2.85	3.16
11	Image 11	3.65	3.94	3.64	4.00
12	Image 12	2.65	2.91	2.64	2.96
13	Image 13	2.43	2.71	2.42	2.74
14	Image 14	6.46	7.06	6.50	7.00
15	Image 15	6.46	7.06	6.50	7.00
16	Image 16	6.74	7.23	6.76	7.28
17	Image 17	6.21	6.94	6.28	6.94
18	Image 18	5.79	6.90	5.90	6.70
19	Image 19	5.79	6.77	5.86	6.61
20	Image 20	5.15	6.14	5.25	5.87
21	Image 21	5.76	6.94	5.84	6.86
22	Image 22	5.41	6.53	5.46	6.27
23	Image 23	5.46	6.59	5.54	6.31

In order to validate our method, we have compared EBCM and entropy values obtained from our method with the unsharp masking and CLAHE methods. The higher values of EBCM show that the contrast has been increased at the edges and low values indicate that edges are not preserved. From the table 5.1 we can observe that EBCM values are high for all cases compared to unsharp masking and CLAHE methods. Another measure that we have used to show the performance of the proposed method is entropy. The entropy values indicate, how the enhanced image preserves the information of original image without any variation. The quantitative measures shown in Tables 5.2 indicate that the entropy values of enhanced image with our method are slightly

higher compared to original image. The difference of entropy values of original and proposed method enhanced images is approximately 0.14 for KIMS database and approximately 1.3 for DDSM database. These values clearly indicate that the KIMS mammograms have less variation from original images when compared to DDSM database after image enhancement. But, in case of CLAHE and unsharp method the difference is almost 0.2 for KIMS and 1.4 for DDSM database. The above results confirm that our method is better when compared to the state of art methods. In our work, for the implementation of proposed enhancement method we have used Daubechies wavelet and max-max method for the wavelet fusion. In CLAHE, we have used uniform distribution, 256 bins and 8*8 tiles. We have used vision assistant in LABVIEW for implementing Gaussian pyramid and Matlab R2015b for wavelet fusion.

5.2 Image Segmentation

The proposed method for image segmentation is evaluated on seventeen mammograms with mass as abnormality. In these mammograms, six images have benign masses and eleven images have malignant masses. These are taken from DDSM database. The overlay file in database gives the information of mass location in mammogram which is segmented by expert radiologist. Objective evaluation is performed with the segmented mass of the proposed method against the ground truth mass. Dice coefficient is used as an evaluation measure for segmentation. It is defined in equation (5.1)

$$Dice\ Coefficient = \frac{GT \& SI}{GT \cup SI} \quad (5.1)$$

where GT indicates the ground truth and SI indicates the segmented image. The original mammogram, ground truth mass and output segmented mass are shown in Figure 5.2. The Dice coefficient of each image is tabulated in Table 5.3. In Table 5.3, quantitative results show that the proposed method achieved average dice coefficient of 0.5654 and 0.6249 for benign and

malignant masses respectively. In the proposed method, thresholding and morphological techniques have been applied to segment mass in the mammogram. The advantage of our method is that the segmentation process is carried out from a complete region of mammogram.

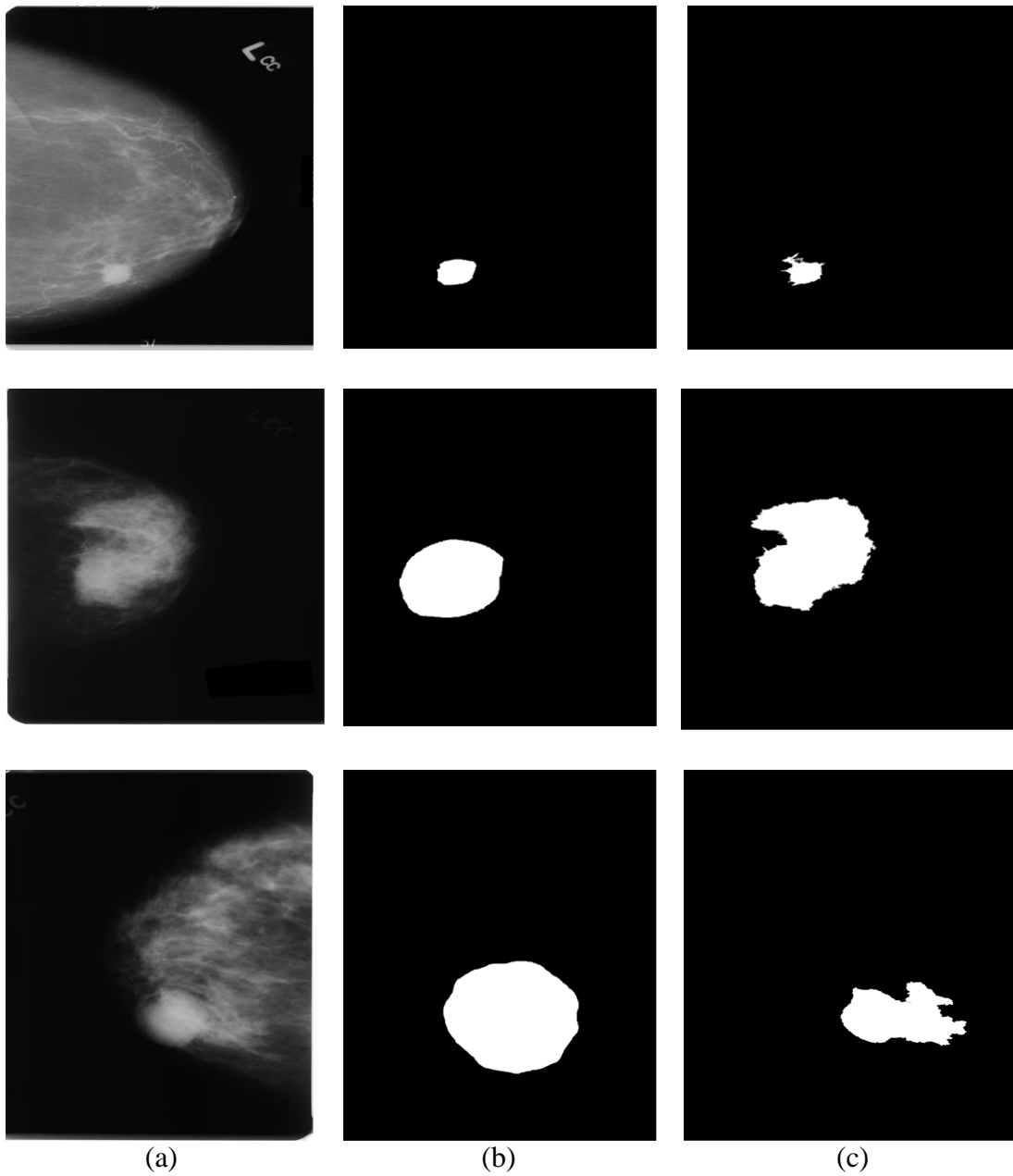


Figure 5.2: (a) Original images selected from DDSM,(b) Ground truth segmented mass ,(c) Output Segmented mass

Table 5.3: Dice Coefficients of benign and malignant mass

Patients	Dice Coefficients
Benign 1	0.623
Benign 2	0.2745
Benign 3	0.6192
Benign 4	0.6554
Benign 5	0.6086
Benign 6	0.6118
Average of benign	0.5654
Malignant 1	0.6843
Malignant 2	0.4771
Malignant 3	0.4198
Malignant 4	0.693
Malignant 5	0.903
Malignant 6	0.8021
Malignant 7	0.5006
Malignant 8	0.6015
Malignant 9	0.6027
Malignant 10	0.6787
Malignant 11	0.511
Average of malignant	0.6249

5.3 Analysis of mammograms into normal-abnormal

In this section, we have studied the implementation of LBP, RILF and SFTA feature extraction techniques to classify mammogram ROIs into normal-abnormal. These extracted features are then fed to a SVM classifier (linear kernel) via ten-fold cross validation method to assess the significance of the feature extraction techniques.

5.3.1 Database

The dataset used in this work is taken from Image retrieval in medical applications (IRMA) project database. The database comprises of ROIs taken from digital mammogram images of selected patients from the Radiological Diagnosis Department of the University of Aachen. IRMA database contains 2777 ROIs of which 932 are normal, 1157 are mass and 688 are microcalcification. The images in the dataset are of 128*128 pixel size and at 256 gray levels.

These ROIs undergo feature extraction process using RILF, LBP and SFTA techniques. Then, these features are given to classifiers to detect abnormal ROIs. There are two types of classification. First type separates ROIs into mass-normal and second type delineates ROIs into microcalcification-normal. To carry out these classifications, we have considered two datasets from IRMA database. Dataset 1 contains 1157 mass and 932 normal ROIs. These ROIs are used for both testing and training in mass-normal classification. Dataset 2 encompasses 688 microcalcification and 932 normal ROIs for microcalcification-normal classification.

5.3.2 Results of Method 1

In this method, ROIs are tested using different LBP and its variants, SFTA and RILF techniques and the results are discussed below.

LBP and its variants: In the implementation of LBP^{riu2} , LBP^{ri} and LBP^{u2} , we have used different neighborhoods ((8, 1), (12, 2) and (16, 2)) and different encoding (LTP and LQP) for testing. The superscripts $riu2$, $u2$ and ri indicates rotation invariant uniform, uniform and rotation invariant patterns respectively.

Results for mass-normal classification: In LBP, uniform LBP resulted in best values (accuracy of 88.18%, sensitivity of 93.50% and specificity of 83.88 %) with (16,2) neighborhood as shown in Figure 5.3. In LTP, uniform LTP (LTP_{u2}) gave best values (accuracy of 91.47%, sensitivity of 99.22% and specificity of 85.23%) with (8,1) neighborhood as shown in Figure 5.4. In Figure.5.5, we can observe that in LQP, LQP^{ri} resulted in best values (accuracy of 91.83 %, sensitivity of 99.16 % and specificity of 85.92 %) with (12,2)

Results for microcalcification-normal classification: In LBP, uniform LBP resulted in best values (accuracy of 82.52%, sensitivity of 86.51% and specificity of 77.12%) with (8,1) neighborhood as shown in Figure 5.6. In LTP, uniform LTP (LTP^{u2}) has obtained the best

accuracy of 87.80%, with 97.03% sensitivity and of 75.32% specificity using (8,1) neighborhood as shown in Figure 5.7. In LQP, uniform rotation invariant LQP (LQP^{riu2}) has resulted best values (accuracy of 86.41%, sensitivity of 94.44% and specificity of 75.56 %) with (16,2) neighborhood as shown in Figure 5.8. Among all the LBP variants, rotational invariant LQP (LQP^r) gave highest accuracy of 91.83% for mass-normal classification. Uniform LTP gave highest accuracy of 87.80% for microcalcification-normal. The highest accuracy is achieved using LTP and LQP then standard LBP because they are more robust to noise due to their encodings.

SFTA

Results for mass-normal: Figure 5.9 shows that SFTA method has achieved the best accuracy of 89.70%, sensitivity of 95.31% and specificity of 41.10% with $n_t=3$.

Results for microcalcification-normal: Figure 5.10 shows the accuracy attained by SFTA method is 86.44%, sensitivity of 88.94% and specificity of 81.28% for microcalcification-normal classification with $n_t=3$.

RILF: The rotation invariant local frequency magnitude descriptors are calculated by considering 2D spectrum of two low frequency channels. This method is tested on ROIs by considering five neighborhoods $((N,R) = (8,1), (16,2), (24,3), (28,4) \text{ and } (32,5))$. We have extracted features from both channel 1(ch1) and channel 2 (ch2). Figure 5.11 shows that RILF method gives best values (accuracy of 93.53%, sensitivity of 99.52% and specificity of 88.71%) with neighborhood (8,1) and both channels for classifying mass-normal ROIs. Figure 5.12 shows the results for microcalcification-normal classification. It depicts that the RILF gives best values (accuracy of 91.11%, sensitivity of 97.45% and specificity of 82.50 %) with neighborhood (8,1) and channel 2.

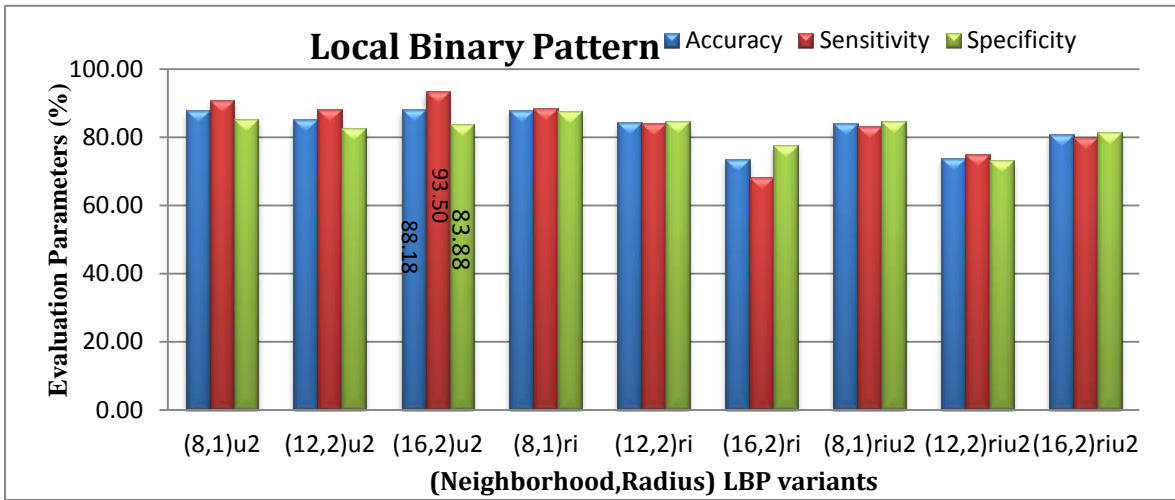


Figure 5.3: Results of LBP (mass-normal)

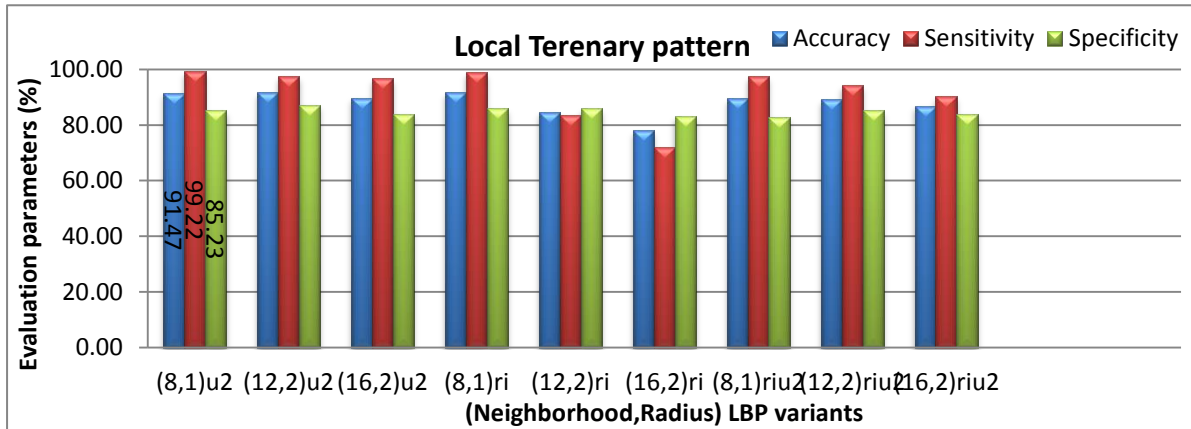


Figure 5.4: Results of LTP (mass-normal)

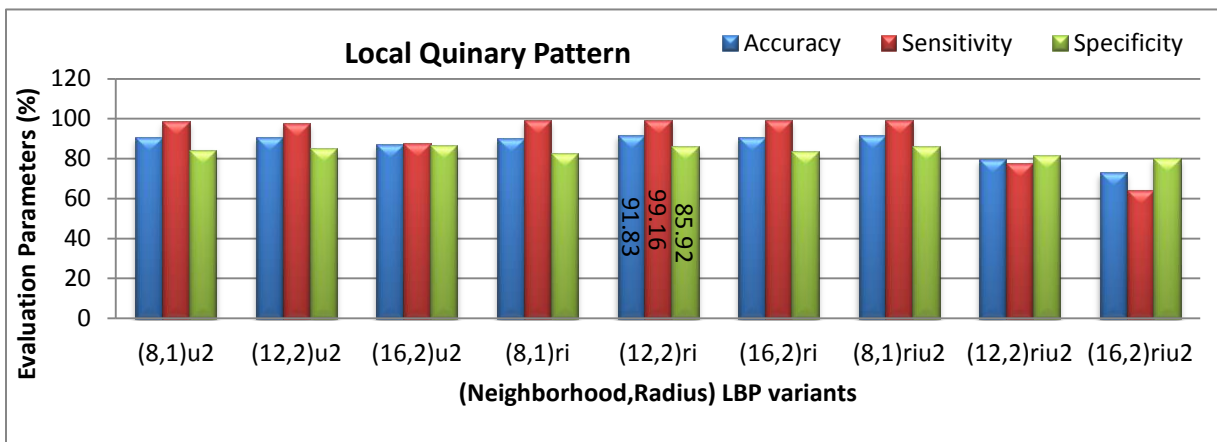


Figure 5.5: Results of LQP (mass-normal)

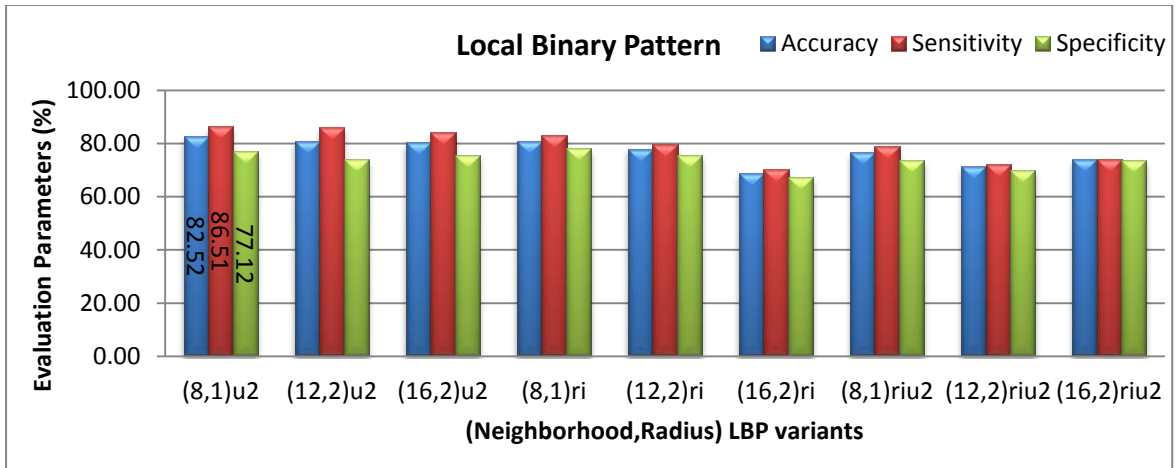


Figure 5.6: Results of LBP (microcalcification-normal)

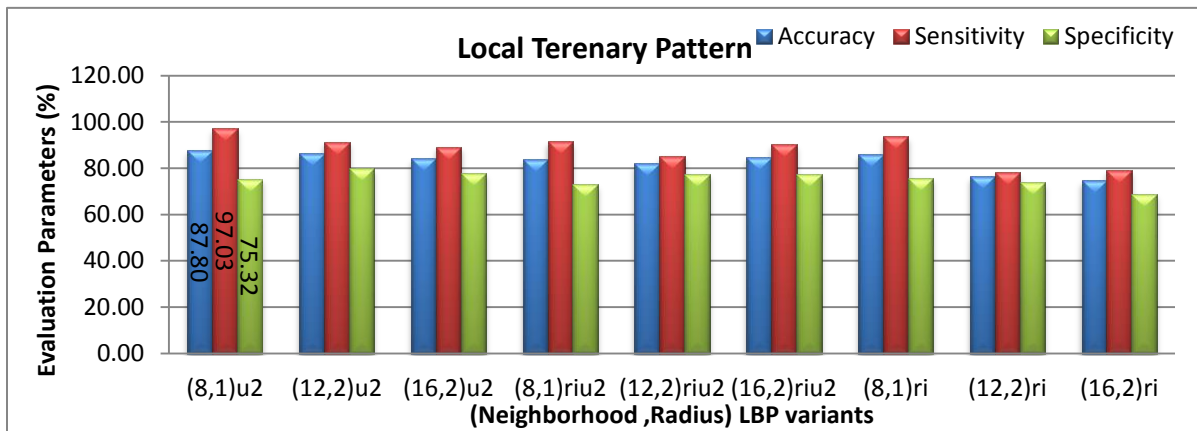


Figure 5.7: Results of LTP (microcalcification-normal)

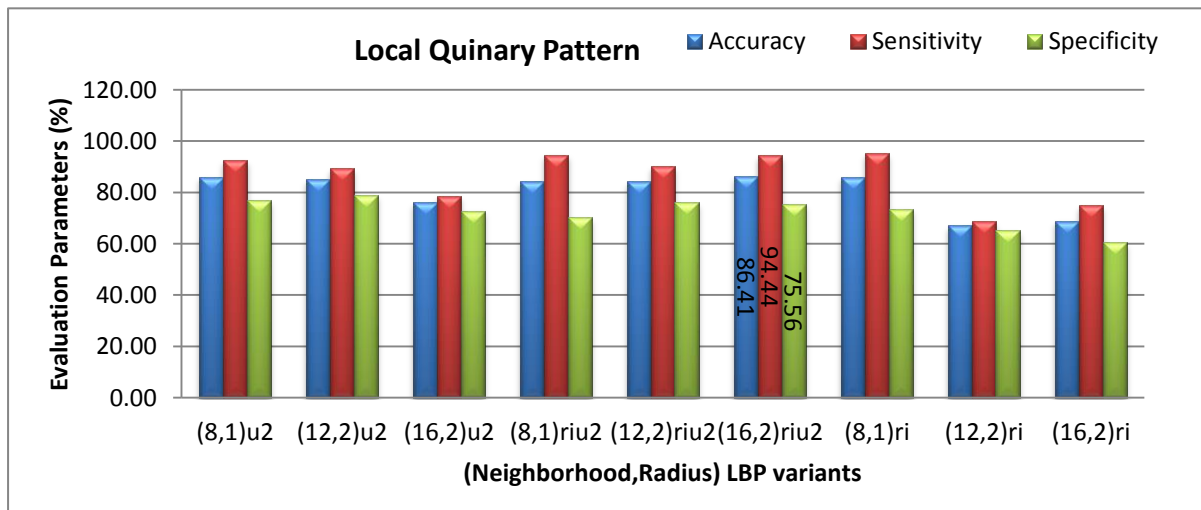


Figure 5.8: Results of LQP (microcalcification-normal)

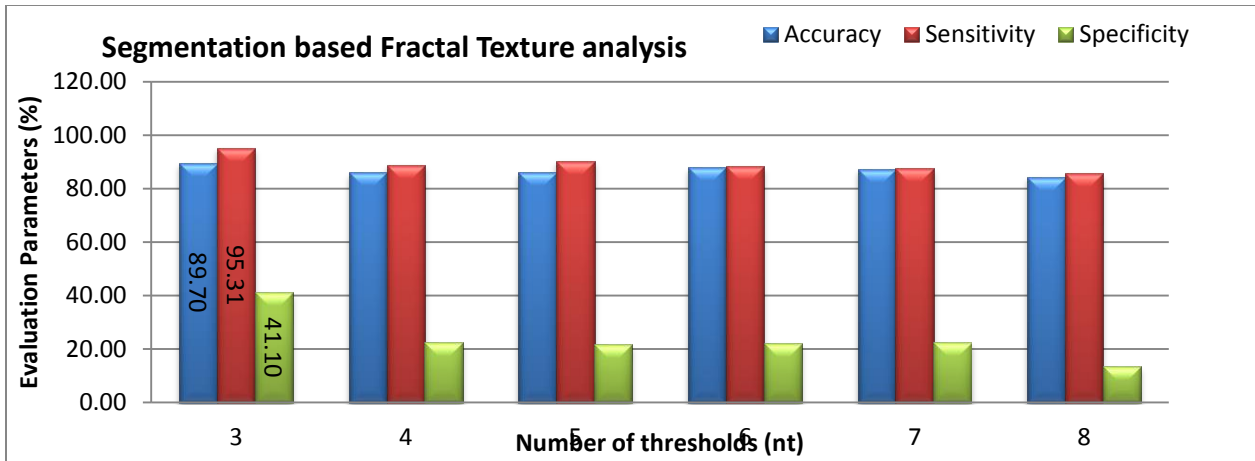


Figure 5.9: Results of SFTA(mass-normal)

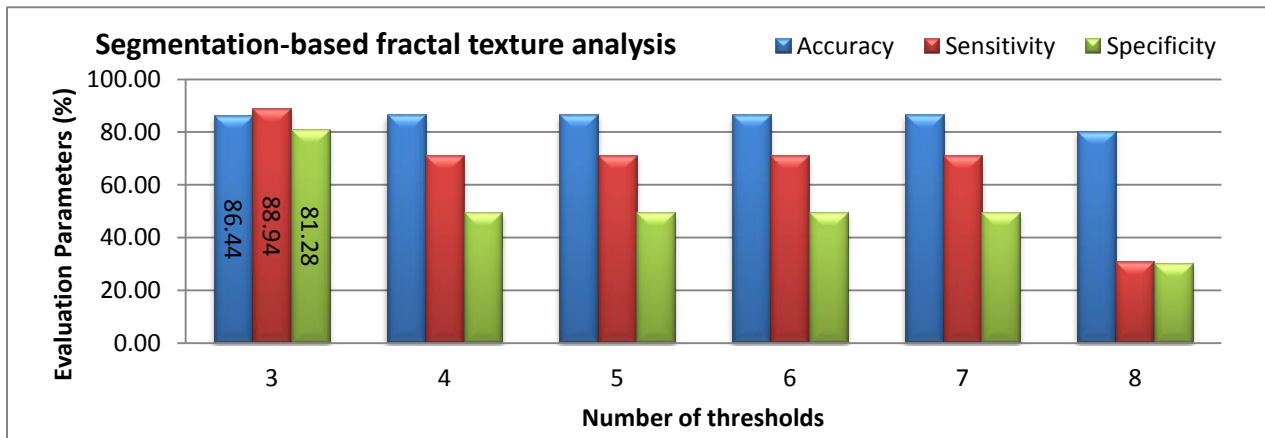


Figure 5.10: Results of SFTA(microcalcification-normal)

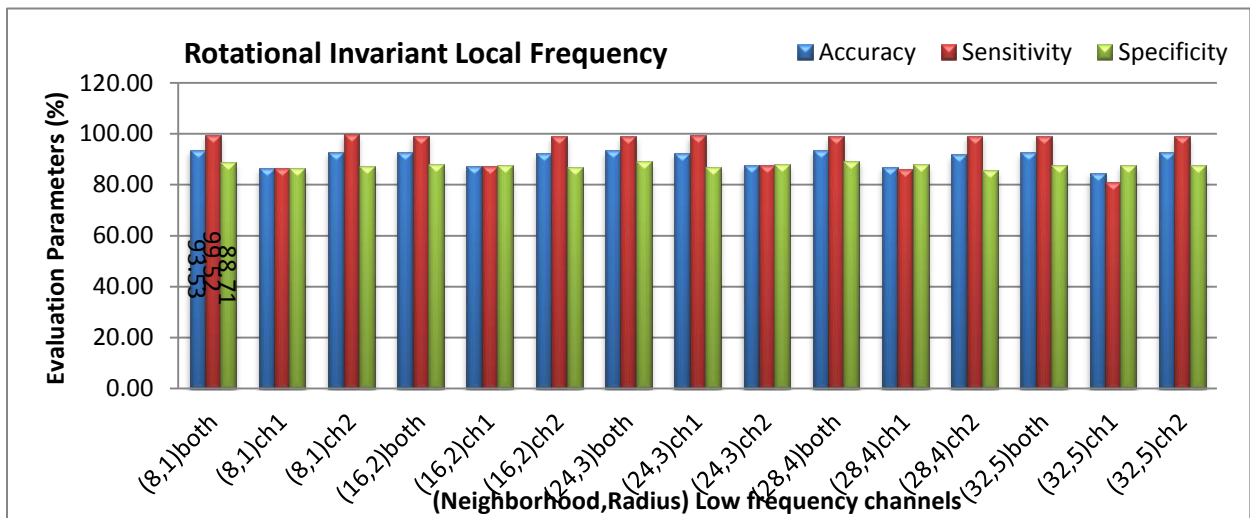


Figure 5.11: Results of RILF (mass-normal)

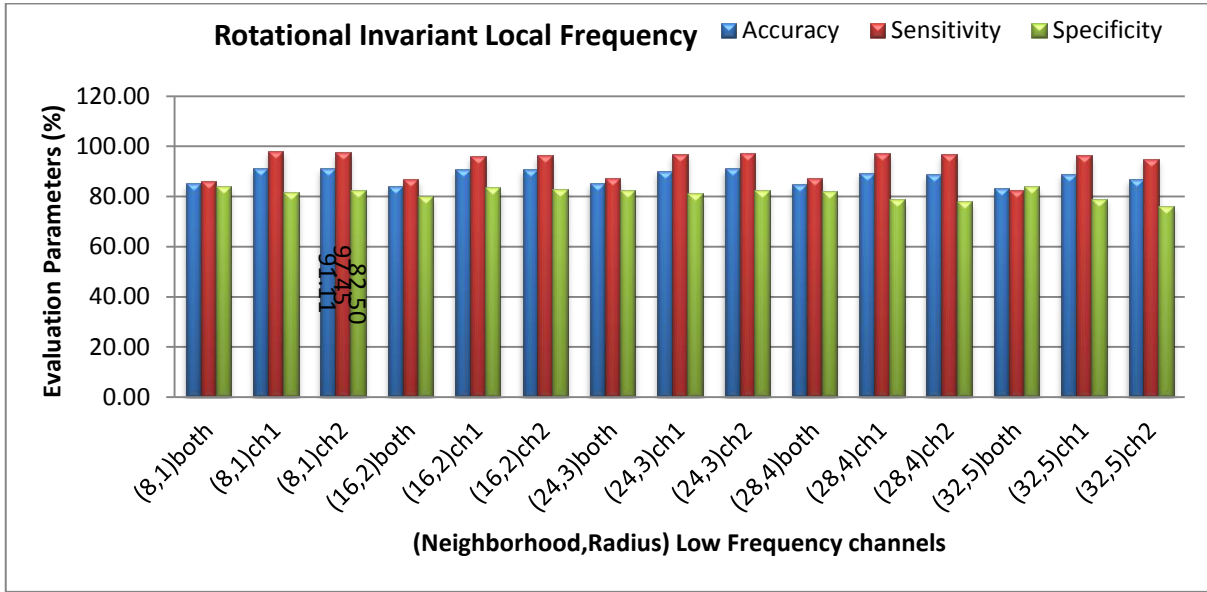


Figure 5.12: Results of RILF (microcalcification-normal)

5.3.3 Results of Method 2

In this method, ROIs are subdivided into N*N blocks. Then the features are extracted from each block and fused together to form a feature vector. LBP, RILF and SFTA techniques are used to implement this method and the results are given below. The parameters considered to implement these techniques are same as that of method 1.

Blockwise local binary pattern (BLBP): Uniform rotation invariant LBP (LBP^{riu2}) was tested using this method 2 with number of blocks (N= 2, 4, 8 and 16).

Results for mass-normal: The best accuracy of 87.67% is achieved by BLBP with 2*2 blocks per ROI and neighborhood of (8,1). Figure 5.13 depicts the sensitivity and specificity with BLBP is 94.74% and 81.99% respectively

Results for microcalcification-normal: In this classification, best accuracy of 79.25% is achieved with 4*4 blocks per ROI and neighborhood (8,1). From Figure 5.14 we can observe that BLBP features achieved sensitivity of 82.26% and specificity of 75.17% respectively.

Blockwise segmentation based fractal texture analysis (BSFTA): SFTA is applied to each block of ROI to extract features using method 2 with $N=2, 4, 8$. The features extracted from each block are concatenated to form enhanced feature vector.

Results for mass-normal: The results shown in Figure.5.15 makes it clear that BSFTA gives best accuracy of 88.47%, sensitivity of 89.163%,and specificity of 42.54% with $N=2$ and number of thresholds (n_t) =3 for mass-normal.

Results for microcalcification-normal: From Figure 5.16, we can observe that BSFTA has achieved accuracy of 88.55%, sensitivity of 84.88% and specificity of 77.24% with $N=2$ and $n_t=3$ for microcalcification-normal.

Block wise rotational invariant local frequency (BRILF): RILF is used to extract features using method 2 with $N=2$ and $N=4$. The features extracted using this method is given to classifier for further validation.

Results for mass-normal: BRILF method has achieved the best accuracy of 92.49%, sensitivity of 98.96% and specificity of 87.26% using channel 2(ch1), $N=2$ and with neighborhood of (16,2) as shown in Figure 5.17

Results for microcalcification-normal: From Figure 5.18, we can observe that BRILF method has achieved accuracy of 90.42%%, sensitivity of 96.55% and specificity of 82.15% with $N=4$, and used channel 2(ch2) with (8,1) neighborhood.

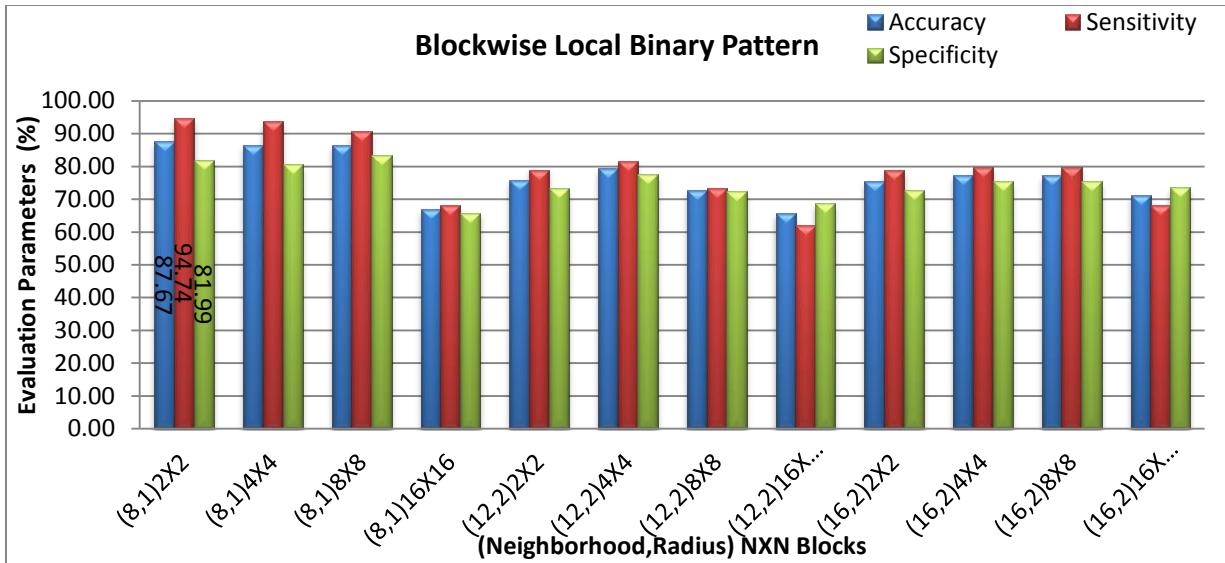


Figure 5.13: Results of BLBP (mass-normal)

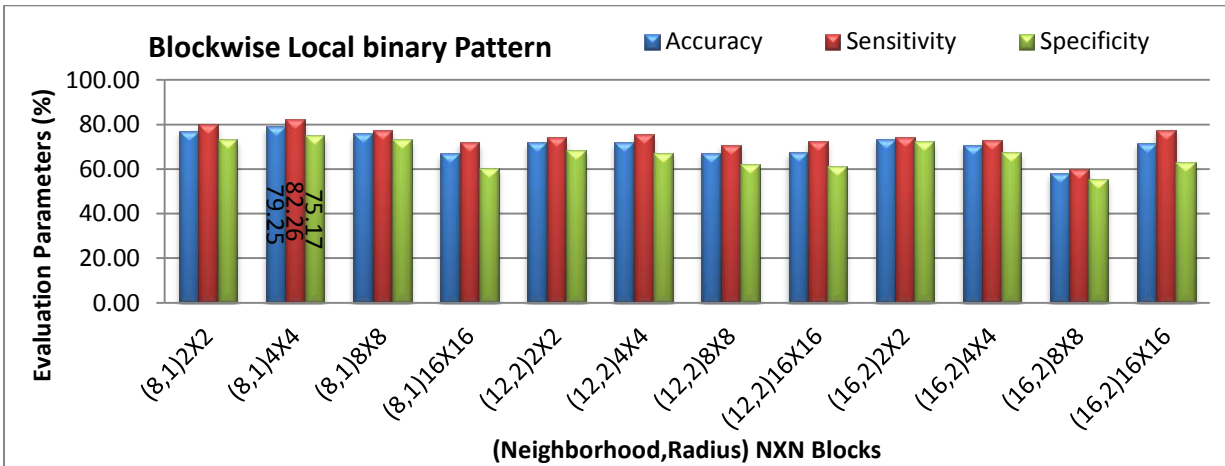


Figure 5.14: Results of BLBP (microcalcification-normal)

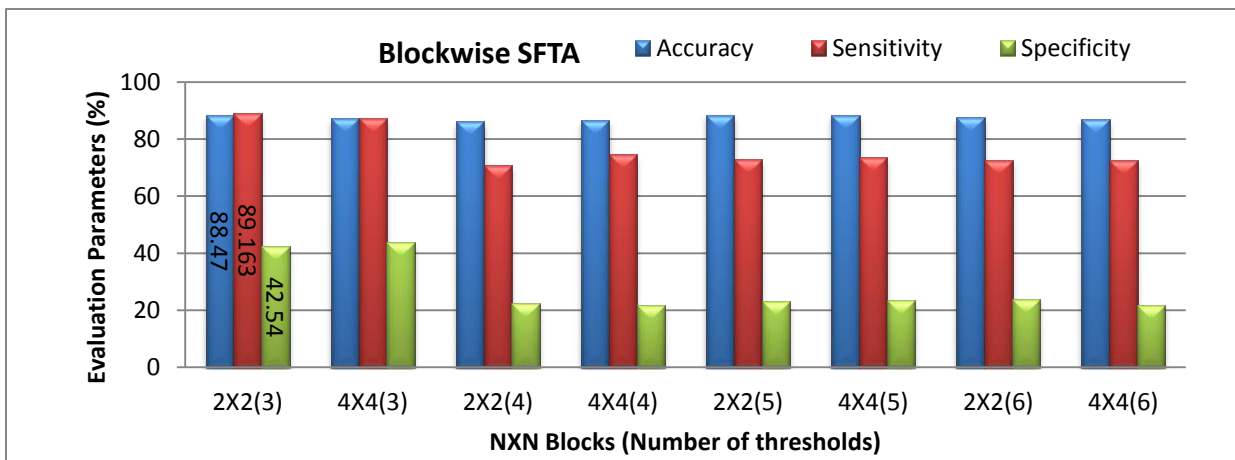


Figure 5.15: Results of BSFTA (mass-normal)

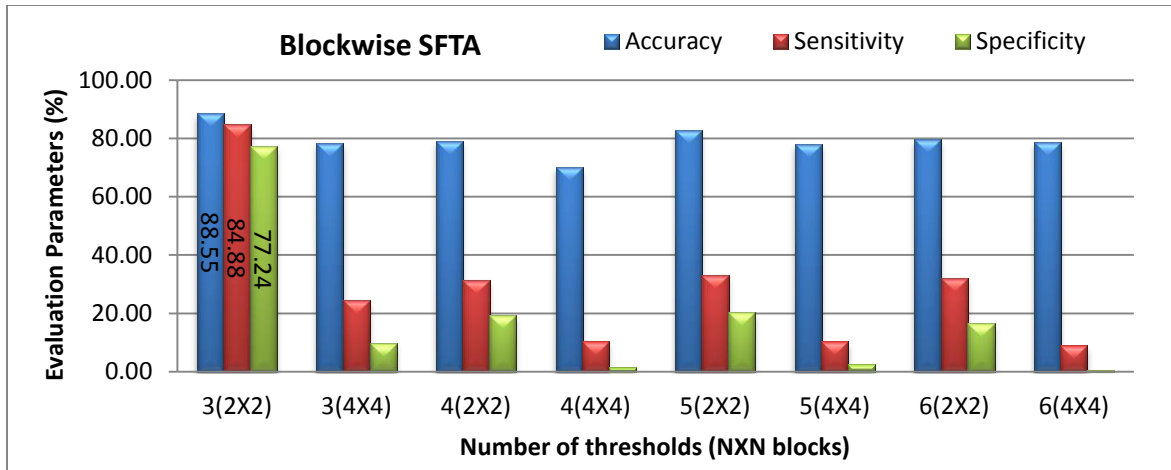


Figure 5.16: Results of BSFTA (microcalcification -normal)

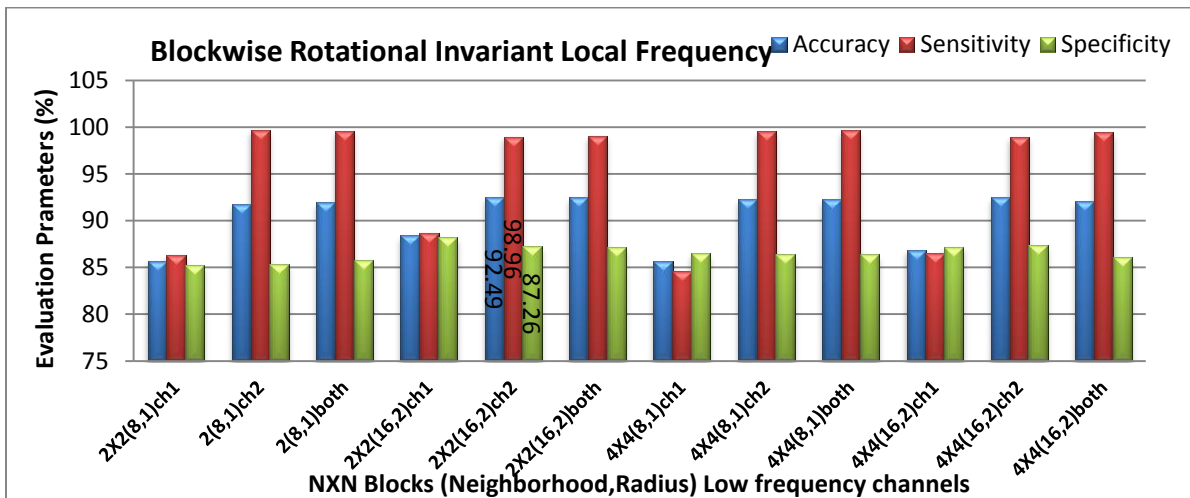


Figure 5.17: Results of BRILF(mass-normal)

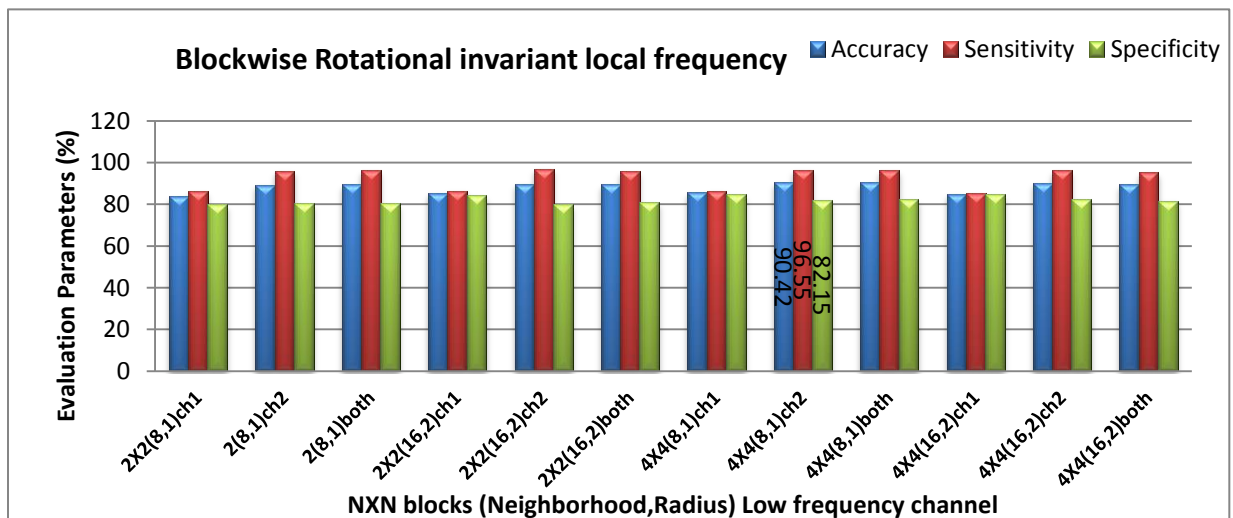


Figure 5.18: Results of BRILF(microcalcification-normal)

5.3.4 Results of Method 3

Method 3 utilizes LBP, SFTA and RILF techniques to classify ROIs. In this method, ROIs are decomposed into different levels and features extracted from each level are combined to form feature vector. The results obtained by method 3 using these techniques are given below

Pyramidal local binary pattern (PLBP): PLBP features are extracted from ROIs using three decomposition levels 2, 3 and 4 with three neighborhoods (8,1), (12,2) and (16,Bl02).

Results for mass-normal: PLBP resulted in best accuracy of 83.35% with neighborhood (8,1) and three level decomposition. From Figure 5.19, we can observe that it has sensitivity of 86.52% and specificity of 80.79%.

Results for microcalcification-normal: In PLBP, the best accuracy of 76.94% is obtained when ROIs are decomposed into three levels and with neighborhood of (8,1). PLBP method achieved sensitivity of 79.38% and specificity of 73.62% as shown in Figure 5.20

PSFTA: SFTA has been applied to three decomposition levels that are 2, 3, 4 and considered number of thresholds ($n_t = 3, 4, 5, 6, 7, 8$) for evaluation.

Results for mass-normal: Figure 5.21 shows that PSFTA gives best accuracy of 91.98%, sensitivity of 97.29%, and specificity of 87.70% by decomposing the images into two levels and selecting $n_t = 3$ for SFTA feature extraction.

Results for microcalcification-normal: From Figure 5.22, we can observe that PSFTA has achieved the accuracy of 89.82%, sensitivity of 91.60% and specificity of 83.90% at decomposition level 3 and number of threshold $n_t = 4$.

From the statistics obtained from three methods, we have observed that specificity in case of mass-normal is very less because SFTA method gave inconclusive results for many normal ROIs

for all the thresholds ($n_t = 3, 4, 5, 6, 7, 8$). The reason for such results is, SFTA method performs global segmentation and does not consider illumination and local changes in the image.

Pyramidal rotational invariant local frequency (PRILF): PRILF features are extracted by decomposition of ROIs using pyramidal filters. Then RILF is applied to three levels

Results for mass-normal: PRILF gives best accuracy of 94.36%, sensitivity of 99.45%, and specificity of 90.26% by decomposing the images into three levels and selecting channel 2 with neighborhood (16,2). It is shown in Figure 5.23.

Results for microcalcification-normal:

The best accuracy of 91.59%, sensitivity of 98.19% and specificity of 82.67% is achieved using PRILF with decomposition level equal to three and selecting both the channels with neighborhood (8,1). The results of PRILF are shown in Figure 5.24

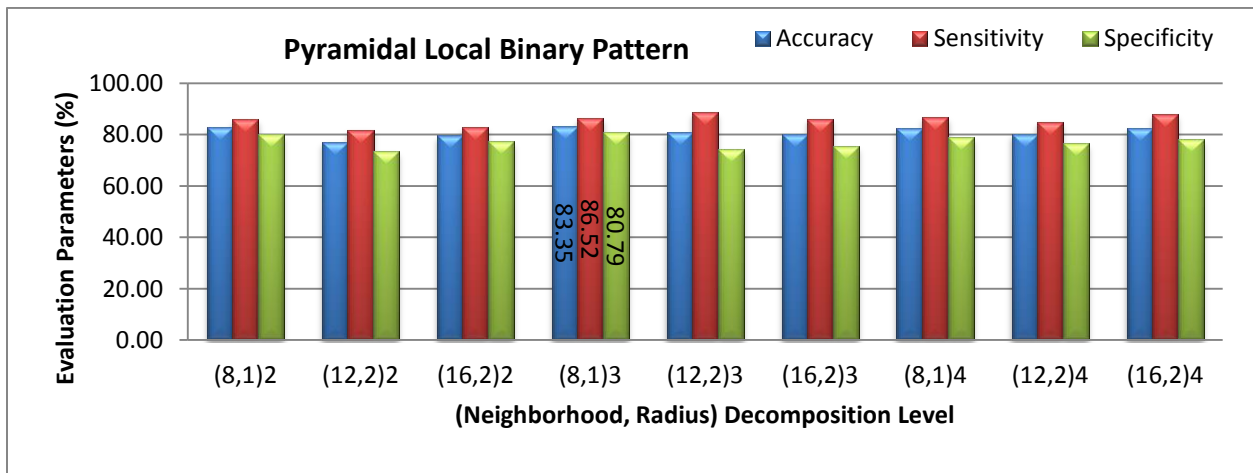


Figure 5.19: Results of Pyramidal Local binary Pattern (mass-normal)

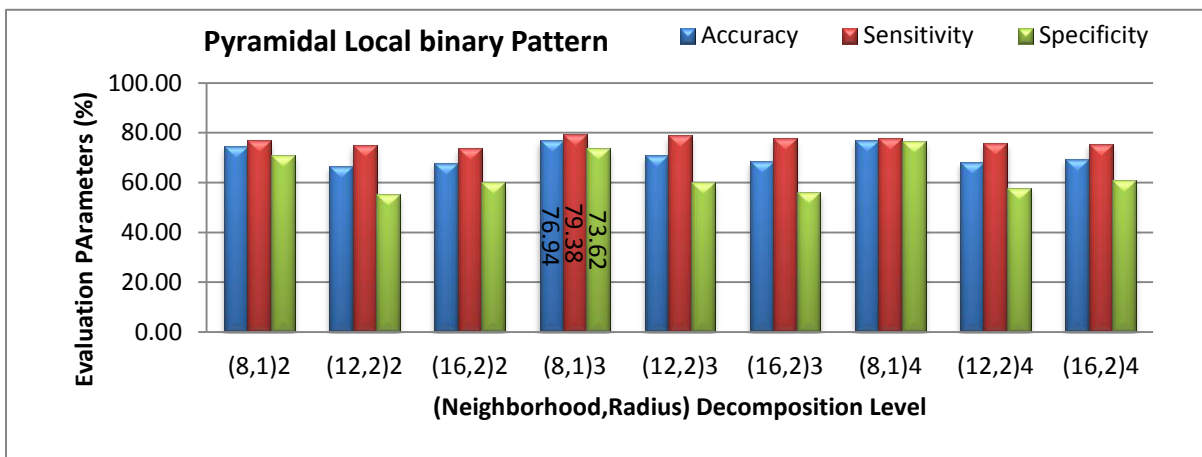


Figure 5.20: Results of Pyramidal Local binary Pattern (microcalcification-normal)

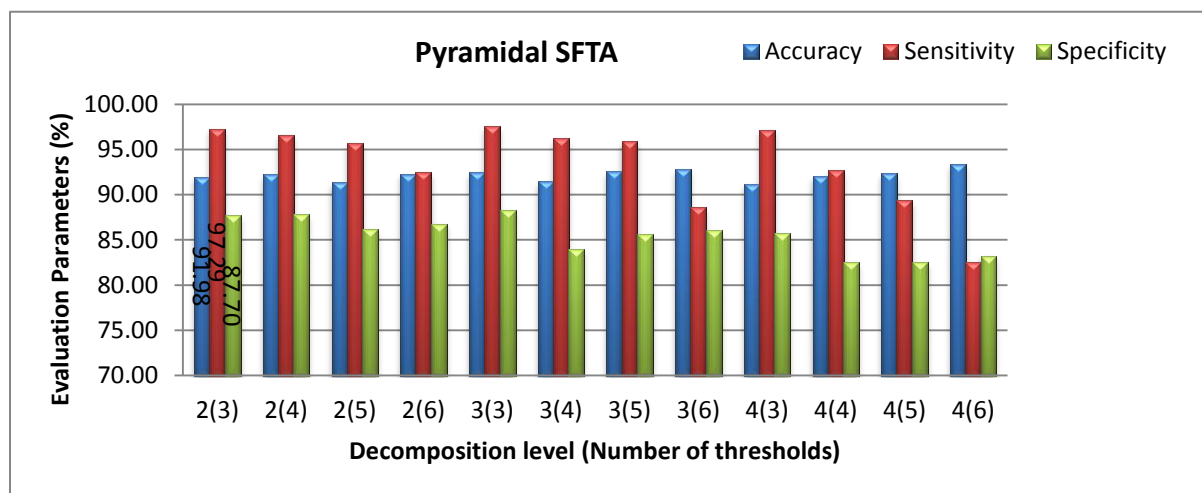


Figure 5.21: Results of Pyramidal SFTA (mass-normal)

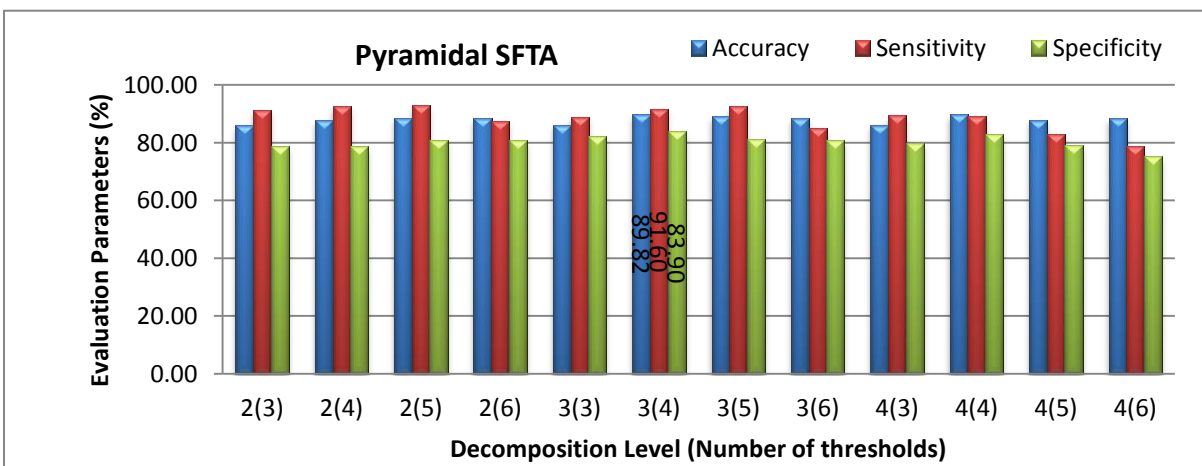


Figure 5.22: Results of Pyramidal SFTA (microcalcification-normal)

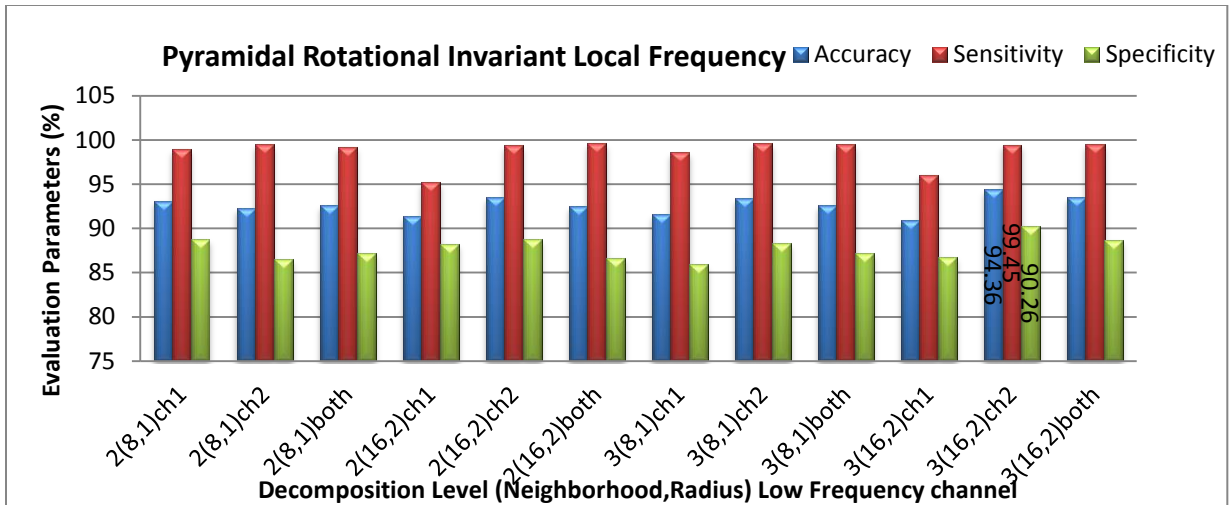


Figure 5.23: Results of Pyramidal RILF(mass-normal)

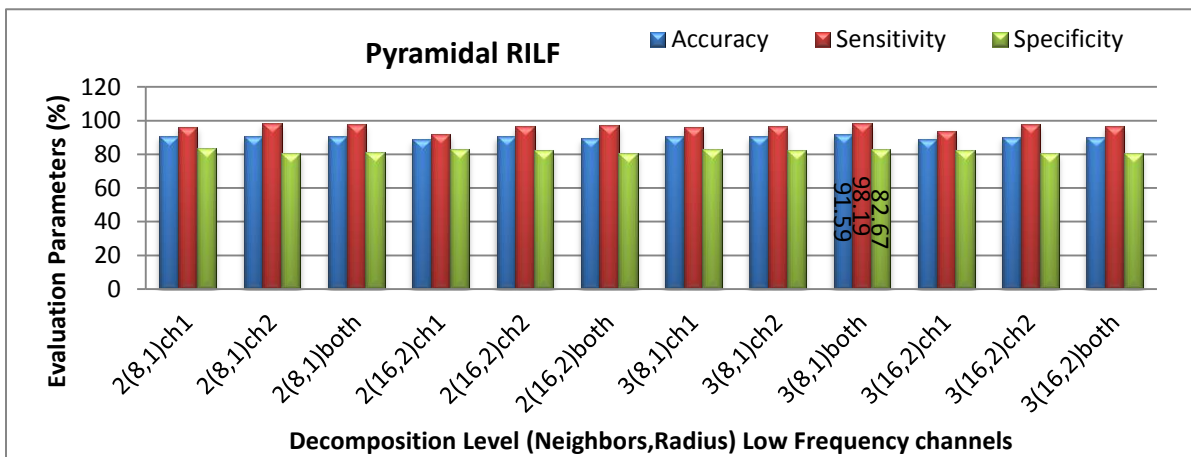


Figure 5.24: Results of Pyramidal RILF (microcalcification-normal)

5.3.5 Summary

The experimental results of the three well known techniques show that RILF gives highest accuracy with all the three methods mentioned in sections 5.3.2, 5.3.3 and 5.3.4. RILF not only dominates in the number of features compared to LBP and SFTA techniques, but also RILF technique is more robust to noise [61]. This robustness to noise is achieved by selection of low frequency channels which has maximum texture energy and less noise information. In LBP, thresholding and binomial factor 2^n on neighbors compromise textural information. To preserve this information in RILF method the gray level values surrounding the center pixel are converted

into frequency components. In LBP, by changing N and R values, the number of features exponentially increases. However, RILF features can be extracted at different scales with same number of features. In case of SFTA, the local features of the image are not considered, as the thresholding is performed globally using Otsu's method and number of thresholds are to be manually selected. We observed that in SFTA method as the number of thresholds increases, the performance is reduced drastically. Many cases were unpredicted in SFTA method as no extra texture patterns were identified with the increase in number of thresholds.

5.4 Analysis of mammograms into benign-malignant

Mass contours are defined by its shape. Malignant contours have ill-defined and spiculated shape. But, where as benign masses have well-defined and circumscribed shape. To describe the properties of masses, contour based features have been extracted to classify masses into benign-malignant. Polynomial regression, TAR, BAS and geometric features based methods are implemented on contours to extract features. Then, these features are submitted to different classifier for analysis.

5.4.1 Geometric features

Geometric features of 2D mass boundary such as area (A), perimeter (P), P:A ratio, circularity Index, L:S (ratio of the major (long) axis to the minor (short) axis of the equivalent ellipse of the mass), E: N (Elliptical normalized circumference) were extracted and evaluated using different classifiers.

5.4.2 Polynomial Regression

In this frame work we have applied polynomial fit with gradient descent algorithm. It yielded different degree of polynomial to fit the signatures of mass contours. Correlation coefficient and mean absolute error were taken as features for classification with different values of n (degree) =

7, 10, 15, 25 and 60. Performance evaluation measures were calculated by using three different classifiers 1) SVM classifier using radial basis function with sigma=0.7, 2) Linear discriminate analysis and 3) Bayes Linear classifier. Evaluation of the proposed method is carried out using 57 mammograms from DDSM database. Among them, 32 images were malignant and 25 were benign contours. These were randomly selected and separated into two sets: 29 images for training and 28 images for testing. The evaluation of our mass detection algorithm is performed by applying a hold-out methodology. The proposed new feature extraction method for delineating malignant and benign masses is evaluated by accuracy, sensitivity, specificity, confusion matrix, positive predictive value, negative predictive value and AUC

D	Table 5.4: Evaluation Parameters With SVM (%)					
	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
7	89.28	93.33	84.6199	87.5	91.666	0.927083
10	85.71	92.86	78.57	81.25	91.666	0.90625
15	96.42	100	92.31	93.75	100	0.9739
25	82.1429	86.67	76.92	81.25	83.333	0.927083
60	60.71	72.72	52.94	50	75	0.5494

D: Degree of Polynomial Hypothesis

D	Table 5.5: Evaluation Parameters With LDA (%)					
	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
7	89.2857	83.33	93.75	90.9091	88.2353	0.9583
10	89.2857	100	84.21	75	100	0.9792
15	92.85	91.67	93.75	91.667	93.75	0.9688
25	89.28	75	100	100	84.21	0.9635
60	64.2857	33	87.5	66.667	63.63	0.6458

D	Table 5.6: Evaluation Parameters With Bayes Linear Classifier					
	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
7	92.59	100	84.16	87.5	100	0.9505
10	88.88	85.71	92.31	92.3077	85.7143	0.9560
15	96.29	91.67	100	100	93.75	0.9833
25	85.18	82.35	90	93.333	75	0.9353
60	51.85	100	7.14	50	100	0.6703

Tables 5.4, 5.5 & 5.6 show the performance evaluation parameters of three different classifiers. Among this, polynomial hypothesis of degree 15 gives good recognition accuracy greater than 90% with all three classifiers. The best result for recognition accuracy (96.42%) is attained by Bayes classifier with polynomial hypothesis of degree 15. We have observed that, if the degree of polynomial increases the accuracy decreases. Figure.5.25 illustrates performance evaluation parameters of three different classifiers

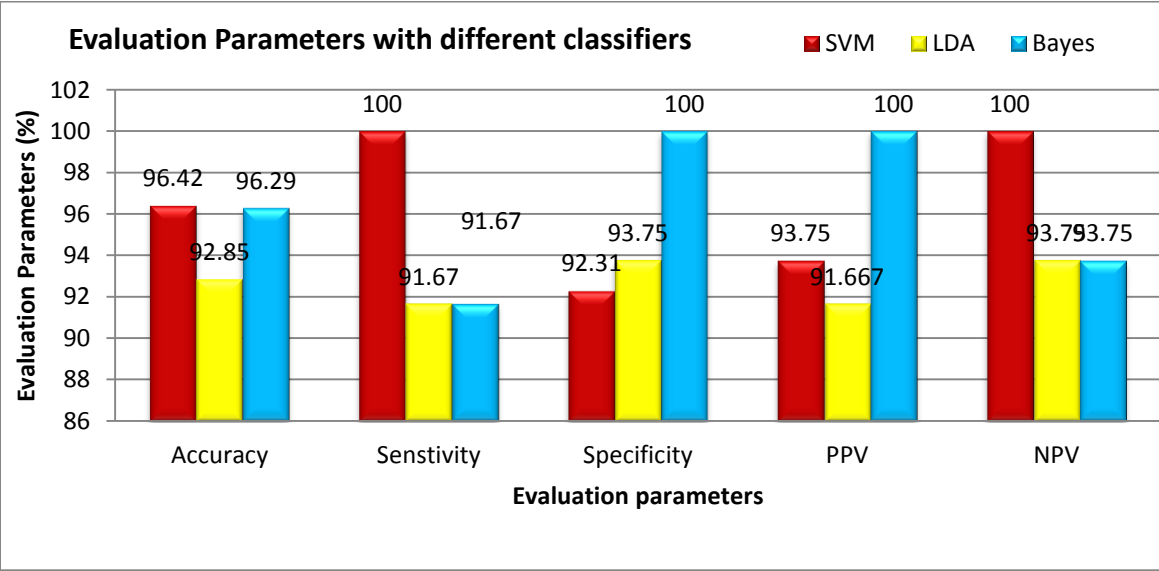


Figure 5.25: Plot of performance evaluation parameters with different classifiers

The better recognition accuracy (96.29%) and area under curve (0.9833) with our new proposed feature extraction framework is achieved with Bayes classifier.

5.4.3 Triangular area representation (TAR)

To extract shape complexity feature from the mass contour, TAR method is employed to convert 2D mass contour to 1D signature and then features are extracted from 1D signature to classify masses into benign and malignant. This method is tested on 148 mass contour images taken from DDSM database. For evaluation, hold-out methodology is applied on 126 images for training and 22 images for testing using different classifiers. The simulations have been carried out using MATLAB 2015a .The personal computer has an Intel Core i7-6500U processor and 8 GB Random Access Memory (RAM).

As TAR signature values gives number of concave and convex points, the ratio of number of concave points to number of convex points is taken as a feature vector for different values of t_s (t_s is the distance between the two points on boundary) and the ratio is represented as R which is given below in equation (5.2)

$$R = \frac{\text{Number of concave points}}{\text{Number of convex points}} \quad (5.2)$$

From the value of R, we can discriminate mass contour or boundaries. To give more discrimination power to feature vector, we have calculated the area, perimeter and eccentricity to mass contours.

We have extracted feature vectors by considering values of t_s as $t_s=1$ to3, $t_s=1$ to5, $t_s=1$ to7, $t_s=1$ to 10 and $t_s= 1$ to 20 where t_s is the distance between two neighbors. R values are extracted for each value of t_s and given to classifiers for further validation. We obtained highest accuracy of 90.9% with $t_s=1$ to 7. The length of this feature vector for each mass contour is ten. This feature vector includes seven R values which are calculated from TAR signatures and other three values

are area, perimeter and eccentricity of mass contour. Table 5.7 gives the accuracies of all five cases using SVM (linear), SVM (quadratic), simple tree, weighted KNN and ensemble adaboost classifiers.

We also considered different ratios of testing/training mass contour images in Table. 5.8 with feature descriptor obtained from $t_s=1$ to 7. The distribution of 148 mass contours for training and testing of the SVM model were performed in four different ways: first, we have used 15% for testing and 85% for training (composed of 148 mass contours with 22 testing images and 126 training images) and obtained accuracy of 90.9% with SVM (quadratic kernel). Second, we have used 30% for testing and 70% for training (composed of 148 mass contours with 45 testing images and 103 training images) and obtained accuracy of 82.2% with SVM (quadratic kernel). Third, we have used 50% for testing and 50% for training (composed of 148 mass contours with 74 testing images and 74 training images) and obtained accuracy of 82.4% with SVM (quadratic kernel) and finally we have used 40% for testing and 60% for training (composed of 148 mass contours with 59 testing images and 89 training images) and achieved accuracy of 83.1% with SVM (quadratic kernel). Therefore, from Table 5.8, we can conclude that the accuracies obtained with different number of testing images is above 80%

Table 5.7: Accuracies in(%) with different classifiers

Classifiers	$t_s=1$ to3	$t_s= 1$ to5	$t_s= 1$ to7	$t_s=1$ to10	$t_s=1$ to20
SVM(linear)	54.5	72.7	86.4	54.5	63.6
SVM(Quadratic)	54.5	72.7	90.9	54.5	50
Simple tree	63.6	63.6	81.8	59.1	72.7
Weighted KNN	72.7	59.1	81.8	63.6	63.6
Ensemble Adaboost	59.1	63.6	81.8	59.1	77.3

Table 5.8: Results achieved with different testing/training configuration using 148 mass contours

Testing/Training: (No. of images used for testing)/(No of images for training)	Acc(%)	Classifier	TN	TP	FP	FN
15/85:(22)/(126)	90.9	SVM (quadratic)	10	10	1	1
30/ 70 :(45)/(103)	82.2	SVM (quadratic)	19	18	3	5
50 /50:(74)/(74)	82.4	SVM (quadratic)	26	35	11	2
40 /60:(59)/(89)	83.1	SVM (quadratic)	21	28	8	2

Acc: Accuracy

From the above Tables 5.7 & 5.8, and as shown in Figure.5.26, we can infer that our proposed method achieved best accuracy of 90.1% using SVM (quadratic kernel) classifier and AUC value of 0.950413 with 15% of testing images and 85% of training images Figure. 5.27 gives confusion matrix i.e., True negative (TN), true positive (TP), false positive (FP) and false negative (FN) with our classifier. This figure indicates that we have obtained sensitivity of 90.9% and specificity of 90.9% with SVM classifier.

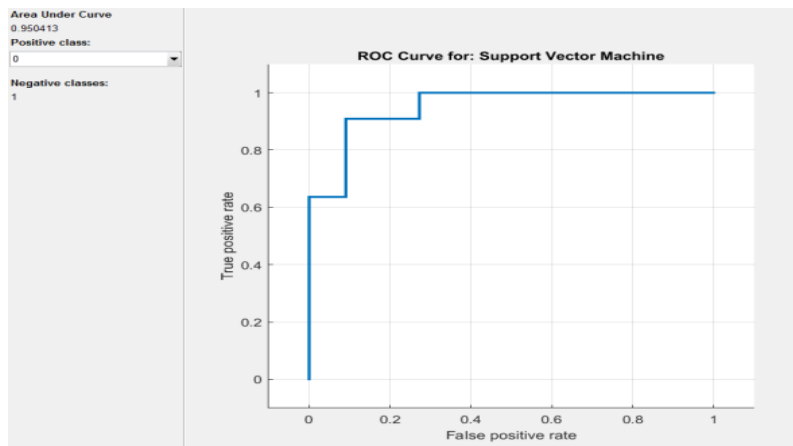


Figure.5.26: ROC analysis with SVM classifier

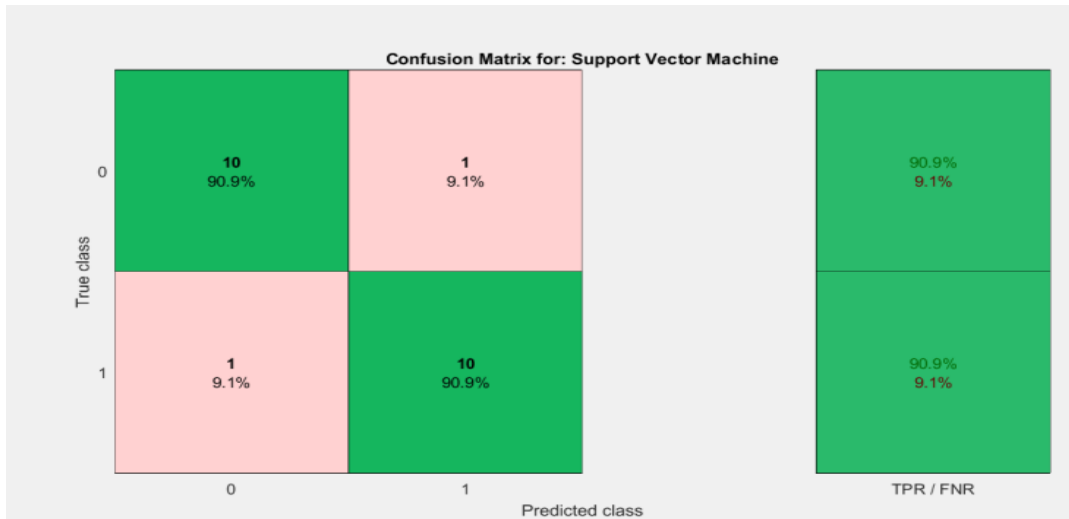


Figure 5.27: Confusion matrix of our SVM(Quadratic) classifier

5.4.4 Beam angle Statistics (BAS)

Beam angle statistics have been performed to convert 2D contour into 1D signature. This procedure describes contour features and irregularity of mass contours. We extracted features from the whole 1D signatures including Fourier descriptors (FD), detail wavelet coefficients or wavelet descriptors (WD), root mean square slope, root mean square roughness, and ratio of mean to standard deviation to classify masses.

In this method, we have analyzed the performance by compressing first two moments of 1D BAS function using Fourier descriptor, root mean square (RMS) features and wavelet coefficients. Then, performance of features extracted from 1D signatures of BAS and Centroid distance function (CDF) are compared using different classifiers and these classifiers are given in Table 5.9 and we have compared our method with state of the art methods for the classification of malignant-benign contours.

The proposed method was tested on 148 contours, which include 74 benign and 74 malignant contours. These contours were taken from digital database of screening mammography (DDSM) database.

Performance evaluation of compact features extracted from BAS descriptor: Table 5.10 presents the effect of compact (Fourier, wavelet and RMS) feature vectors. These are obtained from first and second order moments of BAS shape descriptor in selected database. These features are fed to SVM classifier for further classification. In the first case, we have truncated Fourier spectrum after ‘T’ Fourier coefficients. The different values that are taken for T are 8, 16, 32 and 64. We have obtained the highest accuracy of 80.55% for second order BAS moment with T=8 and both moments with T=64. In second case, i.e. in discrete wavelet transform method, we have considered only detail wavelet coefficients as feature vector. The reason for considering detail wavelet coefficients is that, they give high frequency content of a signal. These values are high, if they are many concave and convex points in the contour which was extracted by BAS descriptor. Too many concave and convex points illustrate the fluctuation of malignant contour. In this case also, we have truncated the wavelet coefficients to ‘T’ which can have values of 8, 16, 32 and 64. The highest accuracy achieved with wavelet coefficients is 80.55% for first order BAS moment with T=64.

Comparison of BAS and CDF: We have extracted compact feature vectors from 1D BAS moments and 1D CDF signature. Out of BAS and CDF, the results in Table 5.11 shows that the compact features extracted by BAS method gave better results than CDF using SVM classifier. It is observed that the highest accuracy achieved by CDF and BAS are 83.33% and 88.88% respectively. The reason is that 1D BAS function gathers information at all scales and gives global discriminative features to each boundary point by using all the other boundary points. In addition to that, it is insensitive to noise by preserving concave and convex points. So, it is very helpful in discriminating benign and malignant contours

Performance of 1D BAS functions Compact features with different classifiers: Table 5.12 shows the performance of three compact feature extraction methods of 1D BAS functions with four different classifiers. FD (T=16, 1st and 2nd moment of BAS) features with SVM classifier achieved highest accuracy of 83.33%. WD (T=64, 1st moment of BAS) features achieved highest accuracy of 86.11% with KNN classifiers. RMS features (rms slope, rms roughness and mean to standard deviation ratio (M/sd)) (1st and 2nd moment of BAS) achieved highest accuracy of 88.88% with SVM classifiers. Out of all the three types of features, RMS features gave better accuracy with all classifiers. Among all the classifiers SVM gave better accuracies for FD and RMS features.

Summary: This section gives the summary of the method used to classify benign and malignant contours. This work focus is on the study of 1D shape descriptor that distinguishes benign and malignant masses. At first, the 2D mass contours are converted into one dimensional BAS moment functions. These 1D BAS functions give perceptual information using statistics of beam angles. The valleys and hills of the first moment of BAS function correspond to concavities and convexities of the object shape [75]. The 1D BAS function are further compressed into FD, WD and RMS features (rms slope, rms roughness and M/sd ratio). These features are fed to different classifiers for further validation. The proposed method achieved an accuracy of 88.8% for benign-malignant classification with RMS features and SVM classifier. To further evaluate our quality of work, we have compared BAS with CDF method. CDF is another method which transforms 2D contour into 1D signature. It is the only one dimensional transformation method applied on mass contours till now in the literature [74].

Table 5.9: Different Classifiers and its input parameters

Serial No	Classifiers and its parameters
1	SVM(Gaussian kernel size=5, Box constraint =0.61)
2	KNN, distance metric =square inverse, correlation, number of neighbors =10
3	Adaboost decision tree, number of learners=100, learning rate=0.2
4	ANN, scaled conjugate gradient back propagation network, Number of hidden neurons=10

Table 5.10: Accuracies of 1D BAS functions with FD, DWC and RMS

Feature Descriptors	Parameters	1 st moment of BAS	2 nd moment of BAS	Both
Fourier Descriptor	T=8	72.22	80.55	77.77
	T=16	69.44	77.77	72.22
	T=32	66.667	75	75
	T=64	75	80.55	80.55
Wavelet Descriptors	T=8	58	52.77	72.22
	T=16	72	72.22	75
	T=32	69.44	66.66	63.88
	T=64	80.55	63.88	77.77
RMS	RMS roughness	69.44	66.66	72.22
	RM's slope	55.55	63.88	66.66
	M/sd	66.667	52.77	58.33
	All	80.55	77.77	88.88

Table 5.11: Comparison of accuracies with 1D BAS and CDF functions with FD, WD and RMS features

Compact features from BAS	Parameters	CDF (%)	Both moments of BAS (%)
Fourier Descriptor(FD)	T=8	69.44	77.77
	T=16	66.66	72.22
	T=32	72.22	75
	T=64	70	80.55
Wavelet Descriptors	T=8	66.66	72.22
	T=16	63.88	75
	T=32	58.33	63.88
	T=64	61.11	77.77
RMS	RMS roughness	50	72.22
	RM's slope	83.33	66.66
	M/sd	80.55	58.33
	All	75	88.88

Table 5.12: Classification accuracies with different classifiers

Classifiers	RMS	FD(T=16,1 st moment)	WD(T=64, 1 st moment)
SVM(Gaussian kernel size=5, Box constraint =0.61)	88.8	80.55	80.55
KNN	80.55	72.22	86.11
Adaboost decision tree	83.33	72.22	83.33
ANN	83.8	75 T=16	78.4

5.4.5 Comparison between all the developed methods

In this section, we have compared all the developed methods in sections 5.4.1, 5.4.2, 5.4.3 and 5.4.4 using 148 mass contours. Among them, 74 are benign contours and 74 are malignant contours. Different feature extraction techniques that are developed such as geometric features, polynomial regression, TAR and BAS methods are compared and it is shown in Tables 5.13.

Tables 5.13 show the comparison of the developed methods. The evaluation for comparison is carried out with training to testing ratio which is equal to 70:30. Different classifiers such as SVM, KNN, adaboost decision tree and simple tree have been implemented to classify mass contour into benign-malignant. From the Table 5.13, we can observe that TAR method achieved highest accuracy of 90.9% using SVM classifier.

5.4.6 Comparisons of our developed methods and other existing methods

To validate the efficacy of our proposed methods for benign-malignant classification, we compared the accuracy of our methods with the existing methods. These comparisons are shown in Table 5.14.

5.5 Implementation of above algorithms on complete mammograms

In all the above sections, we have presented results by considering complete mammograms for image enhancement and segmentation. Individual ROIs and mass contours were manually extracted for the feature extraction and classification. In this section, we have developed a method for automatic identification and analysis of mass with image enhancement, segmentation, feature extraction and classification methods discussed in chapters 2, 3 and 4. We have considered ten mammograms taken from DDSM database for implementation of the above process. Figures 5.28 and 5.29 gives the block diagram for this complete process and shows the output of all the methods for only three mammograms respectively. The ROIs are manually cropped from enhanced mammograms and features are extracted using LBP, SFTA and RILF techniques to classify mammograms ROIs into normal-abnormal and achieved highest accuracy of 100% using BRILF as shown in Table 5.15. The geometric, polynomial regression, TAR and BAS techniques are used to classify contours of segmented masses into benign-malignant and achieved highest accuracy of 100% using TAR and it is shown in Table 5.16.

Table 5.13: Comparison of benign-malignant feature extraction techniques

Classifiers	Methods	Accuracy	Parameters with specific values
SVM	GEO	62.2	Kernel: Linear Boxconstraint:1.2
	POLY	81.3	Kernel :Linear BoxConstraint:1.2
	TAR	90.9	Kernel: Linear Boxconstraint:1.2
	BAS	88.8	Gaussian kernel size=5, Box constraint =0.61
KNN	GEO	73	Number of neighbors = 4 Distance Weight= Euclidean distance Distance Metric = Square Inverse
	POLY	75.7	Number of neighbors = 10 Distance Weight= Euclidean distance Distance Metric = Equal
	TAR	81.8	Number of neighbors = 4 Distance Weight= Euclidean distance Distance Metric = Square Inverse
	BAS	80.55	Distance metric =square inverse, correlation, number of neighbors =10
Adaboost Decision tree	GEO	64.9	Number of Learners= 175 Learning Rate=0.2
	POLY	73	Number of learners =200 Learning Rate=0.2
	TAR	81.8	Number of Learners= 175 Learning Rate=0.2
	BAS	83.33	Number of learners=100, learning rate=0.2
Simple Tree	GEO	67.6	Maximum Number of Splits:16 Split Criterion: Maximum Deviance Reduction
	POLY	73	Maximum Number of Splits:4 Split Criterion: Twoing
	TAR	81.8	Maximum Number of Splits:16 Split Criterion: Maximum Deviance Reduction
	BAS	83.8	No of splits=20 Split criteria GDI, , surrogates=10

Table 5.14: Comparison of benign-malignant feature extraction techniques with the state of the art methods

Feature extraction method	Database	Samples	Acc	Sens	Spec	AUC
GaborPCA [92]	MIAS	114	80%	-	-	-
Degree of spiculation of a mass and relative gradient orientation of pixels that correspond to possible spicules[93]	DDSM and MIAS	319	81%	-	-	-
Fractional concavity and spiculation index [94]	MIAS	111	82%			0.79
Fractal dimension using ruler method and fractional concavity [40]	MIAS	111	-	-	-	0.82
Geometric Features and ensemble classifier [95]	DDSM	-	72%			
LTP [28]	DDSM	-	-	-	-	0.97
Zernike moments[96]	DDSM	-	96.43%	-	-	0.97
Proposed methods	DDSM					
Geometric features		148	73			
Polynomial Regression		148	81.3			
Triangle Area Representation		148	90.0			
Beam Angle Statistics		148	88.8			

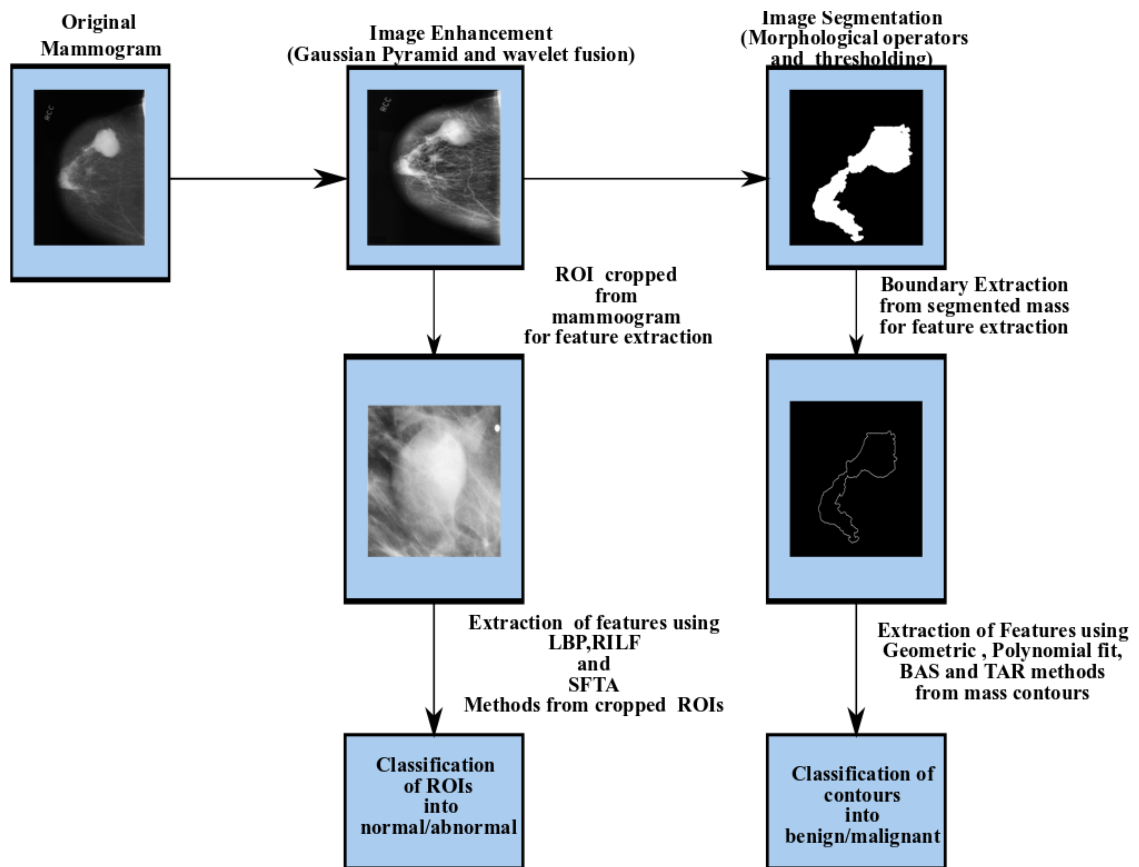
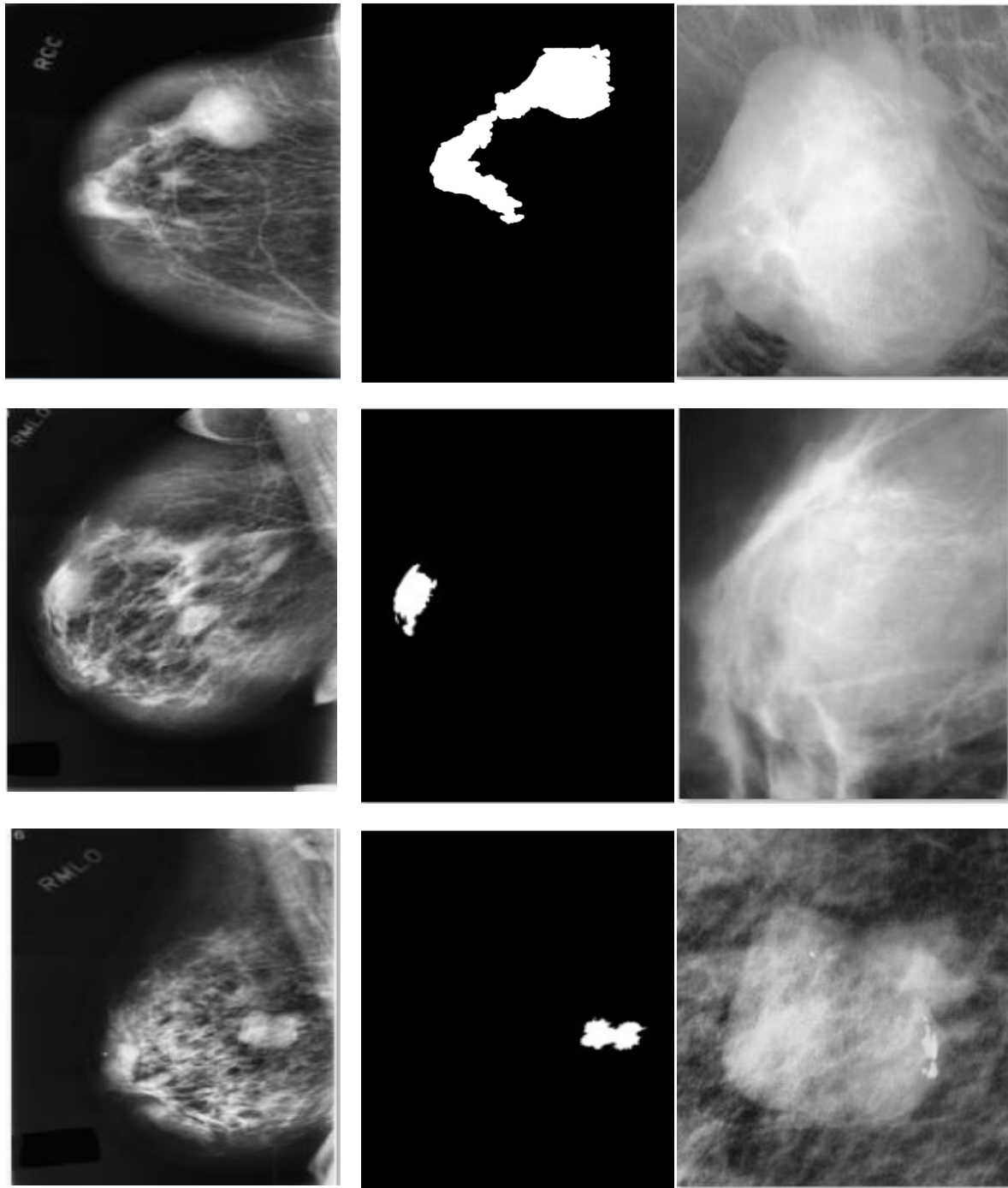


Figure 5.28: Block diagram for identification and analysis of mass in mammograms



(a)

(b)

(c)

Figure 5.29: a) Enhanced Mammograms, b) Segmented Masses for benign-malignant classification, c) ROIs for normal-abnormal classification

Table 5.15 : Accuracy for the normal-abnormal classification using complete mammogram with the proposed methods

Method	Parameters	Accuracy (%)
LBP	(16,2)u2	78.97
LTP	(8,1)u2	95.68
LQP	(12,2)riu2	94.11
SFTA	th =3	88.37
RILF	(8,1)both	84.77
BLBP	(8,1)2*2 riu2	82.86
BSFTA	2*2(3)	88
BRILF	4*4(16,2)ch2	100
PLBP	3(8,1)	85.97
PSFTA	2(3)	96.17
PRILF	3(16,2)ch2	84.96

Table 5.16: Accuracy for the benign-malignant classification using complete mammogram with the proposed methods

Methods	SVM(linear)
Geo	80%
poly (Degree =7)	90%
BAS(RMS features from 1D BAS functions)	93%
TAR(ts = 1 to 7)	100%

5.6 Conclusion

In this chapter, the results and analysis of all the developed methods have been discussed. The results include the validation of image enhancement and segmentation methods. Further, the performance of LBP, SFTA and RILF features extraction techniques have been tested for normal-abnormal classification. The performance of geometric features, polynomial regression, TAR and BAS methods were also tested for the classification of mass contours into benign-malignant and finally we identified and analyzed mass automatically with all the methods discussed in chapter 2, 3 and 4. All these proposed methods are validated using mammograms taken from DDSM, IRMA and KIMS. The results are compared with the existing methods.

Chapter 6

Automatic detection and analysis of abnormality in Mammogram

Automatic detection and classification of masses is a key issue for breast cancer diagnosis in digital mammograms. It helps the radiologists in detecting the masses accurately in less time.

In this chapter, a heuristic method is developed for detection and analysis of mammograms. A block diagram of the developed model is shown in Figure 6.1. For the automatic detection of mass, we applied Dunn index based automatic k-means (DAK) algorithm. In this method, the number of clusters for individual image is calculated based on the Dunn index value and the cluster with abnormality is detected using homogeneity and entropy without any user intervention. For feature extraction step modified segmented fractal texture analysis (MSFTA) features are extracted. The main advantage of this feature extraction method is that it generates less number of features to decrease computational time. The extracted features are given to SVM classifier for classifying the detected abnormality into mass-non-mass and benign-malignant. Further, three class system was also developed to classify the abnormalities into non-mass-benign-malignant.

The rest of the chapter is organized as follows: pre-processing of mammogram is described in section 6.1. Section 6.2 explains the segmentation of mass using DAK algorithm. The extraction of features using MSFTA is explained in section 6.3 followed by classification in section 6.4. Section 6.5 describes the experimental results of our developed work in this chapter and further conclusion is given in section 6.6.

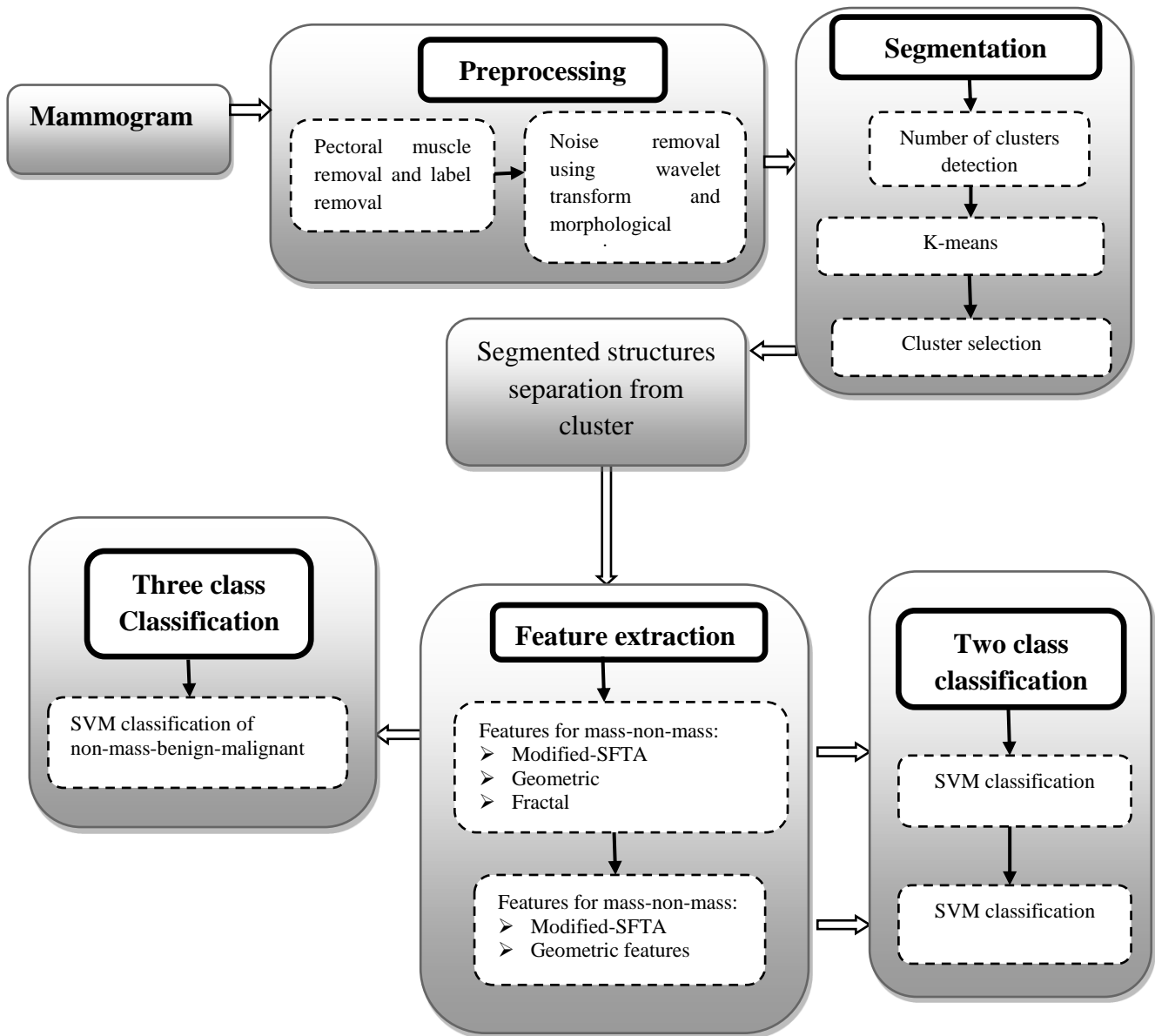


Figure 6.1: Block diagram of the proposed model for automatic detection and analysis

6.1 Preprocessing

Mammogram preprocessing helps to detect abnormalities in less time. Mammograms contain noise during the acquisition process and some portions of breast regions contain labels and background which is superimposed with mammogram. Thus, mammograms require initial pre-

processing to remove the noise and to extract breast area. The following preprocessing steps are used and explained below.

6.1.1 Breast area region extraction and label removal

Mammograms cover only 30% of breast area. The computation time for analysis increases, if the complete mammogram is considered. Due to this, breast area is segmented by applying global thresholding technique for further processing. A label in mammogram also affects the results of mass detection method. So, these labels are removed during the global thresholding process. Figure 6.2a shows the original mammogram.

6.1.2 Pectoral muscle removal

Pectoral muscle is a dense region in mammogram which is attached to the chest wall. The pectoral muscle possesses high intensity values as that of mass which results in a higher rate of false positives. Thus, removal of pectoral muscle is an important step. The pectoral muscle is removed by identifying the breast orientation and then using connected component labeling algorithm [82].

6.1.3 Noise removal

The mammograms contain noise during acquisition. So, the noise is removed by applying wavelet denoising, guided filter and adaptive histogram equalization. Along with the removal of noise, mammogram edges are also enhanced by morphological operations to highlight abnormal areas in mammogram. The matlab code for mammographic enhancement is taken from [83]. The preprocessed mammogram with all these three steps is shown in Figure 6.2b

6.2 Segmentation of suspicious areas

In this study, suspicious areas are segmented using Dunn index based automatic k-means (DAK) algorithm. It uses the traditional unsupervised k-means clustering algorithm for segmentation

[84]. The two advantages that separate k-means and DAK algorithm are, (i) the number of clusters and the dense cluster with mass is selected without any manual help and (ii) the clustering was carried out according to the local conditions (3*3 neighbourhood) of the image.

The three steps for DAK algorithm is discussed below.

- In the first step, for each pixel p in the image, 3*3, sliding window is placed such that the pixel p coincides with the center location of the window. Intensity based features (mean, entropy and skewness) were calculated inside the window of each pixel. The mean, entropy and skewness were chosen because mass pixels and its surrounding values are brighter and mass regions have very little contrast. This results in high mean value, positive skewness and low entropy values in that mass region. The size of the window 3*3 was selected because, we observed that the shape, size and texture of the mass depend on its surrounding pixels. The three intensity based features were calculated for each pixel in 3*3 neighbourhood region. At the end, matrix (M) of size $L*3$ is obtained, where L is the number of pixels in the image.

- The second step involves the selection of number of clusters (k) for grouping the matrix M . The value of k in this step is selected based on the Dunn index. It is defined as the ratio of minimal inter-cluster distance to maximal intra-cluster distance [85]. Unsupervised k-means algorithm is applied on the matrix (M) by varying the number of clusters in the range from 3 to 9. At the end of each run, Dunn index is evaluated. Then optimal k is chosen as the first maximal peak of Dunn index values and further matrix M is grouped using k-means clustering with optimal k value. The output of k-means clustering is shown in Figure 6.2c. The graph of Dunn index vs no of clusters is shown in Figure 6.3.

- Third step involves the selection of a dense cluster with mass. The selection of dense clusters involves the concept of homogeneity and entropy. Each cluster obtained from DAK algorithm is

analysed by homogeneity and entropy features. Homogeneity measures the presence of intra-regional changes in the cluster and entropy reflects the non-uniformity and complexity in the cluster [86]. The expression for homogeneity and entropy are given below in equations (6.1) and (6.2) Figure 6.2d shows the dense cluster with mass.

$$\text{Homogeneity} = \sum_{i=1}^G \sum_{j=1}^G \frac{\text{cooc}(i,j)}{1+|i-j|} \quad (6.1)$$

$$\text{Entropy} = \sum_{i=1}^G \sum_{j=1}^G \text{cooc}(i,j) \log(\text{cooc}(i,j)) \quad (6.2)$$

Where $\text{cooc}(i,j)$ is a co-occurrence matrix of image with maximum gray level value of G. First the two clusters whose homogeneity values less than other clusters were selected. In the next step, the cluster with high entropy value is selected as a dense cluster having mass. Usually the dense cluster has more texture information and less local changes inside the cluster. Hence, the dense cluster has maximum entropy and minimum homogeneity than other clusters. Flow chart in Figure 6.4 explains the process of selecting dense cluster with mass.

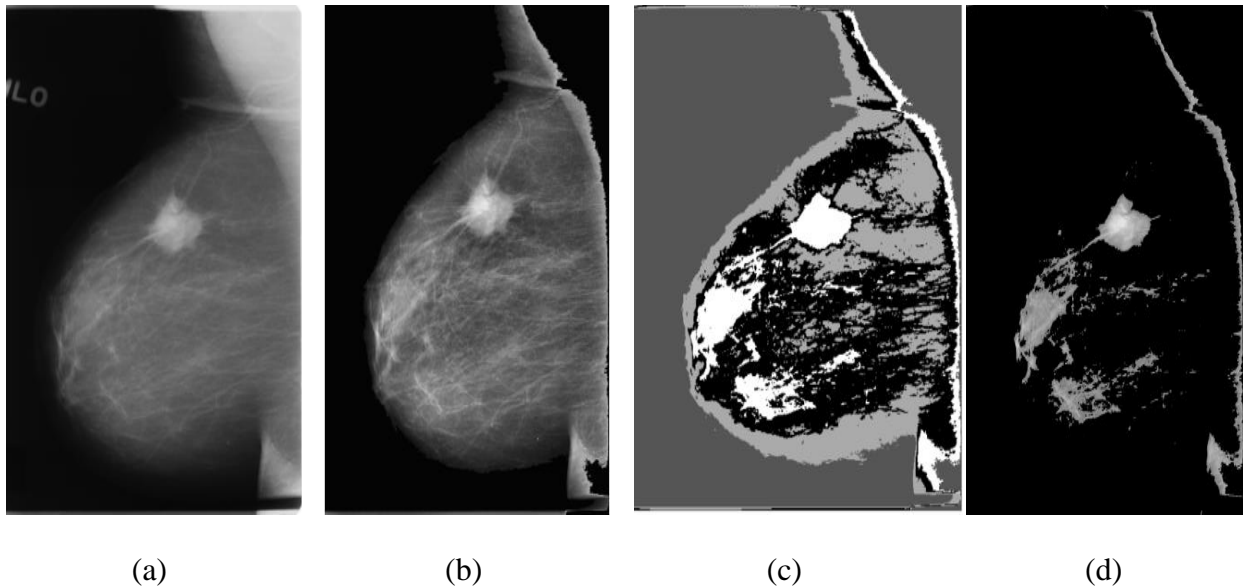


Figure 6.2: a) Original mammogram, b)Pre-processed mammogram, c) Output after DAK algorithm, d) Dense cluster with mass

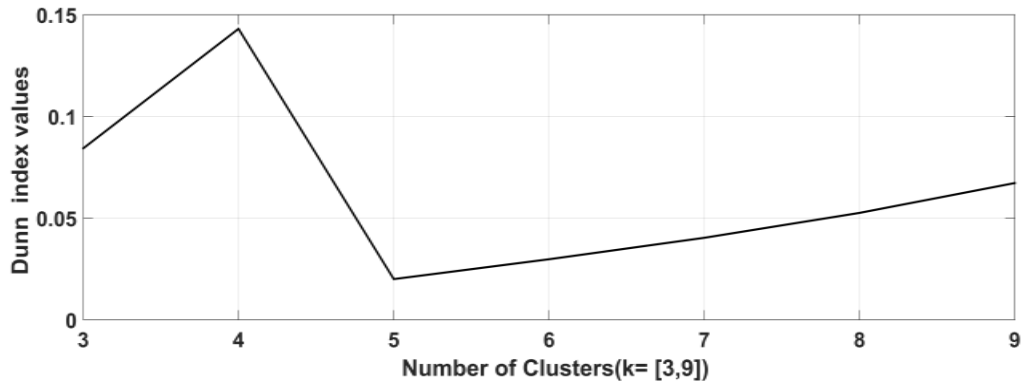


Figure 6.3: Plot of Dunn index values vs number of clusters

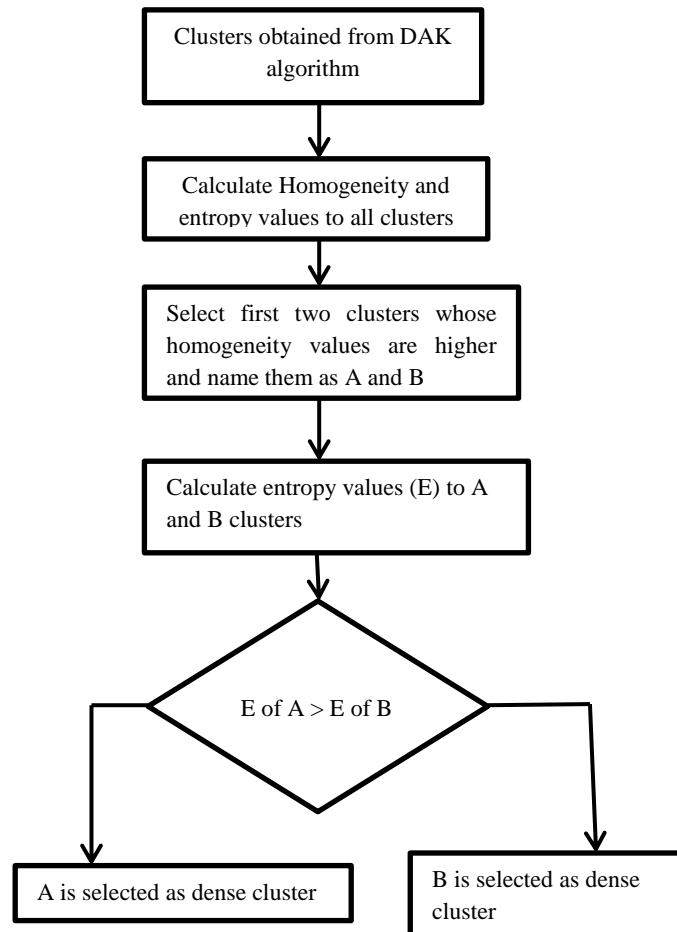


Figure 6.4: Flow chart for the selection of Dense cluster

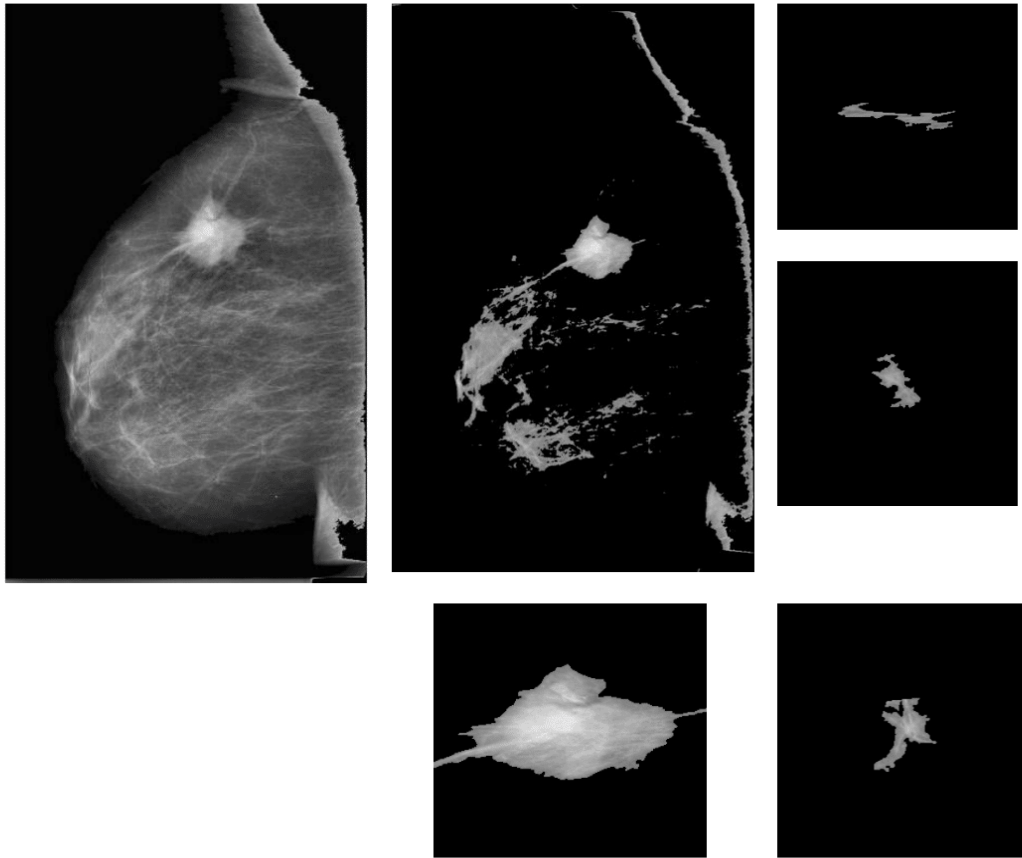


Figure.6.5: Original mammogram, dense cluster, suspicious areas selection from dense cluster

6.3 Feature extraction

The dense cluster obtained from DAK has many segmented structures along with mass as shown in Figure.6.5. These are to be analysed for further to classification of segmented structures into mass-non-mass and mass in dense cluster into benign-malignant. This section gives the description of MSFTA feature extraction method.

6.3.1 MSFTA

MSFTA method uses the basis of segmented fractal texture analysis (SFTA) for feature extraction which is discussed in chapter 3. The main advantage of SFTA is it gives texture information from both interior and exterior regions of segmented structures. We have

implemented SFTA with some changes and named it as MSFTA. The procedure for MSFTA method is discussed below.

Step 1: The steps for extraction of features using MSFTA is almost the same as SFTA method. First step involves decomposing of segmented structures into binary images using Otsu thresholding, which is similar to SFTA. However in the second step, additional features have been extracted in order to add more discrimination power to our feature vector.

Step 2: Features like fractal dimension using box counting method, pixel count from binary images and mean gray value from the original image were computed in SFTA method. But in MSFTA method along with SFTA features, we also extracted standard deviation as an extra feature from gray-level values of original. We can obtain only three feature vectors from decomposed binary images (fractal dimension, pixel count and mean gray-level value) using SFTA method. So, the size of feature vector in SFTA method is $n_t \times 6$ (n_t is number of thresholds). But, after applying MSFTA, the size of the feature vector is $n_t \times 8$ as we have computed extra features from binary images and segmented gray level image.

6.3.2 Feature extraction for mass-non-mass classification

Mass is more circular and thicker where as non-mass is thinner and longer [87]. MSFTA features were computed to extract texture information from segmented structures. Further geometric [88], polynomial regression and fractal features [89] were also calculated to define shape characteristics of segmented structures. The following are the different types of features used for mass-non-mass classification

MSFTA features: MSFTA features are extracted from segmented structures by considering $n_t = 3$. The best value of n_t was selected empirically by testing with different values in range from 3 to 8. The reason for considering this range is that, as the number of thresholds (n_t) increase, no

extra texture patterns were identified and accuracy had reduced drastically with MSFTA method. The number of features obtained with MSFTA for mass-non-mass is equal to twenty four.

Polynomial regression Features

Polynomial regression features give the shape characteristics of mass. So, we have computed polynomial regression features with 1D signature of segmented structures boundary and the method was discussed in section 4.3 of chapter 4

Geometric Features: The geometric features were extracted to distinguish segmented structures into mass-non-mass. The geometric features are area, perimeter, solidity, maximal radial length, bending energy and thinness ratio to quantify shape characteristics of the segmented structures.

Fractal Features: Fractal analysis of a contour or boundary helps to find roughness and complexity of contours. So, we have calculated fractal features from the boundary of segmented structures. The following three fractal features from segmented structures are

1. 2D Hausdroff dimension
2. 1D fractal dimension using box counting method
3. 2D fractal dimension using box counting method

6.3.3 Feature Extraction from breast masses for malignant-benign classification

Breast masses show different gray-level characteristics and shapes. A typical benign mass is round with smooth and well-defined (well-circumscribed) boundary, whereas a malignant mass is spiculated and rough with a blurry boundary. However, for any breast mass, the characteristics of its interior and exterior parts are different [90]. So, MSFTA features are extracted from masses to classify them. The main advantage for these features is the computation of gray-level characteristics from interior and exterior regions of mass with multithreshold segmentation. In

addition to MSFTA features, we also calculated geometric features to extract shape information. The following are the features extracted to classify benign and malignant masses

MSFTA features: MSFTA features are extracted from masses with n_t (number of thresholds) values equal to 3 and 8. The n_t was chosen after many tests by varying the values from 3 to 8. We obtained better accuracy by fusing the MSFTA features obtained with two n_t values that are 3 and 8 with geometric features. The number of MSFTA features computed for each mass is equal to eighty eight.

Geometric features: The extracted geometric features and their definitions are given below

Solidity(S) is the ratio of the area within shape to the ratio of convex hull area and it is given by equation (6.3)

$$S = \frac{\text{Area}}{\text{Convex hull Area}} \quad (6.3)$$

Maximal radial length (MRL) is the maximum Euclidean distance from the centroid of mass to the pixel on the boundary and it is given by the equation (6.4)

$$\text{MRL} = \max(r(k)) \quad (6.4)$$

$r(k)$ is the Euclidean distance from the centroid of mass to the pixel on the boundary. In the above equation k varies from 1, 2, 3... N and N is the number of points on the contour.

Bending Energy (BE) is the energy stored in the contour and it is given by equation (6.5)

$$\text{BE} = 1/N \sum_{s=0}^{N-1} (K(s)) \quad (6.5)$$

Where $K(s)$ denotes the curvature of the shape parameterized by arc length. N is the number of points on a boundary and s is the arc length parameter [91].

Thinness ratio (TR): It distinguishes line from other shapes.

$$\text{TR} = \frac{4*\pi*Area}{perimeter^2} \quad (6.6)$$

6.4 Classification

Three types of classification have been carried out for the automatic detection and analysis of abnormality in mammogram. One classification, distinguishes the segmented regions into mass-non-mass. Second classification delineates the segmented mass regions into benign-malignant. Third classification distinguishes segmented structures into three classes that are non-mass-benign-malignant.

6.5 Experimental analysis

6.5.1 Data set

To evaluate the proposed CAD system, we have considered data set of 120 abnormal mammograms having at least one mass. From the data set, mass was detected only in 85 mammograms. The accuracy of detecting actual mass was 70.8% with DAK. In other 35 mammograms, the mass was detected, but it was connected to other high density structures. Further, 352 segmented structures (88 masses and 264 non-masses) are separated out from 85 high density clusters obtained from DAK algorithm Hence, we have considered 88 segmented structures with mass and 264 segmented structures with non-mass for further classification In addition to segmentation of high density cluster, features are extracted from segmented structures and submitted to SVM classifier for validation. The classification was performed using ten-fold cross validation method. The best classification rate for mass-non-mass is obtained by using a combination of features shown in Table 6.1

Table 6.1: Features for mass-non-mass

Features	Parameters
Modified SFTA	Number of thresholds ($n_t=3$)
Geometric, Fractal 1D, 2D	-
Polynomial Regression	Degree=15

6.5.2 Evaluation for mass-non-mass classification

Two class classification for mass-non-mass: To evaluate the effectiveness of our developed method we have computed accuracy, sensitivity, specificity and AUC (Area under curve). Six different feature sets were calculated for the 352 segmented structures. Those feature sets are polynomial regression features (P), geometric features (G), fractal features (F), MSFTA features ($n_r=3$), OSFTA-PGF (original SFTA combined with polynomial, geometric and fractal features) and MSFTA-PGF (MSFTA combined with polynomial, geometric and fractal features). The average accuracies with all these feature sets were calculated and shown in Table.6.2 These computed accuracy values are 71.77%, 80.76%, 80.07%, 87.45%, 93.02%, and 97.93 % for above mentioned six feature sets respectively. Sensitivity and specificity values of six feature sets were compared in Figure.6.6.

As seen in Table.6.3, we can observe that our method has achieved 98% sensitivity, 97.75% specificity, 97.93% classification accuracy and from Figure.6.7 the AUC value obtained from ROC curve is 0.9921. These values are obtained with the MSFTA-PGF feature set. The reason for attaining the highest accuracy with MSFTA-PGF features because of the additional gray level characteristics of MSFTA features. However, geometric, polynomial regression and fractal features gave shape information which added more discriminative power to our feature set. Thus, the combined feature set (MSFTA-PGF) performed well compared to other feature sets.

Table.6.2: Average classification accuracy(%) with different features sets using tenfold cross validation with SVM for mass-non-mass

Classifier folds	Poly10(P)	Geo(G)	Modified SFTA (MSFTA)	Frac (F)	Original SFTA+P+G+F (OSFTA-PGF)	Modified SFTA+P+G+F (MSFTA-PGF) Proposed
1.Fold	71.88	81.82	80.65	96.88	100.00	100.00
2.Fold	70.31	80.00	84.13	92.19	92.06	100.00
3.Fold	68.75	80.41	82.98	91.67	92.63	100.00
4.Fold	70.31	80.62	80.16	90.63	91.34	98.43
5.Fold	71.88	80.25	78.48	88.75	92.50	98.11
6.Fold	75.00	79.90	79.58	89.06	91.10	97.38
7.Fold	72.77	81.33	79.37	89.33	91.93	97.32
8.Fold	71.88	81.71	78.52	89.88	92.55	96.88
9.Fold	72.47	80.97	78.13	90.63	93.03	96.18
10.Fold	72.50	80.63	78.75	90.63	92.81	95.00
Average	71.77	80.76	80.07	87.45	93.02	97.93

Table. 6.3: Average classification accuracy ,sensitivity and specificity with MSFTA-PGF feature set (%)

Fold	Accuracy1	Sensitivity1	Specificity1
1fold	100.00	100.00	100.00
2fold	100.00	100.00	100.00
3fold	100.00	100.00	100.00
4fold	98.43	98.91	97.14
5fold	98.11	98.26	97.73
6fold	97.38	97.10	98.11
7fold	97.32	97.53	96.77
8fold	96.88	96.76	97.18
9fold	96.18	96.15	96.25
10fold	95.00	95.26	94.32
Average	97.93	98.00	97.75

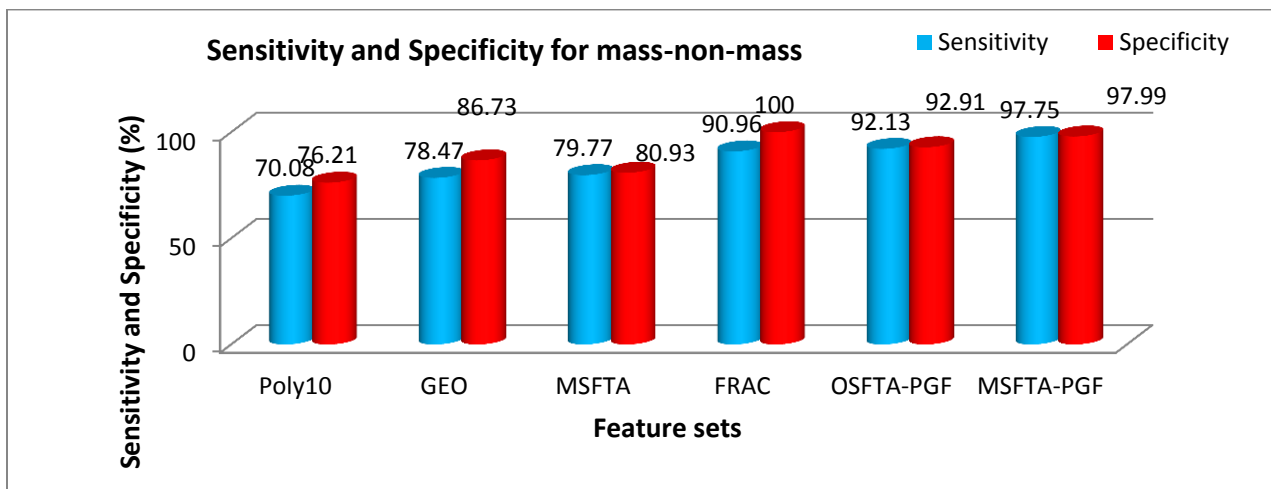


Figure.6.6: Plot of sensitivity and specificity for different feature sets

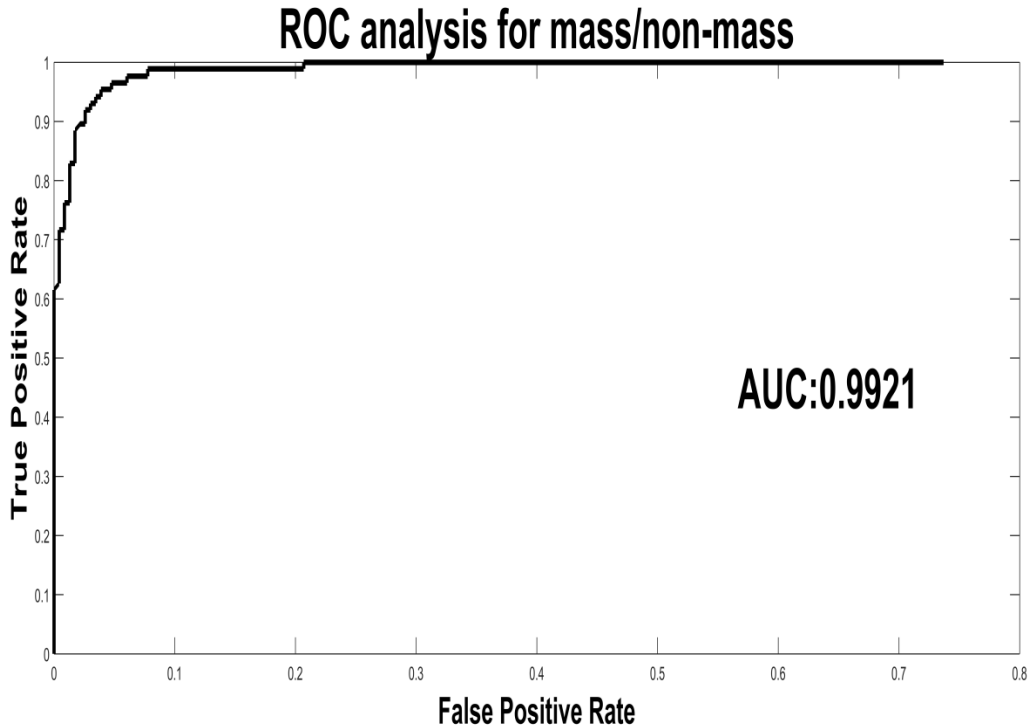


Figure 6.7: ROC analysis for mass-non-mass

Comparison of with the existing methods: To evaluate the efficiency of the developed method, it is tested with existing algorithms with our dataset using ten-fold cross validation. However, different methods use different database and number of images for evaluation. The selected methods are Local binary pattern (LBP) [27] and Haralick method [45]. The evaluation parameters used for effective comparison were accuracy, sensitivity, specificity, and AUC and computational time. The computational time is calculated for a personal computer with Intel core i7-4770K CPU, 64-bit operating system and 12GB Random Access Memory (RAM). The evaluation process for this comparison is carried out using MATLAB 2015a.

The comparison for mass-non-mass classification is done with LBP [27] and Haralick [45] methods. For LBP, variants specified in paper were tested on manually cropped regions of the mammogram which does not include segmentation step. Moreover, only gray level characteristics of segmented structures are calculated using LBP method.

In Haralick method [45], GLCM and shape based features were extracted with different orientation scales and directions. This method includes segmentation step to extract features from multiple orientations. Due to this long process, it requires high computation time. Tables 6.4 and 6.5 shows that classification performance in terms of accuracy, sensitivity, specificity and AUC is high for the developed method. Table 6.6 reveals that the time taken by the developed method for extracting features from one image is slightly greater than LBP method and the difference is 0.0028s which is acceptable and is better than Haralick method.

Table 6.4: Comparison for benign-malignant classification in (%) using SVM (linear) kernel with different methods

Methods	Accuracy	Sensitivity	Specificity
LBP	81.33	93.7	48.8
Haralick +Geo	78.5	79.15	76.79
Proposed method	97.93	98.00	97.75

Table.6.5: AUC with different methods for mass-non-mass

Feature extraction Methods	AUC
LBP	0.814
Haralick	0.836
Proposed	0.9915

Table 6.6: The computational time evaluation (mass-non-mass)

	Time in seconds/ image	No of features
LBP	0.2008	1888
Haralick	77.8908	291
Proposed	0.2036	37

6.5.3 Evaluation for benign-malignant classification

Two class classification for benign-malignant: The developed method was implemented on 88 segmented structures for mass classification. These segmented structures contain 35 benign and 53 malignant masses. For mass classification, ten-fold cross validation is carried out using SVM classifier. The best classification rate for benign-malignant is obtained by using a combination of

features shown in Table 6.7. The developed feature set consists of MSFTA features with a number of thresholds ($n_t= 3$ and 8) and geometric features. We classify our segmented structures using four feature sets and they include: geometric features, MSFTA features, original SFTA-geometric features (OSFTA-G) and MSFTA-geometric features (MSFTA-G). The performance of our approach is validated by determining accuracy, sensitivity and specificity. The average classification accuracies of all these feature sets are 78.17%, 55.05%, 76.99% and 88.67 % respectively as shown in Table 6.8. The comparison of sensitivity and specificity with these four feature sets is shown in Figure.6.8. From Table.6.8, we can observe that average classification rate is 78.17% with MSFTA features. However, by the addition of geometric features to the MSFTA features, classification accuracy had increased significantly by 10%, i.e 88.67% respectively as shown in Table.6.8. The reason for attaining the highest accuracy with MSFTA-G features can be explained as follows: MSFTA gave more texture information compared to original SFTA due to the computation of extra texture feature i.e, standard deviation. Moreover, MSFTA extracts texture information from both interior and exterior regions of mass and geometric features contribute mass shape information which added more discriminative power to our feature set. Thus, the combined feature set (MSFTA-G) attained the highest accuracy, sensitivity and specificity than other feature sets. We obtained the highest average accuracy of 88.67%, sensitivity of 93.30% and specificity of 85.51% from ten-fold classification as shown in Table 6.9. Figure 6.9 shows that AUC value for benign-malignant is 0.8551.

Classification scheme for non-mass-malignant and non-mass-benign: Two class systems were developed to classify 352 segmented structures into non-mass-malignant and non-mass-benign. The feature set used for this classification is shown in Table. 6.7. We have plotted ROC curves in Figures 6.10 and 6.11. From these figures, we can observe that the AUC values are

0.9921 and 0.9979 for non-mass-malignant and non-mass-benign classification respectively. To further validate our method, ten-fold cross validation is applied using SVM classifier. As shown in Tables 6.10 and 6.11, it is evident that our method has achieved higher average values for all parameters. We obtained accuracy of 93.00%, sensitivity of 91.38%, specificity of 100% for non-mass-malignant and accuracy of 94.62%, sensitivity of 93.82%, specificity of 100% for benign-non-mass using SVM (linear kernel).

Three class study (non-mass-malignant-benign): Three class system was also developed to classify 352 segmented structures into non-mass-benign-malignant. We have used the same feature set as shown in Table 6.7 for three class classification. Ten-fold cross validation is applied on five classifiers for average classification accuracy and it is shown in Table 6.12. We have attained accuracy of 88.16%. by considering SVM classifier (RBF: radial basis function kernel and $\sigma=4.5$).

Comparison of feature extraction methods for benign-malignant: In mammograms, malignant masses have homogeneous texture and benign masses have heterogeneous structure. So, gray level characteristics of mass region plays significant role for analyzing masses. In addition to this, masses also present shape characteristics. Thus, mass shape and gray level characteristics in the features help to analyze masses accurately. For comparison of benign-malignant classification, Zernike moments [96] and LQP [28] were tested with our 88 segmented mass regions. For, Zernike moments method, lower order moments were extracted from segmented mass regions. This method considers shape and margin features only. In, LQP method, different variants were tested to obtain texture features from segmented mass regions. Tables 6.13 and 6.14 reveal that our method achieved good performance in terms of accuracy, sensitivity, specificity and AUC. Table 6.15 shows that our proposed method takes less

computational time for extracting features from single image than LPQ method. The number of features and time taken to calculate Zernike moments is less compared to our proposed method. However, the difference in time is very less and is acceptable. The reason for high classification accuracy of our proposed feature set is that, it has both gray level and shape characteristics of mass. The additional advantage of MSFTA feature set is, it extracts gray level characteristics from interior and exterior regions of mass with multithreshold segmentation.

Table 6.7: Features for benign-malignant

Features	Parameters
MSFTA	$n_t=3$ and 8
Geometric	-

Table 6.8: Average classification accuracy with different features using tenfold cross validation with SVM for benign-malignant (%)

Classifier Fold	Modified SFTA($n_t=3,8$) (MSFTA)	Geo (G)	Original SFTA(OSFTA) ($n_t=3,8$)+GEO (OSFTA-G)	Modified SFTA ($n_t=3,8$)+GEO (MSFTA-G)
1.Fold	88.89	55.56	77.78	100.00
2.Fold	76.47	55.56	77.78	88.89
3.Fold	72.00	55.56	73.08	88.89
4.Fold	80.00	55.56	74.29	91.67
5.Fold	81.40	54.55	75.56	93.18
6.Fold	78.85	52.83	77.78	90.57
7.Fold	78.69	54.84	79.03	85.25
8.Fold	75.71	53.52	80.00	84.29
9.Fold	74.68	55.70	78.48	83.33
10.Fold	75.00	56.82	76.14	80.68
Average	78.17	55.05	76.99	88.67

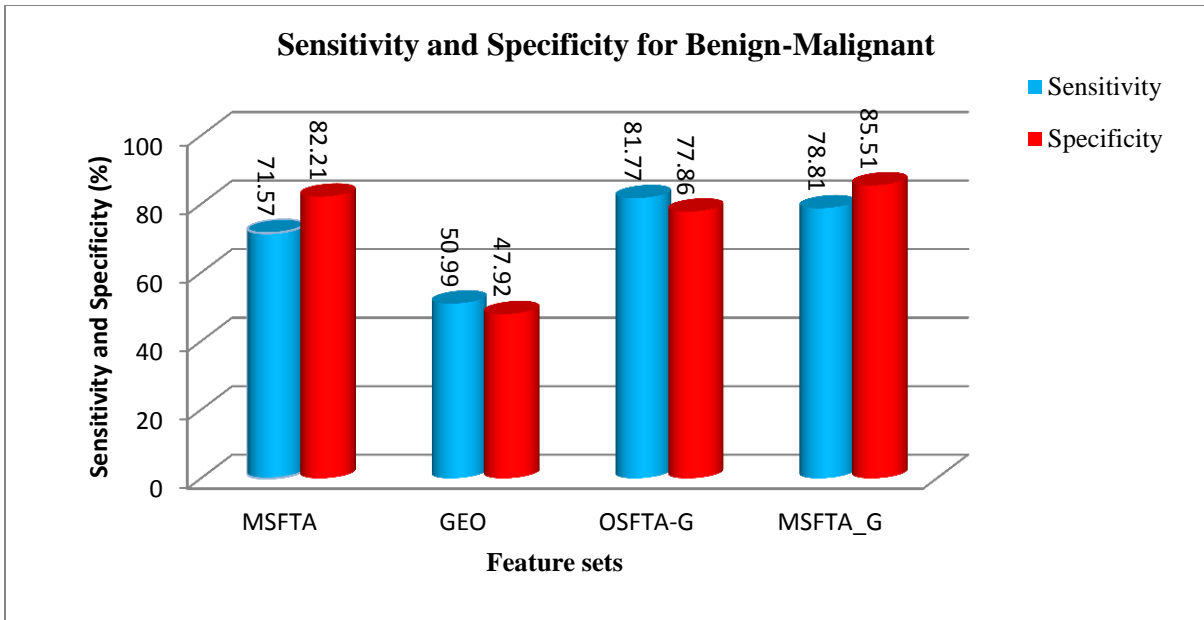


Figure 6.8: Plot of sensitivity and specificity for different feature sets

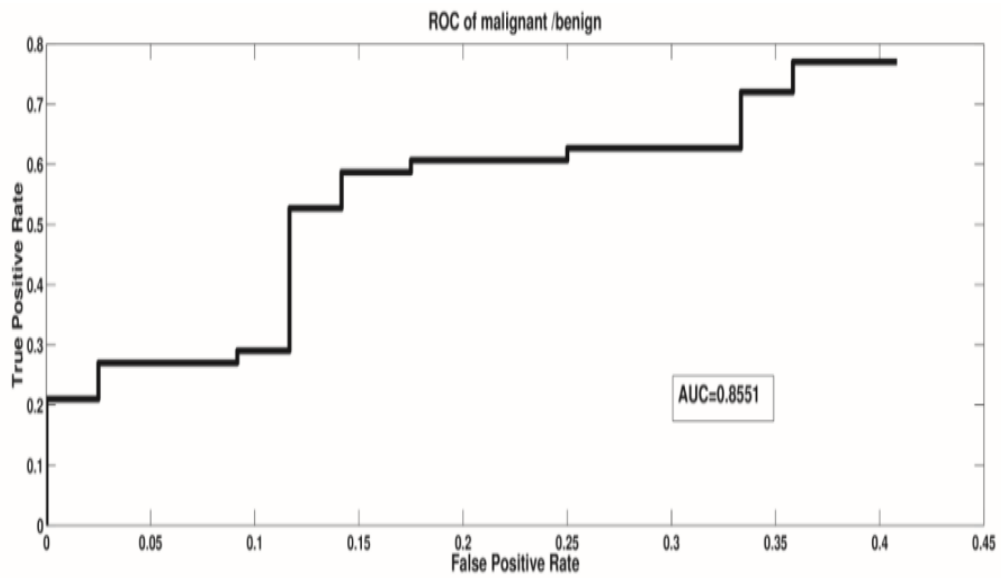


Figure 6.9: ROC analysis for benign-malignant

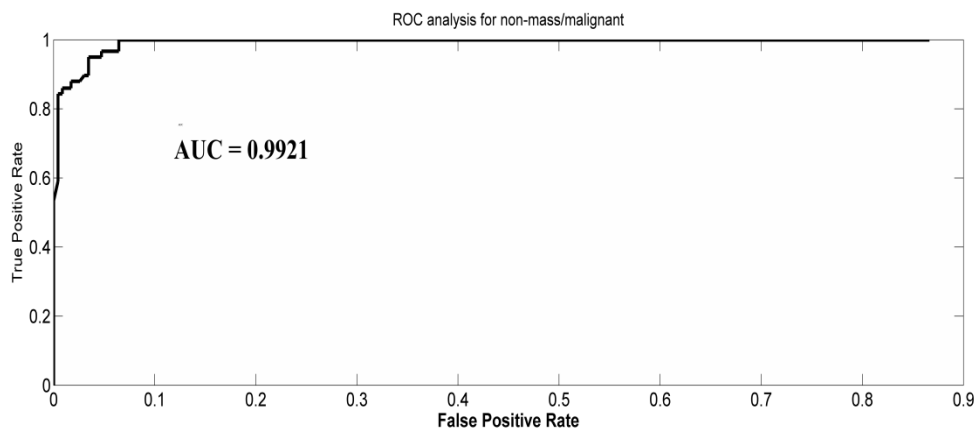


Figure 6.10: ROC analysis for non-mass-malignant classification

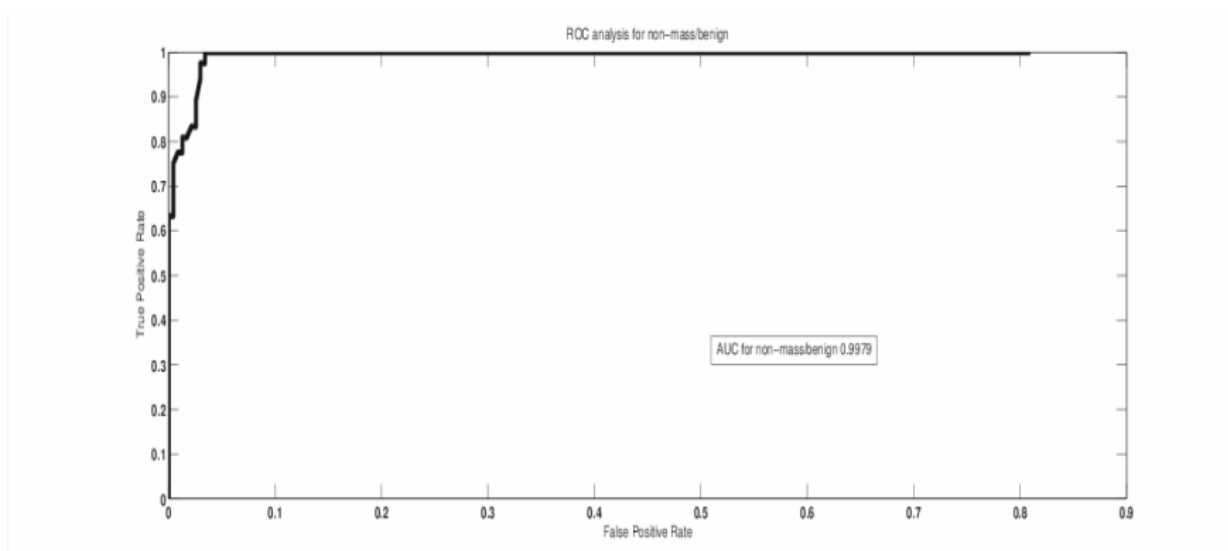


Figure 6.11: ROC analysis for non-mass-benign classification

Table 6.9: Average classification accuracy, sensitivity and specificity with MSFTA-PGF feature set (%)

Classifiers fold	Accuracy1	Sensitivity1	Specificity1
1fold	100.00	100.00	100.00
2fold	88.89	100.00	81.82
3fold	88.89	100.00	81.25
4fold	91.67	100.00	85.71
5fold	93.18	100.00	88.46
6fold	90.57	95.45	87.10
7fold	85.25	88.00	83.33
8fold	84.29	85.71	83.33
9fold	83.33	83.87	82.98
10fold	80.68	80.00	81.13
Average	88.67	93.30	85.51

Table 6.10: Mean evaluation parameters for non-mass-malignant

Classifier folds	Acc(%)	Sens(%)	Spec(%)
1fold	100.00	100.00	100.00
2fold	96.49	95.65	100.00
3fold	94.12	92.75	100.00
4fold	92.98	91.30	100.00
5fold	90.85	88.70	100.00
6fold	90.64	88.49	100.00
7fold	91.5	89.57	100.00
8fold	90.79	88.71	100.00
9fold	91.02	89.00	100.00
10fold	91.58	89.66	100.00
Average	93.00	91.38	100.00

Table 6.11: Mean evaluation parameters for non-mass-benign

Classifier folds	Acc(%)	Sens(%)	Spec(%)
1fold	96.15	95.65	100.00
2fold	96.23	95.65	100.00
3fold	96.25	95.71	100.00
4fold	95.33	94.62	100.00
5fold	94.07	93.16	100.00
6fold	93.83	92.86	100.00
7fold	94.15	93.25	100.00
8fold	93.93	93.01	100.00
9fold	94.17	93.3	100.00
10fold	92.13	90.95	100.00
Average	94.62	93.82	100.00

Table 6.12 : Average accuracies for three class study using ten fold cross validation using different classifier in (%)

Classifiers fold	SVM(RBF)	SVM(Linear)	KNN	LDA	Bayes
1fold	93.75	93.75	60.61	75.76	75.00
2fold	84.85	81.25	71.88	81.25	81.25
3fold	93.55	77.42	69.7	75.00	61.29
4fold	87.1	80.65	62.5	68.75	63.64
5fold	78.79	87.5	68.75	78.13	42.42
6fold	87.1	87.5	68.75	81.25	54.84
7fold	96.88	82.35	61.29	74.19	64.52
8fold	84.38	75.76	74.19	81.25	63.64
9fold	84.38	80.65	71.88	90.63	63.64
10fold	90.91	87.5	65.63	90.63	70.97
Average	88.16	83.43	67.52	79.69	64.11

Table 6.13 : Comparison for benign-malignant classification in (%) with different methods

Methods	Accuracy	Sensitivity	Specificity
Zernike Moments [35]	61.68	81.49	48.63
LPQ[11]	65.56	51.06	74.96
Proposed method	88.67	93.30	85.51

Table 6.14 : Comparison of AUC with different methods (benign-malignant)

benign-malignant	AUC
Zernike	0.5672
LPQ	0.6025
Proposed	0.8551

Table. 6.15: Computational time Evaluation (benign-malignant)

	Per one image(s)	No of features
Zernike	0.239413	32
LPQ	1.380255	26880
Proposed	0.45464	92

6.5.4 Discussion

This section gives the summary of our developed computer aided detection and diagnosis system of mammograms. In this study, we developed a heuristic method that detects suspicious areas in mammogram using DAK algorithm without any user intervention. It also analyses the detected area as mass-non-mass and benign-malignant. Our work also focuses on the method that generates less number of features, which does not require feature selection process. In addition to that, an automatic three class system is developed to classify segmented structures into non-mass -malignant-benign. The developed method achieved an accuracy of 97.93% for mass-non-mass and 88.67% accuracy for benign-malignant classification. The accuracy obtained for three class system is 88.16%.

To further evaluate our quality of work, the comparison is done in two ways: one way is, testing with well-known selected methods using our dataset (352 segmented structures). The selected methods are: LBP and Haralick methods described in [27], [45] for mas-non-mass classification. Zernike moments method (Ref [96]) and LQP method (In Ref. [28]) for benign-malignant classification. For comparison, we have determined accuracy, sensitivity, specificity, number of features, AUC and computation time. The results in Tables 6.4, 6.5, 6.6, 6.13, 6.14 and 6.15

show that our proposed method performed well. Other way is, we compared our presented results with other state of the art methods. However, there exists difficulty in direct comparison of our proposed method with existing methods because they use different database, number of the images, preprocessing steps and classification methods. Tables 6.16 and 6.17 summarize the performance of our method along with existing methods. We can see Table 6.17 that our method's accuracy is less than some methods in case of benign-malignant. But, these methods investigated on manually cropped regions without any previous segmentation (a more challenging task).

The main contribution of the present investigation is as follows

- Adaption of DAK algorithm and MSFTA for detection and analysis of suspicious areas without any user intervention.
- MSFTA features extract information from the interior and exterior regions of mass with Otsu's multithresholding technique.
- This approach generates less number of features compared to many multiresolution methods like wavelets, curvelets, spherical wavelets etc. which are developed in recent works of breast cancer[97,41,98,92,66]
- Many works in literature have implemented two class systems. Only few research papers worked on the three-class system for classification of normal-benign-malignant [99, 65]. In this work, we have developed three class system for non-mass-benign-malignant.

Table 6.16: Comparison of feature extraction methods for the classification of mass-non-mass

Method	Database	Detection	Manual intervention	Samples	Acc (%)	Sens (%)	Spec (%)	AUC
Local seed region growing spherical wavelet transform shape, local seed region growing, spherical wavelet transform, SVM [41] 2013	MIAS	Yes	yes	60	96	-	-	-
Contourlet-based mammography mass classification using the SVM family [98] 2010	DDSM	Yes	No(algorithm applicable for bigger size masses)	90	96.6	-	-	-
LBP+SVM [27] 2009	DDSM	No	Yes	1792		-	-	0.94
GLCM texture features, Neural network architecture[100] 2011	DDSM	No	Yes		96	-	-	-
K-means, GLCM and SVM [45] 2009	DDSM	Yes	Yes	1177	92.63	86	94.61	-
Shape features [88] 2010	DDSM			300	95	-	-	-
Quality thresholding, correlogram and shape features [25] 2014	DDSM	Yes	Yes	2033	83.53	92.3	82.2	-
Proposed method	DDSM	Yes	No	352	97.93	98	97.75	0.99

Table.6.17: Comparison of benign-malignant feature extraction techniques

Feature extraction method	Database	Samples	Acc	Sens	Spec	AUC
GaborPCA [92]	MIAS	114	80	-	-	-
Degree of spiculation of a mass and relative gradient orientation of pixels that correspond to possible spicules[93]	DDSM and MIAS	319	81	-	-	-
Fractional concavity and spiculation index [94]	MIAS	111	82			0.79
Fractal dimension using ruler method and fractional concavity [40]	MIAS	111	-	-	-	0.82
Geometric Features and ensemble classifier [95]	DDSM	-	72	-	-	-
LTP [28]	DDSM	-	-	-	-	0.97
Zernike moments[96]	DDSM	-	96.4	-	-	0.97
Proposed method	DDSM		88.67	93.3	85.51	0.8551

6.6 Conclusion

In this chapter, we have developed heuristic method consisting of preprocessing operations, DAK and MSFTA methods for detection and analysis of suspicious areas in mammograms without user intervention. We also presented three class system for the classification of mammograms into non-mass-benign-malignant. DAK algorithm generates number of clusters for k-means and selects dense cluster with mass automatically without any user intervention. The segmented structures of dense cluster are further analysed to reduce false positives. MSFTA feature extraction method is applied to the segmented structures for further classification.

Chapter 7

Conclusions and Future Work

7.1 Introduction

In this section, the work accomplished during the research is summarized. The main objective of the research is to develop a simple software algorithm that detects abnormalities in mammogram and classify them into cancer (malignant)-non-cancer (benign) without any manual intervention. The research has been started with the intention of assisting radiologists to decrease inter-observer and intra-observer variability. The other advantage of the software algorithm developed is that it can also help the people in rural area, where there is inadequate medical facility and the accessibility of expert radiologists is very less compared to urban areas. In this thesis, new algorithms have been developed and validated on image enhancement, segmentation, feature extraction and classification techniques.

7.2 Implementations of the developed methods and future work

The developed methods contributions have been distributed in six chapters. In chapter 2, a mammogram enhancement and segmentation for abnormality detection schemes have been developed to assist radiologists. The image enhancement scheme employs Gaussian pyramid and wavelet fusion methods. In chapter 3, feature extraction techniques such as LBP, SFTA and RILF techniques have been studied. These features are further submitted to SVM classifier to delineate mammogram ROIs into normal-abnormal. Chapter 4 deals with classification of mammogram into benign-malignant using mass contours. Different feature extraction methods based on geometric, poly regression, BAS and TAR techniques have been implemented. The

features extracted from the above methods are given to different classifiers such as SVM, KNN, adaboost and ANN to classify mammograms into benign-malignant. All the above chapters utilize ROIs of mammograms as a dataset for validation. Chapter 5 deals with the simulation results of image enhancement, segmentation, feature extraction methods for mammogram classification into normal-abnormal and benign-malignant. To validate image enhancement method, quality measures such as EBCM and entropy are calculated for original and enhanced mammograms. Along with the image enhancement method, we also developed image segmentation for the detection of mass using thresholding and morphological techniques. Dice coefficient has been calculated to compare the segmented mass with the ground truth information of mass. The results obtained by the proposed feature extraction techniques in chapter 3 and chapter 4 are good. Among them, RILF performed well compared to LBP and SFTA techniques and TAR achieved best result compared to geometrical, polynomial regression and BAS techniques for the classification of mammograms into normal-abnormal and benign-malignant respectively. In the end of this chapter, we implemented the proposed techniques in chapter 2, chapter 3 and chapter 4 on complete mammogram and the results were presented. In chapter 6, a complete mammogram is submitted to software algorithm for automatic detection and classification of masses. It helps the radiologists in detecting the masses accurately in less time. This chapter presents heuristic method for segmentation, feature extraction and classification of masses in digital mammograms for computer aided diagnosis (CAD) system. The primary advantage of this method is to automatically detect and analyze the abnormalities. This algorithm helps the radiologists when confronting with more number of mammograms. It also helps to make accurate evaluation without any manual intervention. The proposed algorithm is a three step method. First step is to detect abnormality in mammogram using Dunn index based

automated k-means algorithm(DAK), Second step incorporates the novel modified-segmented fractal texture analysis(MSFTA) algorithm which we developed for feature extraction and third step involves classification of abnormalities as mass-non-mass and benign-malignant using support vector machine (SVM). The training and testing samples taken to validate are not the same for all proposed methods. However, comparison between our methods and also with the existing methods is performed on the same dataset

The objective of each scheme suggested in this thesis is to classify only one type of abnormality. It does not consider the other type of abnormalities such as microcalcification, architectural distortion and bilateral asymmetry. The abnormalities in mammograms such as mass and microcalcification have different morphology. The masses have different shapes like lobular, circumscribed speculated, ill-defined, lobulated etc. The microcalcifications are ring shaped, punctuate, linear, needle shaped etc. The detection of morphology of mass or microcalcification assists the radiologists in detecting the grade of the tumor in CAD system which is one of the limitations of this thesis. In future, we would like to work on multi class classification of masses and microcalcification.

Bibliography

- [1]. Deserno TM. “Fundamentals of biomedical image processing”. In Biomedical Image Processing 2010 (pp. 1-51). Springer, Berlin, Heidelberg.
- [2]. H. Ostensen, Diagnostic imaging: What is it? When and how to use it where resources are limited?" World Health Organization (WHO), 2001.
- [3]. Chan V, Perlas A. “Basics of ultrasound imaging. In Atlas of ultrasound-guided procedures in interventional pain management”, Springer, New York, pp. 13-19, 2011.
- [4]. Sree SV, Ng EY, Acharya RU, Faust O. “Breast imaging: A survey. World journal of clinical oncology”, vol.2 no.4 :171 Apr 10; 2011.
- [5]. Zhu H. Medical image processing overview. University of Calgary, Summer School Program-Introduction to Mathematical Medicine, held at the University of Waterloo. 2003.
- [6]. <http://www.who.int/cancer/detection/breastcancer/en/index1.html>
- [7]. http://www.breastcancerindia.net/statistics/stat_global.html
- [8]. Alghaib HA, Scott M, Adhami RR. “An overview of mammogram analysis. IEEE Potentials”. Vol 35, no.6, pp21-8, Nov 2016.
- [9]. Hela, B., Hela, M., Kamel, H., Sana, B. and Najla, M. “Breast cancer detection: A review on mammograms analysis techniques”. in *Systems, Signals & Devices (SSD), 2013 10th International Multi-Conference on*, pp. 1-6. IEEE, March 2013.
- [10]. Jalalian, A., Mashohor, S.B., Mahmud, H.R., Saripan, M.I.B., Ramli, A.R.B. and Karasfi, B. “Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review” *Clinical imaging*, vol 37, no.3, pp.420-426, 2013.

- [11]. <http://marathon.csee.usf.edu/Mammography/Database.html>
- [12]. <http://irma-project.org/>
- [13]. Hadjiiski, Lubomir, Berkman Sahiner, and Heang-Ping Chan. "Advances in CAD for diagnosis of breast cancer." *Current opinion in obstetrics & gynecology*, vol 18, no.1,64 2006.
- [14]. Eltoukhy, M. M., Faye, I., & Samir, B. B. "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation." *Computers in biology and medicine*, vol.42, no.1, pp.123-128, 2012.
- [15]. Abdellatif, H.; Taha, T.E.; Zahran, O.F.; Al-Nauimy, W.; El-Samie, F.E.A.: K2. "Automatic pectoral muscle boundary detection in mammograms using eigenvectors segmentation," in *Radio Science Conference (NRSC), 2012 29th National* , pp.633-640, April, 2012.
- [16]. Li Liu, Qian Liu, Wei Lu "Pectoral Muscle Detection in Mammograms Using Local Statistical Features" *Journal of Digital Imaging* , Vol.27, no. 5, pp.633-641, 2014.
- [17]. Mustra, M.; Bozek, J.; Grgic, M. "Nipple detection in craniocaudal digital mammograms" in *ELMAR '09. International Symposium*, pp.15-18, 28-30 Sept. 2009.
- [18]. Zhe Wu; Julong Yuan; Binghai Lv; Xiaofeng Zheng, "Digital mammography image enhancement using improved unsharp masking approach," in *Image and Signal Processing (CISP), 2010 3rd International Congress on* , vol.2, no., pp.668-672, 16-18 Oct. 2010
- [19]. Sundaram, M., K. Ramar, N. Arumugam, and G. Prabin. "Histogram modified local contrast enhancement for mammogram images." *Applied Soft Computing*, vol.11, no.8 pp:5809-5816, 2011.

- [20]. Rahmati, Peyman, Andy Adler, and Ghassan Hamarneh. "Mammography segmentation with maximum likelihood active contours." *Medical image analysis*, vol.16, no.6,pp: 1167-1186, 2012.
- [21]. Liu, Jun, et al. "Mass segmentation using a combined method for cancer detection." *BMC systems biology* vol.5, no.3.Suppl 3 2011.
- [22]. Hong, Byung-Woo, and Bong-Soo Sohn. "Segmentation of regions of interest in mammograms in a topographic approach." *Information Technology in Biomedicine, IEEE Transactions on* vol.14, no.1, pp:129-139, 2010.
- [23]. Gedik, Nebi, and Ayten Atasoy. "A computer-aided diagnosis system for breast cancer detection by using a curvelet transform." *Turkish Journal of Electrical Engineering & Computer Sciences* vol.21, no.4, pp: 1002-14, 2013.
- [24]. Liu, Xiaoming, and Zhigang Zeng. "A new automatic mass detection method for breast cancer with false positive reduction." *Neurocomputing*, Vol. 152, pp: 388-402, 2015.
- [25]. de Nazaré Silva, J., de Carvalho Filho, A.O., Silva, A.C., De Paiva, A.C. and Gattass, M., "Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and SVM." *Journal of digital imaging*, vol.28, no.3, pp.323-337, 2015.
- [26]. Sharma, Shubhi, and Pritee Khanna. "Computer-Aided Diagnosis of Malignant Mammograms using Zernike Moments and SVM." *Journal of digital imaging* vol.28, no.1,pp:77-90, 2015.
- [27]. Lladó, Xavier, Arnau Oliver, Jordi Freixenet, Robert Martí, and Joan Martí "A textural approach for mass false positive reduction in mammography." *Computerized Medical Imaging and Graphics*, vol. 33, no.6, pp: 415-422, 2009.

- [28]. Nanni, Loris, Sheryl Brahnham, and Alessandra Lumini. "A very high performing system to discriminate tissues in mammograms as benign and malignant." *Expert Systems with Applications* vol. 39, no.2, pp: 1968-1971, 2012.
- [29]. Tai, Shen-Chuan, Zih-Siou Chen, and Wei-Ting Tsai. "An automatic mass detection system in mammograms based on complex texture features" *Biomedical and Health Informatics, IEEE Journal of* vol.18, no.2, pp: 618-627, 2014.
- [30]. Cao, Ying, et al. "An adaptive region growing algorithm for breast masses in mammograms." *Frontiers of Electrical and Electronic Engineering in China* vol.5, no.2 pp: 128-136, 2010.
- [31]. Dong, Min, Xiangyu Lu, Yide Ma, Yanan Guo, Yurun Ma, and Keju Wang. "An efficient approach for automated mass segmentation and classification in mammograms." *Journal of digital imaging*, vol.28, no. 5, pp: 613-625, 2015.
- [32]. Jen, Chun-Chu, and Shyr-Shen Yu. "Automatic detection of abnormal mammograms in mammographic images." *Expert Systems with Applications*, vol.42, no.6 , pp:3048-3055, 2015.
- [33]. Beura, Shradhananda, Banshidhar Majhi, and Ratnakar Dash. "Automatic Characterization of Mammograms using Fractal Texture Analysis and Fast Correlation Based Filter Method." *Proceedings of the 2nd International Conference on Perception and Machine Intelligence*. ACM, 2015.
- [34]. Costa, Alceu Ferraz, Gabriel Humpire-Mamani, and Agma Juci Machado Traina. "An efficient algorithm for fractal analysis of textures." *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*. IEEE, 2012.

- [35]. Kendall, Edward J., Michael G. Barnett, and Krista Chytk-Praznik. "Automatic detection of anomalies in screening mammograms." *BMC medical imaging* 13.1 (2013): 43.
- [36]. Mencattini, Arianna, Marcello Salmeri, Paola Casti, Grazia Raguso, Samuela L'Abbate, Loredana Chieppa, Antonietta Ancona, Fabio Mangieri, and Maria Luisa Pepe. "Automatic breast masses boundary extraction in digital mammography using spatial fuzzy c-means clustering and active contour models." In *Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on*, pp. 632-637. IEEE, 2011
- [37]. Do Nascimento, Marcelo Zanchetta, Alessandro Santana Martins, Leandro Alves Neves, Rodrigo Pereira Ramos, Edna LúCia Flores, and Gilberto Arantes Carrijo. "Classification of masses in mammographic image using wavelet domain features and polynomial classifier." *Expert Systems with Applications* vol 40, no.15 pp: 6213-6221, 2013.
- [38]. Bator, Marcin, and Mariusz Nieniewski. "Detection of cancerous masses in mammograms by template matching: Optimization of template brightness distribution by means of evolutionary algorithm." *Journal of digital imaging*, vol .25, no.1, pp: 162-172,2012.
- [39]. Ericeira, Daniel Rodrigues, AristóFanes CorrêA Silva, Anselmo Cardoso De Paiva, and Marcelo Gattass "Detection of masses based on asymmetric regions of digital bilateral mammograms using spatial description with variogram and cross-variogram functions." *Computers in biology and medicine*, vol. 43, no.8, pp: 987-999, 2013.
- [40]. Rangayyan, Rangaraj M., and Thanh M. Nguyen. "Fractal analysis of contours of breast masses in mammograms." *Journal of Digital Imaging*, Vol 20, no.3, pp: 223-237, 2007.

- [41]. Görgel, Pelin, Ahmet Sertbas, and Osman N. Ucan. "Mammographical mass detection and classification using local seed region growing–spherical wavelet transform (lsrg–swt) hybrid scheme." *Computers in biology and medicine* , vol.43, no.6, 765-774, 2013.
- [42]. Pereira, Danilo Cesar, Rodrigo Pereira Ramos, and Marcelo Zanchetta do Nascimento. "Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm." *Computer methods and programs in biomedicine*, vol.114, no.1, 88-101, 2014.
- [43]. Mohanty, A. K., Senapati, M. R., Beberta, S., & Lenka, S. K. "Texture-based features for classification of mammograms using decision tree." *Neural Computing and Applications*, vol. 23, no. 3-4, pp:1011-1017, 2013.
- [44]. Dubey, R. B., Madasu Hanmandlu, and S. K. Gupta. "A comparison of two methods for the segmentation of masses in the digital mammograms." *Computerized Medical Imaging and Graphics vol 34. no.3*, pp: 185-191,2010
- [45]. Oliveira Martins, Leonardo, et al. "Detection of masses in digital mammograms using K-means and support vector machine." *ELCVIA: electronic letters on computer vision and image analysis, vol.8, no.2*, pp: 39-50,2009.
- [46]. Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks. Machine learning." Vol.20, no.3 pp: 273-297,1995.
- [47]. <http://www.macmillan.org.uk/Cancerinformation/Causesriskfactors/Preancerous/Breastcalfications.aspx>.
- [48]. Jose Salvado, Bruno Roque "Detection of Calcifications in Digital Mammograms using Wavelet analysis and Contrast Enhancement", *IEEE Int. Work shop in signal processing*,pp.200-205,2005,Portuga.l

- [49]. Sheila Timp, Celia Varela, and Nico Karssemeijer “Temporal Change Analysis for Characterization of Mass Lesions in Mammography” *IEEE Int..Trans. in Medical imaging*, Vol 26, no.7, pp.945-953, 2007.
- [50]. <http://www.breastCancer.ca/screening/mammogramabnormalities>
- [51]. Wang, W. and Chang, F “A Multi-focus Image Fusion Method Based on Laplacian Pyramid.” *JCP*, vol.6 ,no.12, pp.2559-2566, 2011.
- [52]. C. Gonzalez and E. Woods, *Digital Image Processing*.
- [53]. Airouche, M., Bentabet, L. and Zelmat, M., “Image segmentation using active contour model and level set method applied to detect oil spills.” In *Proceedings of the World Congress on Engineering Lecture Notes in Engineering and Computer Science*.Vol. 1, No. 1, pp. 1-3, 2009, July.
- [54]. Sampaio, W.B., Diniz, E.M., Silva, A.C., De Paiva, A.C. and Gattass, M. “ Detection of masses in mammogram images using CNN, geostatistic functions and SVM”. *Computers in biology and medicine*, vol.41, no.8, pp.653-664,2011.
- [55]. Rodrigues R, Braz R, Pereira M, Moutinho J, Pinheiro AM. A two-step segmentation method for breast ultrasound masses based on multi-resolution analysis. *Ultrasound in Medicine and Biology*. Vol.41, no.6, pp:1737-48 2015
- [56]. Huang, Qing-Hua, Su-Ying Lee, Long-Zhong Liu, Min-Hua Lu, Lian-Wen Jin, and An-Hua Li. "A robust graph-based segmentation method for breast tumors in ultrasound images." *Ultrasonics* , vol.52, no. 2, pp: 266-275,2015.
- [57]. Wang L, Pan C. Robust level set image segmentation via a local correntropy-based K-means clustering. *Pattern Recognition*. Vol.47, no.5, pp:1917-25, 2014.

- [58]. Kumar HS, Raja KB, Venugopal KR, Patnaik LM. “Automatic image segmentation using wavelets. *International Journal of Computer Science and Network Security.*” Vol.9, no.2, pp:305-13, 2009.
- [59]. Ojala, Timo, T., Valkealahti, K., Oja, E., & Pietikäinen, M ‘Texture discrimination with multidimensional distributions of signed gray-level differences(2001)’ *Pattern Recognition*, Vol. 34, No.3, pp.727-739.
- [60]. Paci, M., Nanni, L., Lahti, A., Aalto-Setälä, K., Hyttinen, J., & Severi, S. “Non-binary coding for texture descriptors in sub-cellular and stem cell image classification” *Current Bioinformatics*, vol.8 , no.2, pp:208-219, 2013.
- [61]. Rouzbeh Maani, Sanjay Kalra, and Yee-Hong Yang “Rotation invariant local frequency descriptors for texture classification” *IEEE Transactions on Image Processing*, Vol.22, No.6, pp.2409-2419, 2013.
- [62]. Wang, Wei, Weimin Chen, and Dongxia Xu ‘Pyramid-based multi-scale lbp features for face recognition’, In *Multimedia and Signal Processing (CMSP) International Conference on*, vol. 1, pp. 151-155,2011.
- [63]. Wajid, Summrina Kanwal, and Amir Hussain “Local energy-based shape histogram feature extraction technique for breast cancer diagnosis” *Expert Systems with Applications* , Vol.42, No. 20, pp.6990-6999,2015.
- [64]. Sampat MP, Bovik AC, Markey MK. Classification of mammographic lesions into BI-RADS shape categories using the beamlet transform. In *Medical Imaging 2005: Image Processing 2005 Apr 29* (Vol. 5747, pp. 16-26). International Society for Optics and Photonics.

- [65]. Narváez F, Díaz G, Poveda C, Romero E. “An automatic BI-RADS description of mammographic masses by fusing multiresolution features.” *Expert Systems with Applications*, vol.74, pp:82-95, 2017.
- [66]. Bruno DO, do Nascimento MZ, Ramos RP, Batista VR, Neves LA, Martins AS. “LBP operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues. *Expert Systems with Applications*”. Vol.55, pp:329-40, 2016.
- [67]. Verma B, McLeod P, Klevansky A. “Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer”. *Expert systems with applications*. vol.37, no.4, pp: 3344-51, 2010.
- [68]. Gedik N. A new feature extraction method based on multi-resolution representations of mammograms. *Applied Soft Computing*, vol.44, pp: 128-33,2016.
- [69]. Rabidas R, Chakraborty J, Midya A. “Analysis of 2D singularities for mammographic mass classification.”, *IET Computer Vision*, vol.11,no.1, pp:22-32, 2016.
- [70]. Li Y, Chen H, Rohde GK, Yao C, Cheng L. “Texton analysis for mass classification in mammograms. *Pattern Recognition Letters*; vol.52, pp:87-93, 2015.
- [71]. Chen, C.M., Chou, Y.H., Han, K.C., Hung, G.S., Tiu, C.M., Chiou, H.J. and Chiou, S.Y., “Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks.”, *Radiology*, vol.22, no.2, pp.504-514, 2003.
- [72]. Rahil Garnavi, Mohammad Aldeen, James Bailey. “Computer-Aided Diagnosis of Melanoma Using Border- and Wavelet-Based Texture Analysis” *IEEE Transactions on Information Technology in BioMedicine*, vol.16, no..6, pp: 1239-125,2012
- [73]. <http://zone.ni.com/devzone/cda/epd/p/id/1517>.

- [74]. Rangarai.M, Rangayam, Faraz Oloumi, Thanh M. Nguyen, “Fractal Analysis of Contours of Breast Masses in Mammograms via the Power Spectra of the Signatures”, *IEEE Annual International On Engineering in Medicine and Biology, EMBS*, pp:6737-6740,2010.
- [75]. Arica N, Vural FT. BAS: a perceptual shape descriptor based on the beam angle statistics. *Pattern Recognition Letters*. Vol. (9-10):1627-39, 2003.
- [76]. Rabidas R, Chakraborty J, Midya A. Analysis of 2D singularities for mammographic mass classification. *IET Computer Vision*, vol.11, no.1,pp:22-32, 2016 Jul 27.
- [77]. Li H, Meng X, Wang T, Tang Y, Yin Y. “Breast masses in mammography classification with local contour features.”, *Biomedical engineering online*. vol.16, no.1, pp:44, 2017.
- [78]. Alajlan N, El Rube I, Kamel MS, Freeman G. “Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recognition*.” vol.40, no.7, pp:1911-20, 2007
- [79]. Pawar, M.M. and Talbar, S.N.“ Local entropy maximization based image fusion for contrast enhancement of mammogram.” *Journal of King Saud University-Computer and Information Sciences*, 2018.
- [80]. Sundaram, M., Ramar, K., Arumugam, N. and Prabin, G., 2011. “Histogram modified local contrast enhancement for mammogram images.” *Applied soft computing*, vol.11, no.8, pp:5809-5816, 2011.
- [81]. Rangarai.M, Rangayam, Faraz Oloumi, Thanh M. Nguyen, “Fractal Analysis of Contours of Breast Masses in Mammograms via the Power Spectra of the Signatures”, *IEEE Annual International On Engineering in Medicine and Biology, EMBS*, 2010, pp:6737-6740.

- [82]. Boss, R., K. Thangavel, and D. Daniel. "Automatic mammogram image breast region extraction and removal of pectoral muscle." arXiv preprint arXiv: 1307.7474 (2013).
- [83]. <http://www.mathworks.com/matlabcentral/fileexchange/50281-mammogram-image-enhancement>
- [84]. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R. and Wu, A.Y., "An efficient k-means clustering algorithm: Analysis and implementation." *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.7, pp.881-892, 2002.
- [85]. Arakeri, Megha P., and G. Ram Mohana Reddy. "An intelligent content-based image retrieval system for clinical decision support in brain tumor diagnosis." *International Journal of Multimedia Information Retrieval* vol.2, no. 3 , pp:175-188, 2013.
- [86]. Zhao, Qian, Chang-Zheng Shi, and Liang-Ping Luo. "Role of the texture features of images in the diagnosis of solitary pulmonary nodules in different sizes." *Chinese Journal of Cancer Research* vol.26, no. 4, pp:451, 2014.
- [87]. Özekes, Serhat, Onur Osman, and A. Yilmaz Camurcu. "Mammographic mass detection using a mass template." *Korean journal of radiology* , vol.6, no. 4, pp: 221-228, 2005.
- [88]. Surendiran, B., and A. Vadivel. "Mammogram mass classification using various geometric shape and margin features for early detection of breast cancer." *International Journal of Medical Engineering and Informatics* , vol.4, no.1, pp: 36-54,2012
- [89]. Raguso, Grazia, Antonietta Ancona, Loredana Chieppa, Samuela L. Abbate, Maria Luisa Pepe, Fabio Mangieri, Miriam De Palo, and Rangaraj M. Rangayyan. "Application of fractal analysis to mammography." In *Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE*, pp. 3182-3185, 2010.

- [90]. Li, Y., Chen, H., Wei, X., Peng, Y., & Cheng, L. "Mass classification in mammograms based on two-concentric masks and discriminating texton. Pattern Recognition", vol.60, 648-656,2016.
- [91]. Yang, Mingqiang, Kidiyo Kpalma, and Joseph Ronsin. "A survey of shape feature extraction techniques." Pattern recognition, pp: 43-90, 2008.
- [92]. Buciu, Ioan, and Alexandru Gacsadi. "Directional features for automatic tumor classification of mammogram images." Biomedical Signal Processing and Control, vol.6, no. 4, pp :370-378, 2011.
- [93]. Rojas-Domínguez, Alfonso, and Asoke K. Nandi. "Development of tolerant features for characterization of masses in mammograms." Computers in Biology and Medicine, vol.39, no. 8 ,pp :678-688, 2009.
- [94]. Rangayyan, Rangaraj M., Naga R. Mudigonda, and J. EL Desautels. "Boundary modelling and shape analysis methods for classification of mammographic masses." Medical and Biological Engineering and Computing, vol.38, no. 5, pp: 487-496, 2000.
- [95]. Zhang, Yu, Noriko Tomuro, Jacob Furst, and Daniela Stan Raicu. "Building an ensemble system for diagnosing masses in mammograms." International journal of computer assisted radiology and surgery, vol.7, no.2, pp: 323-329, 2012.
- [96]. Tahmasbi, Amir, Fatemeh Saki, and Shahriar B. Shokouhi. "Classification of benign and malignant masses based on Zernike moments." Computers in biology and medicine, vol.41, no.8, pp:726-735,2011.

- [97]. Eltoukhy, M. M., Faye, I., & Samir, B. B. "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation." *Computers in biology and medicine*, vol.42, no.1, pp:123-128, 2012.
- [98]. CLMoayedi, Fatemeh, Zohreh Azimifar, Reza Boostani, and Serajodin Katebi. "Contourlet-based mammography mass classification using the SVM family." *Computers in biology and medicine*, vol.40, no. 4 , pp: 373-383,2010.
- [99]. Ergin, Semih, and Onur Kilinc. "A new feature extraction framework based on wavelets for breast cancer diagnosis." *Computers in biology and medicine*, vol.51, pp: 171-182,2014
- [100]. Nithya, R., and B. Santhi. "Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer." *International Journal of Computer Applications*, vol.28, no. 6, pp: 21-25, 2011.

Publications Related to this Thesis

International Conferences

- [1]. Spandana, P., and KMM Rao. "Novel image processing techniques for early detection of breast cancer, mat lab and lab view implementation." In *Point-of-Care Healthcare Technologies (PHT), 2013 IEEE*, pp. 105-108. IEEE, 2013.(Cited by 6)
- [2]. Paramkusham, Spandana, KMM Rao, and BVVSN Prabhakar Rao. "Early stage detection of breast cancer using novel image processing techniques, Matlab and Labview implementation." In *Advanced Computing Technologies (ICACT), 2013 15th International Conference on*, pp. 1-5. IEEE, 2013.(Cited by 2)
- [3]. Paramkusham, Spandana, K. M. M. Rao, and BVVSN Prabhakar Rao. "Automatic Detection of Breast Lesion Contour and Analysis Using fractals through spectral methods." In *Advances in Computer Science, AETACS*. 2013.
- [4]. Paramkusham, Spandana, Patri Shivakshit, K. M. M. Rao, and BVVSN Prabhakar Rao. "A new features extraction method based on polynomial regression for the assessment of breast lesion Contours." In *Industrial Instrumentation and Control (ICIC), 2015 International Conference on*, pp. 579-583. IEEE, 2015.
- [5]. Paramkusham, Spandana, K. M. M. Rao, and BVVSN Prabhakar Rao. "Novel technique for the detection of abnormalities in Mammograms using texture and geometric features." In *Microwave, Optical and Communication Engineering (ICMOCE), 2015 International Conference on*, pp. 150-153. IEEE, 2015

International Journals

- [1]. Paramkusham, Spandana, KMM Rao, and BVVSN Prabhakar Rao. "Comparison of rotation invariant local frequency, LBP and SFTA methods for breast abnormality classification." International Journal of Signal and Imaging Systems Engineering, vol.11, no. 3, pp: 136-150, 2018. **(Emerging Sources Citation Index &Scopus)**
- [2]. Paramkusham, Spandana, KMM Rao, Prabhakar Rao BVVSN, and Shivam Sharma. "Application of TAR signature for breast mass analysis." Biomedical Research, vol.29, no. 10, pp: 2030-2034, 2018. **(Scopus)**
- [3]. Paramkusham, Spandana, Kunda MM Rao, Prabhakar Rao BVVSN, "Automatic detection of abnormality in mammogram using Dunn index based automated k-means and MSFTA" submitted to Sadhana, springer publishers **(under Review)**.

Biographies

Candidate Biography

Paramkusham Spandana received Mtech degree in Digital systems in Computer Electronics from Jawaharlal Nehru Technology University, Hyderabad. Her research interest includes Biomedical image processing, and artificial intelligence.

Supervisor Biography

Prof. B.V.V.S.N.Prabhakar Rao, obtained Mtech and PhD from IIT Delhi. He is currently working as an Associate Professor in EEE Department at Bits Pilani, Hyderabad Campus. His research interests include studies on Solar energy, Energy management, Biomedical signal and image processing areas in which he have more than 20 publications in national/international conferences and published few journal papers. Recently he has published an article on “Noninvasive Biomarker” along with his research team members in a book “Higher Education” by Times of India.

Co-Supervisor Biography

Dr.K.M.M.Rao received his B.E (Honors) and M.E. in Electronics from Birla Institute of Technology and Science, Pilani in 1970 & 1972, respectively and his Ph.D in Image Processing from S.K.University, Ananthapur. Dr.K.M.M.Rao joined Indian Institute of Science, Bangalore in 1973. He was responsible for the design and development of Black and White and Colour Drum Scanner Imager, first time in India, for digitizing and reproducing various types of photographic prints.. He was Deputy Director (Data Processing Area) at NRSC, till December

2006. Currently, he is Adjunct faculty in BITS-Pilani Hyderabad campus. He has designed and developed number of technologies for Image Processing and Remote Sensing applications. He obtained five patents for his designs. His research interests include Satellite Data Processing, Image Processing, Medical Imaging, Data Mining and Photo Writing. Further details are at www.drkmm.com