A HYBRID FEATURE BASED METHOD FOR EARLY DETECTION AND CLASSIFICATION OF MAMMOGRAPHIC DENSITY IN BREAST CANCER DIAGNOSIS

By
Najmus Saltanat

MASTER OF SCIENCE IN
ELECTRICAL AND ELECTRONIC ENGINEERING

**DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY**

October 2011

The thesis entitled **"A Hybrid Feature Based Method For Early Detection and Classification of Mammographic Density in Breast Cancer Diagnosis"** submitted b y N ajmus S altanat, Student No: 080406247P, S ession: A pril, 2008 ha s be en a ccepted a s s atisfactory i n pa rtial fulfillment o f th e re quirement f or th e d egree o f M ASTER O F S CIENCE IN E LECTRICAL AND ELECTRONIC ENGINEERING on October 30, 2011.

BOARD OF EXAMINERS

1._____

(Dr. Aynal Haque)   Chairman
  Professor,   (Supervisor)
  Department of Electrical and Electronic Engineering,
  Bangladesh University of Engineering and Technology
  Dhaka-1000, Bangladesh.

2._____

  (Dr. Md. Saifur Rahman)   Member
  Professor and Head   (Ex-Officio)
  Department of Electrical and Electronic Engineering,
  Bangladesh University of Engineering and Technology
  Dhaka-1000, Bangladesh.

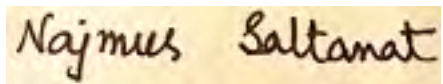3._____

  (Dr. Celia Shahnaz)   Member
  Assistant Professor,
  Department of Electrical and Electronic Engineering,
  Bangladesh University of Engineering and Technology
  Dhaka-1000, Bangladesh.

4._____

  (Dr. Khawza Iftekhar Uddin Ahmed)   Member
  Associate *Professor*,   (External)
  Department of Electrical and Electronic Engineering,
  United International University
  Dhaka-1209, Bangladesh.

CANDIDTE'S DECLARATION

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

Najmus Saltanat

_____

Najmus Saltanat

Dedication

To my parents.

Acknowledgements

I w ould l ike t o e xpress m y gr atitude t o m y supervisor D r. M d. A ynal H aque, Professor, Department o f EEE, BUET, for hi s e ndless pa tience, friendly s upervision a nd i nvaluable assistance in making a difficult task a pleasant one.

I w ould l ike t o e xpress my thanks a nd gr atitude t o D r. A lamgir H ossain, P rofessor, D ept o f Computer S cience, N orthumbria U nversity f or hi s support to s tart t his r esearch w ork unde r Erasmus Mundus E-Link Scholarship.

I would like to express gratitude to Dr. Celia Shahnaz, Assistant Professor, Department of EEE, BUET, for he r endless p atience, valuable a dvice a nd di rections t o c omplete my present w ork successfully.

I would like to convey thanks to my friend, Rasel Mahmud for his mental support and directions during my present research.

I would express my thanks also to the head of the department of EEE, BUET for his support to complete this thesis.

Sincerest thanks t o m y husband, Mynul M ohammad S hah a nd my parents f or t heir co nstant support and criticism of the thesis work.

Abstract

Breast cancer is a common disease with unknown reason. So, early detection or prediction is better way to fight against this disease. A method can be used to detect mass in mammograms and predict breast cancer in its early stage. This early detection of breast cancer will play an important role to decrease the mortality rate due to this disease. The method, in general, consists of some sections like: image pre-processing, pectoral muscle segmentation, suspected area identification, feature extraction and classification. The pectoral muscle extraction and removal from the mammogram is a prerequisite to gain more accurate mammographic density detection and classification. Here, the pectoral muscle segmentation has been performed by region grow method. Then the mammogram without pectoral muscle has been divided into sixteen-by-sixteen blocks and by taking the Gray Level Co-occurrence matrix (GLCM) of the blocks, the region of interest (ROI) containing possible mammographic density has been extracted for further analysis. Intensity-based or area-based or hybrid of both the features based extracted from the segmented region of interest (ROI) of the original mammogram have been used for classification. By using Support Vector Machine (SVM) method or two-layer Neural Network (NN) method, the ROI of the mammographic density was classified as the normal, benign or malignant. Simulations have been performed on Mammographic Image Analysis Society (MIAS) mini database and effectiveness of the method was shown by different performance. Correction rate has been found 96.77% for SVM and 95% for Feed Forward Neural Network. The proposed method of mammogram screening for early detection of breast cancer can open a new era to biomedical field and make it easier to decrease mortality rate caused by breast cancer.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

MC   Microcalcification

ROI   Region of Interest

NN   Neural Network

ROI   Region of Interest

SVM   Support Vector Machine

CAD   Computer Aided Diagnosis

GLCM   Gray-Level Co-Occurrence Matrix

CLAHE   Contrast Limited Adaptive Histogram Equalization

MIAS   Mammographic Image Analysis Society

HDM   Hausdorff Distance Measure

MLO   Mediolateral Object

MST   Minimum Spanning Trees

FP   False Positive

FN   False Negative

TP   True Positive

TN   True Negative

MAEDM   Mean of Absolute Error Distance Measure

MSD   Mean Square Deviation

IDL   Interactive Data Language

SF   Spiculation Filters

ED   Euclidean distance

SOD   Sum of Differences

GA   Genetic Algorithm

ROC   Receiver Operating Curve

BBN   Bayesian Belief Networks

ANN   Artificial Neural Network

Chapter 1
INTRODUCTION

1.1 Background

Breast cancer is the second commonest cancer among women of Bangladesh. Nulliparous women, daughters of breast cancer patients, and those who do not breast-feed their babies and indulge in high-fat diet suffer more from breast cancer. Based on 21,23 8 cancer (Male 14,222 and female 7,076) treated at the Radiotherapy Department of Dhaka Medical College Hospital during the period of 1985-92, about 17% of Bangladeshi women suffer from breast cancer [1].

Causes related to breast cancer are still mysterious. As there is currently no means of preventing breast cancer, the focus in reducing deaths from the disease has been on finding breast cancer as early as possible. Early detection of cancer saves patients from the more aggressive radical treatments and increases the overall survival rate. The introduction of screening mammography in 1963 brought about a major revolution in breast cancer detection and diagnosis. Routine screening of women over the age of 40 with X-ray mammography has been demonstrated to contribute to mortality reduction. Studies of the impact of screening have estimated the magnitude of reduction to be between 17-30% [2].

Until the causes of breast cancer are understood, the identification of risk factors for the disease is particularly important. Mammography is one of the best methods for early detection of breast cancer. It reduces the mortality rate by as much as 41% [3]. However various studies [4] have

demonstrated t hat on average a n estimated 2 1 p ercent o f b reast can cers ar e missed b y radiologists during screening mammography. Research showed that Computer Aided Diagnosis (CAD) systems for breast cancer can improve the detection rate from 4.7% to 19.5% compared to radiologists [5]-[8]. A comprehensive review of techniques and algorithms for CAD systems was presented by many researchers [9]-[10].

Abnormalities ex isting in t he m ammogram ar e t he sy mptoms r elated t o b reast c ancer. T he parameters of breast cancer diagnosis by an automated CAD system are shown in Fig. 1.1[11]. Mammographic density/ micro-calcification has been shown to be a strong risk factor for breast cancer and breast cancer risk can be estimated through quantitative analysis of mammographic density. So mass detection and classification can be done by an automatic hybrid feature based method for early detection and prediction of breast cancer.
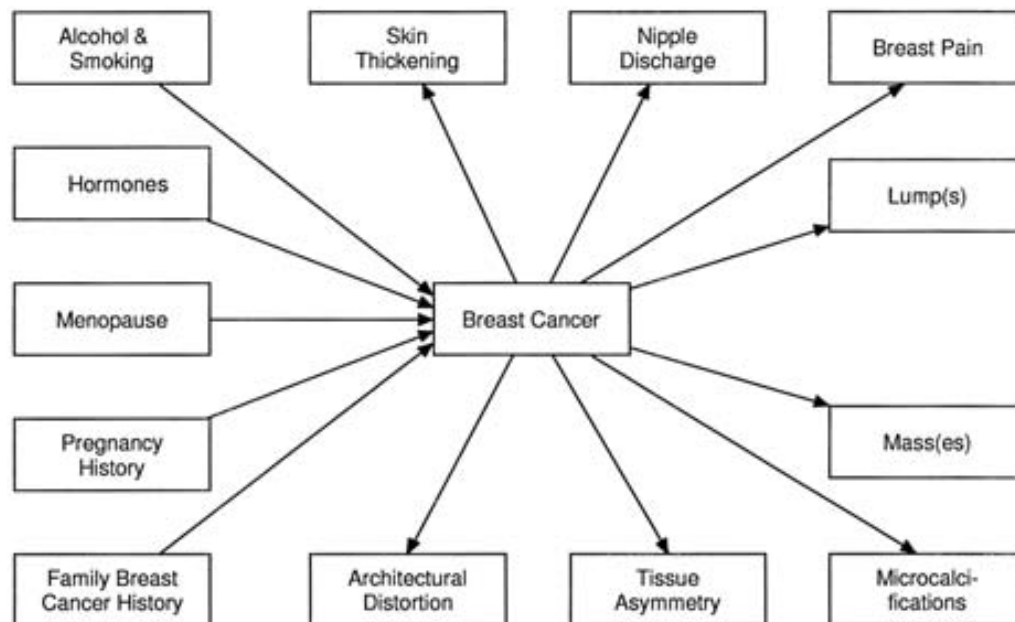
Fig. 1.1: Breast Cancer: Causes and Symptom.

1.2 Terminology

1.2.1 Cancer

The medical term of cancer is malignant neoplasm. It is a class of some diseases like, a group of cells display uncontrolled growth which means division beyond the normal limits, invasion which means intrusion on and destruction of adjacent tissues, and sometimes metastasis which means spread to other locations in the body via lymph or blood. These three malignant properties of cancers differentiate them from benign tumors, which are self-limited, and do not invade or metastasize. Most cancers form a tumor but some, like leukemia, do not [12].

1.2.2 Breast Anatomy

The breast is a mass of glandular, fatty and fibrous tissues positioned over the pectoral muscles of the chest wall. A layer of fatty tissue surrounds the breast glands and extends throughout the breast. Figure below shows the breast anatomical structure and indicates the significant parts [13].
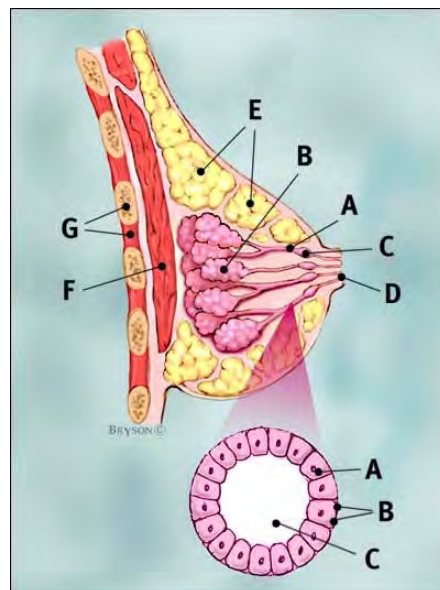
Fig. 1.2: Breast Anatomy.

Breast profile:

A ducts

B lobules

C dilated section of duct to hold milk

D nipple

E fat

F pectoral's major muscle

G chest wall/rib cage

Enlargement:

A normal duct cells

B basement membrane

C lumen (center of duct)

1.2.3 Breast Cancer

Breast cancer is cancers originating from breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk. Cancers originating from ducts are known as ductal carcinomas; those originating from lobules are known as lobular carcinomas. There are many different types of breast cancer, with different stages (spread), aggressiveness, and genetic makeup; survival varies greatly depending on those factors.

1.2.4 Mass

Mass in breast is a localized swelling that feels different from the surrounding breast tissue. It is also known as breast lump. It is a symptom/sign for a variety of conditions. As approximately 10% of breast lumps ultimately lead to a diagnosis of breast cancer.

1.2.5 Microcalcification

Microcalcifications are tiny specks of calcium that can be scattered throughout the mammary gland, or occur in clusters. When found on a mammogram, a radiologist will then decide whether the specks are of concern - usually, this is not the case. Commonly, they simply indicate the presence of tiny benign cysts, but can signify the presence of early breast cancer; for this reason, it is important to attend regular screening sessions, as recommended by health service.

1.2.6 Image Processing

Image processing is one form of signal processing where the input is an image; the output can either be an image or be a set of features related to the input image. Image processing usually refers to digital signal processing. Digital Image Processing is the use of computer algorithms to perform image processing on digital images. Medical images are processed to extract necessary features to use in detection and diagnosis.

1.2.7 Mammography

Mammography [2] is a radiographic examination of the breast. The film mammogram can be considered to be a map of optical density that reflects the composition of the breast. Radiographically the breast consists mainly of two component tissues: fibro-glandular tissue and fat. Fibro-glandular tissue is a mixture of fibrous connective tissue (the stroma) and the functional (or glandular) epithelial cells that line the ducts of the breast (the parenchyma). The remainder of the breast is fat.

1.2.8 Mammographic Density

In terms of X-ray attenuation, fat is more radiolucent than fibroglandular tissue; thus, regions of fat appear darker on a transilluminated radiograph of the breast. Regions of brightness associated

with fibro-glandular tissue are referred to as mammographic density. The hypothesis that breast cancer r isk is as sociated w ith m ammographic d ensity w as p resented in [14] a nd [15]. T his strong relationship between mammographic density and breast cancer risk suggests the causes of breast can cer m ay b e b etter u nderstood b y f inding the f actors t hat asso ciated w ith mammographically dense tissue and how the tissue changes [2].

1.2.9 Classification

Classification of br east l esions a s m alignant or be nign f rom mammograms must be ba sed on information present in the mammogram. Some other clinical data as the Network given in Figure 1 c an a lso be us ed. A c ommon a pproach t o t his t ask is t o e xtract i nformation (f eatures) from mammograms a nd th en u se a m athematical or s tatistical m odel (a c lassifier) to m ake th e malignant versus benign assessment.

1.2.10 Artificial Intelligence

The data extracted and presented for pattern recognition or classification in medical images are most of the time complex, noisy, or incomplete. This leads to the fact that it is difficult to use conventional algorithmic approaches to determine the truth based on a set of predefined rules. So, many m achine l earning m ethods (artificial i ntelligence) ha ve be en t ested a nd used i n the field of medical image processing where computer assisted diagnosis or classification is involved [2]. Machine l earning involves ad aptive m echanisms t hat en able co mputers t o learn f rom experience, l earn by example a nd l earn by a nalogy. L earning c apabilities c an i mprove t he performance of a n i ntelligent s ystem ove r t ime. T he most popul ar a pproaches t o machine learning for medical application are artificial neural networks (ANN) genetic algorithm (GA) and Bayesian belief networks (BBN).
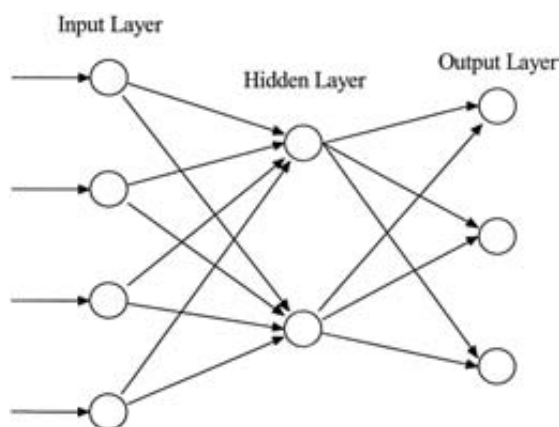
Fig. 1.3: A three-layer feed-forward ANN.

1.2.11 Feature Selection

After t raining an d t esting d atabases es tablished, t he images ar e u sually p reprocessed u sing various techniques of filtering, segmentation and transforming. This is done to define the regions of i nterest (ROI) in the i mage. Then the computer pr ogram c an be us ed t o c ompute or e xtract features from each ROI in the processed images. Many different image features (i.e., intensity-based, geometrical, morphological, fractal di mensional and texture features) ha ve b een us ed in medical i mage p rocessing. F eature ex traction can b e co nsidered as data co mpression that removes irrelevant information and preserves relevant information from the raw data.

1.2.12 Computer aided Diagnosis (CAD) System

Most o f t he p roposed sy stems f ollow a h ierarchical ap proach. A t first, t he C AD sy stem preprocesses the original mammogram to detect the suspicious regions in the breast parenchyma that will be considered for further analysis. This stage is typically designed to operate at a h igh sensitivity to en sure that all possible malignancies are detected accurately. However, this stage also generates some false positive detection. During the second stage, the suspicious areas are scrutinized i n m ore d etail t o eliminate so me o f t he f alse p ositives. T his st age i s designed to

ensure that the results will remain clinically acceptable while reducing false positive rate. Breast masses are common risk factors of breast cancer. It is a challenge to detect the masses accurately both for radiologists and CAD systems. This is due to the great variability in shape, size and contrast.
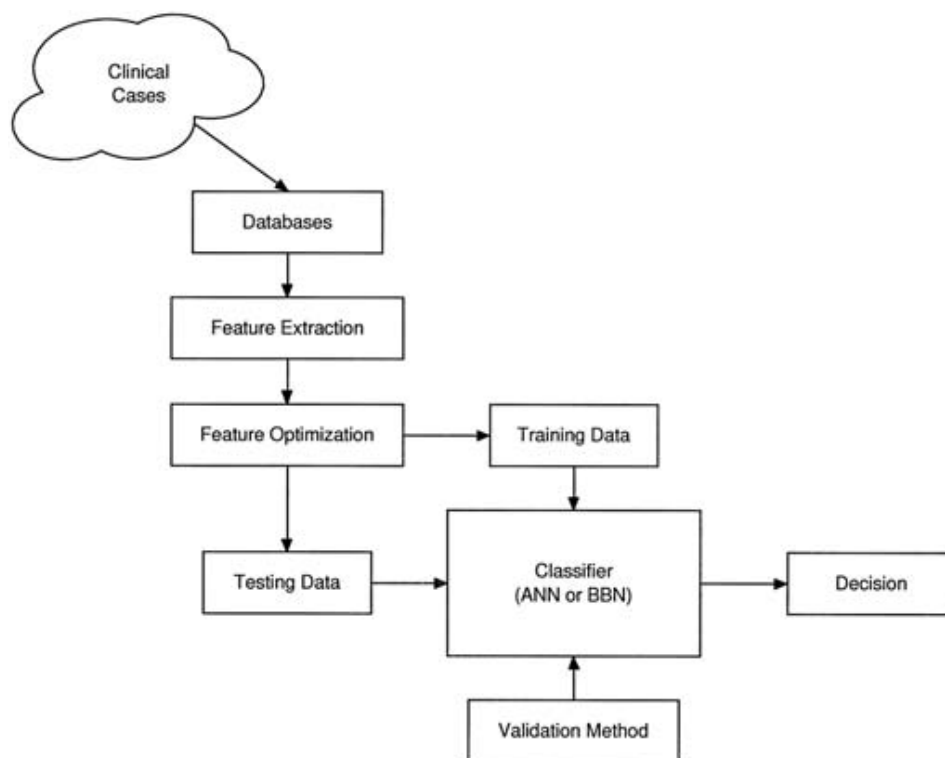
Fig. 1.4: A typical CAD system.

1.3 Aims and Objectives

Mammography is a very important method to analyze breast condition and also to predict breast cancer in its earlier stage. The aim of this thesis is to develop a method for breast cancer detection. To achieve this goal, the total work can be segmented into some sub-sections as follows:

- The breast contour will be first extracted and the artifacts like medical tags etc. will be eliminated from the mammographic images in order to localize the abnormalities.

- The pectoral muscle extraction and removal from the mammogram is a prerequisite to gain more accurate mammographic density detection and classification. Here, the pectoral muscle segmentation will be performed by region grow method.

- Then the mammogram without pectoral muscle will be divided into sixteen-by-sixteen blocks and by taking the Gray Level Co-occurrence matrix (GLCM) of the blocks, the region of interest (ROI) containing possible mammographic density will be extracted for further analysis.

- Intensity-based or area-based or hybrid of both the features based extracted from the segmented region of interest (ROI) of the original mammogram will be used for classification.

- By using Support Vector Machine (SVM) method or two-layer Neural Network (NN) method, the ROI of the mammographic density will be classified as the normal, benign or malignant. Simulations will be performed on Mammographic Image Analysis Society (MIAS) mini database and effectiveness of the method will be shown by different performance parameters and will be compared with some of the existing techniques.

Chapter 2

LITERATURE REVIEW

As there is currently no means of preventing breast cancer, the focus in reducing deaths from the disease has been on finding breast cancer as early as possible. Early detection of cancer saves patients from the more aggressive radical treatments and increases the overall survival rate. The introduction of screening mammography in 1963 has brought about a major revolution in breast cancer detection and diagnosis [16]. It is widely adopted in many countries including Australia as a nationwide public health care program. The decline in the number of breast cancer deaths corresponds directly to an increase in routine mammography screening.

Mammography is one of the best methods for early detection of breast cancer. In order to diagnose breast cancer, it is necessary to detect and classify the Mammographic Density for which segmentation of regions related to cancer is a task of vital importance. Through Region segmentation, the regions, such as skin-line [17], the nipple [18], the denser tissue, and the pectoral muscle having the possibility of carrying cancer are extracted from the original mammograms.

Some previous works related to different region segmentation can be demonstrated under following sub-sections:
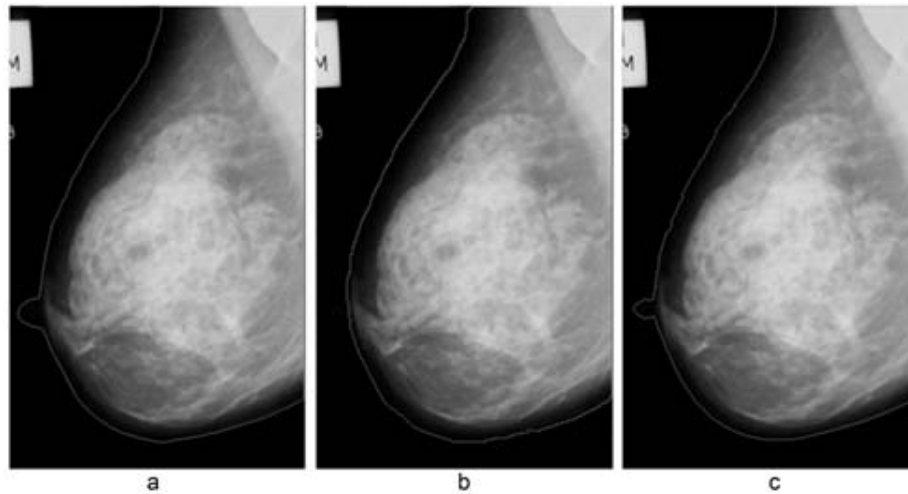
Fig. 2.1: Mammogram mdb125 breast border, (a) extracted by an expert radiologist, (b) estimated by active contours modeling, and (c) detected by local adaptive thresholding method.

2.1 The Skin-line Detection Methods

The skin-line or breast border detection as well as breast segmentation are important key factors which c an i mprove clinical d iagnosis of m ammographic di seases. F urthermore, br east segmentation is an important prerequisite factor of CAD methodology of breast cancer in digital mammograms. Histogram thresholding methods can be considered as one of the first techniques of t he br east r egion e xtraction [19]. I n [20], t hresholding m ethod h as be en improved. A n enhanced s nake a lgorithm ha s be en us ed i n [21]. I n [22], s kin-line h as b een seg mented u sing Hausdorff Distance Measure (HDM) [23], while in [24], an adaptive method has been used. Fig. 2.1 [24] depicts breast contour detection in a clear way.

In the pre-processing part, the mammogram is digitized first. Then breast border line is outlined as the preceding part. As discussed in previous paragraph, only few approach, such as: histogram

thresholding m ethod, s nake a lgorithm, a daptation m ethod e tc ha ve b een us ed i n t his f ield. Anyone of these methods can be used to detect the breast border first.

Table 2.1: Comparison Study of Breast Contour Detection.

Method	Percentage of Accuracy
Wavelet Decomposition [29]	100%
Hausdorff Distance Measurement [22]	98%
Local Adaptive Method[24]	86%

2.2 The Nipple Detection Methods

There are two possible cases of visible nipple positions in mammogram, outside and inside of the breast profile as shown in Fig. 2.2 [25]. The accurate nipple position can be a very important parameter in the process of mammogram registration. In [26], the nipple detection method using combination of searching the breast boundary and the identification using texture convergence analysis ha s be en pr esented. T he ge netic a lgorithm a pproach f or br east bor der a nd ni pple position de tection i n o rder t o r egister m ammograms ha s been pr esented i n [27]. T he R adon-domain ni pple de tection a pproach ha s be en pr esented i n [28]. T he method us ed i n [29], i s intensity based and certain morphological features of breast images has been used.

As asy mmetric n ipple position in left a nd r ight br easts i s c onsidered a s one o f t he main symptoms o f br east c ancer, ni pple de tection c an be done t o pr edict r isk f actor. I n t his r egard, genetic a lgorithm, R adon-domain a pproach/ i ntensity ba sed a pproach, w hich a re us ed a s indicated in section 2.2. A method has been described in [22], that has found breast contour with 98% accuracy while method described in [24], has achieved 86% accuracy in contour detection.

But method used in [24] has an extra facility for nipple extraction with 94% accuracy. Overall nipple detection rate by the method in [29], is 97.92% with the accuracy better than 1 mm for 56.03% of images.

(a)

(b)
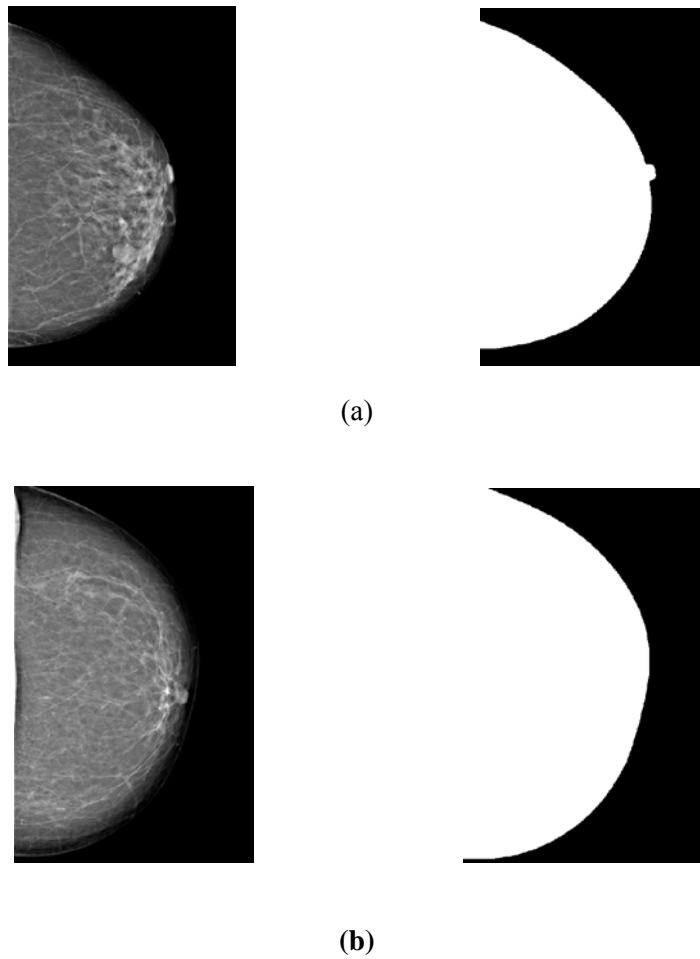
Fig. 2.2: (a) Case of nipple outside the breast profile, (b) Case of nipple inside the breast profile.

Table 2.2: Comparison Study of Nipple Detection.

Method	Percentage of Accuracy
Local Adaptive Method [24]	94%
Intensity Based Method [25]	97.92%

2.3 The Pectoral Muscle Segmentation Methods

A threshold-based algorithm has been presented in [30] for the extraction of the pectoral muscle edge in MLO view mammograms. They have applied minimum cross-entropy thresholding algorithm to local areas around the pectoral muscle to determine a series of thresholds as a function of area size.

Using a model image it has been shown that an inflection point in this function corresponds to a threshold that separated the pectoral muscle from the rest of the breast. Post processing has been performed on mammograms to eliminate false positive points of inflection and a straight line has been fitted to the detected pectoral boundary in order to smooth roughness caused by the non-uniform intensity of the pectoral muscle edge.

A model-based algorithm has been described in [31], to segment the pectoral muscle in the mammograms. After the mammogram has been preprocessed, a set of ROI with different sizes has been applied on it, in each of which, an iterative thresholding technique has been used to gain an optimal threshold correspondingly. All the thresholds has been composed the threshold. Finally, a twice line fitting and polygon approaching technique has been carried out to refine and

approach t he ed ge cu rve o f t he p ectoral m uscle, w hich has b een partly s egmented by t he thresholding.

A new algorithm for automatic segmentation of the pectoral muscle in mammograms has been proposed i n [32]. T heir method, ba sed upon Gabor f ilters, ove rcomes t he l imitation of t he straight-line hypothesis. With reference to the number of pixels in manually demarcated pectoral muscle regions, the segmented regions provided by the Gabor-filter-based method has resulted in average F P a nd F N r ates of 0.58% a nd 5.77% , r espectively. F urthermore, t he r esults of t he Gabor-filter-based m ethod h ave i ndicated l ow H ausdorff d istances w ith r espect t o t he h and-drawn pectoral muscle edges, with the mean and standard deviation being 3.84±1.73mm over 84 images.

A method has been described in [33] for fully automatic segmentation of the pectoral muscle on MLO v iew m ammograms, c ontaining tw o m ajor p arts: (a) s traight lin e e stimation a nd (b) c liff detection. Straight line estimation has included iterative thresholding and straight line fitting with gradient test. This estimate of the pectoral edge has validated at least twice. Cliff detection [34] has be en us ed t o e stimate t he s traight l ine i n conjunction w ith s urface s moothing a nd e dge detection to yield the final segmentation. It has been found that bicubic spline [35] interpolation is u seful in smoothing th e in tensity s urface w hile s igmoidal m odel f itting to in tensity p rofiles across the e stimated ed ge h elp i n l ocating the pectoral ed ge m ore accu rately. T he m ethod i s adaptive t o l arge v ariations in ap pearance o f t he p ectoral m uscle a nd m argin. T he m ethod remains effective when parts of the pectoral edge are obscured by superimposed glandular tissue or a rtifacts. The m ethod ha s b een tested ou t on the 322 di gitized m ammograms of t he M IAS database an d t wo mammographic r adiologists have assess ed the seg mentation r esults. T heir

findings show that segmentation accuracy has been improved after refining the straight line into a curve using iterative cliff detection and that 83.9% of the curve segmentations are adequate or better [36].

A graph theoretical method based on minimum spanning trees (MST) [37] has been used in [38], to recognize the pectoral muscle in screening mammograms. The segmentation found using the MST has been used to initialize an active contour for finding an anatomically reasonable estimate of the boundary of the pectoral muscle. The error has been reported in terms of the number of incorrectly assigned pixels. Out of 83 images, 25 images have error rates less than 5 percent and 56 images have error rates less than 10 percent.

A new approach for pectoral muscle segmentation has been presented in [39], based on nonlinear diffusion filtering. They have applied their method to 90 mammograms from Mammography Image Analysis Society (MIAS) database [40] and have compared obtained results by those recognized by two expert radiologists. To evaluate the accuracy of proposed method, Hausdorff Distance Measure (HDM) [41] and Mean of Absolute Error Distance Measure (MAEDM) are used. Then their results are also compared with two pectoral muscle segmentation methods proposed in [42] and [32]. The first is based on Hough-Transform and the second is based on Gabor-Filters.

A novel pectoral muscle segmentation algorithm has been proposed in [43], which has overcome a few defects of the conventional methods, and has got a generally high precision when dealing with the very different featured pectoral muscles. It has carried out a series of ROIs upon the neighborhood of the pectoral muscle, has computed the optimal threshold curve and local mean

square deviation (MSD) cu rve, h as at tained t he seg mentation t hreshold acco rding t o the g ray-level distribution of the mammogram, has used the zonal Hough transform and polyline fitting to approach the boundary of the pectoral muscle, and finally has applied an elastic thread algorithm to extract the pectoral muscle accurately.
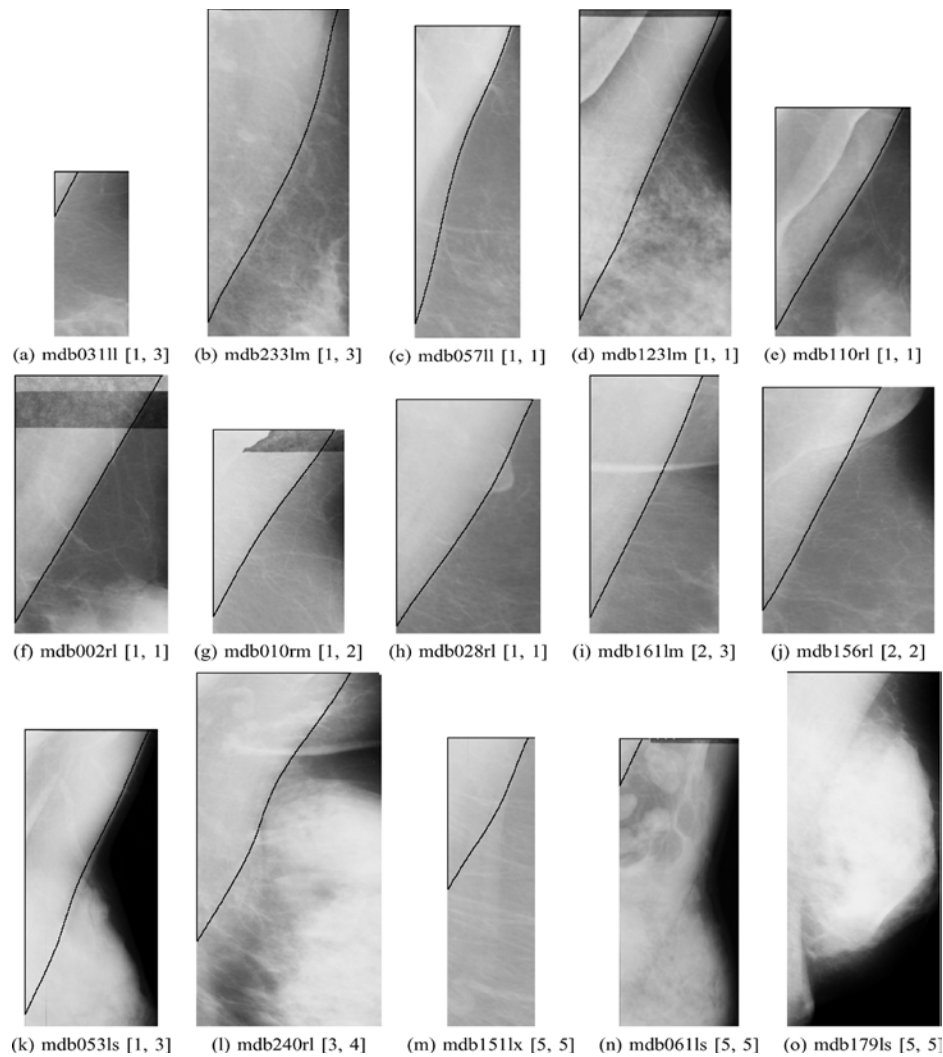
Fig. 2.3: Segmentation results on MLO mammograms from the MIAS database obtained in [40]. The scores of each image rated by two radiologists are shown in $[r_1; r_2]$. All images are shown in the same scale.

Pectoral muscle segmentation follows the pre-processing part. This segmentation is done based on s ome a ssumptions; i) t his m uscle i s a lways pos itioned i n up per left/right c orner of Mediolateral Oblique view (MLO) mammograms, ii) it can be segmented as assuming a straight line a nd t o get be tter r esult a nd i ii) c urve f itting c an be a dapted. B y u sing I CA method [53], average c lassification r ates o f 9 7.3% f or p ectoral m uscle can b e o btained. B y ad aptive cu rve fitting [36], 83.9% accuracy is found in segmenting pectoral muscle. Although method described in [33], ha s got r esult of 94% a ccuracy us ing s traight line a pproximation a nd iterative c liff detection method, the method is not efficient enough for assumptions used. After segmenting this pectoral muscle, the remaining part of the mammogram can be considered as the ROI (region of interest).

Table 2.3: Comparison Study for Pectoral Muscle Detection.

Method	Percentage of Accuracy
Wavelet Decomposition [29]	85%
Curve fitting Method [36]	83.9%

2.4 The Denser tissue Detection Methods

Mammography is a routine examination for screening. It is one of the most important areas in which image processing procedures are usefully applied. One of the challenging aspects of the medical image processing is to extract features from the images that represent efficiently their diagnostic and visual information content. Identification and feature extraction of denser tissue is the core part of the mammographic analysis.

A data set of 290 digitized mammograms has been analyzed in [44], with no lesion and 98 images with malignant mass or micro classifications clusters. The mammograms have been categorized by radiologists using the BI-RADS breast density index 1, 2, 3, or 4 related to fibro glandular tissue composition. They have transferred the dicom images which are digitized by the CAD system first and are analyzed by an neural network, in a PC and are saved in an uncompressed bitmap format in order to be processed by Interactive Data Language (IDL) and/or MATLAB program tools. They have created oval ROIs around a massive lesion while rectangular ROIs around micro-calcifications cluster. The texture features have been derived for each sub-region from an averaged gray level co-occurrence matrix. The Statistics of gray level histograms give parameters for each processed region. They have calculated the co occurrence matrices for 5 separation vectors in the horizontal direction; that is, the mean, variance, skewness, kurtosis, and entropy of the histogram were computed. 10 parameters (homogeneity and energy from 5 co-occurrence matrices) describing each image were calculated and the reduction of factors by principal component analysis results to three factors. Factors mainly depend on energy for 20 pixels distance, homogeneity for 5 and 25 pixel distances. They extracted the two second order statistical parameters (homogeneity, energy) from each matrix and the two first order parameters (mean and standard deviation). An attempt was made to characterize texture by second order statistical parameters and there was correct classification for 84% of total cases. Better classification rate was obtained for non homogeneous texture. Their results show that quantitative analysis of mammography provides an objective way to extract and interpret image data. They indicated the usefulness of texture feature extraction and the ability to quantitatively differentiate healthy from pathological breast tissue.

Texture is a feature that cannot be defined for a point, and the resolution at which an image is observed determines the scale at which the texture is perceived. This leads to the fact that texture is a confusion measurement that depends mainly on the scale which the data are observed. In [45] it is proposed that by using a wavelet transformation of the data, a pattern recognition solution can be devised, where the features are selected directly from the wavelet decomposition, denoised in a tailored way to accomplish a separation between the classes. The Db4 and Haar wavelets are the basis used in the decomposition process and 100 greatest coefficients in magnitude in the first level of decomposition were considered.

A novel technique has been developed in [46], to automatically identify a region of interest (ROI) surrounding a speculated lesion on a mammogram. They have approached to determine the size of the ROI depending on the response of a set of unique Spiculation Filters (SF). They have designed these filters which are based on manually annotated physical characteristics of spicules. Spiculation Filters are a new class of complex quadrature filters made up of the cosine, $fc(r, \theta; r0, \sigma, \omega)$, and sine, $fs(r, \theta; r0, \sigma, \omega)$, components. The spiculation filters have three parameters. 1) radius 2) frequency and 3) σ. The radius of the filter is the size parameter measured in pixels and its value corresponds to the length of the spicules. The frequency corresponds to the number of spicules located per circumference of the central mass region. σ is the standard deviation in pixels. They have calculated the percentage of pixels marked as spicules by the radiologist which are located inside the ROI detected by their algorithm. On average, 94% of spicule pixels were located inside the ROI. Then they have used an area overlap criteria to determine the overlap in area between the ROI identified by their algorithm and the

area that i s m anually i dentified as b est. On av erage, o verlap p ercentage o f t heir R OI identification method is 48%.

A multi-resolution approach to automated classification of mammograms has been proposed in [47]. They have used Gabor filters of different frequencies and orientations to extract textual patterns of mammograms. Statistic t-test and its p-values for feature selection and weighting are proposed to increase classification efficiency and reduce feature space. A Gabor filter can be seen as a sinusoidal plane of a particular frequency and orientation, modulated by a Gaussian envelops [48]. A Gabor filter tuned to that frequency range exhibits a strong response, but a significantly weaker response of a different texture [49]. Therefore, an important step for mammogram classification using texture is to select appropriate Gabor filters that can tell the texture differences between normal and abnormal mammograms. But it is difficult to select appropriate filters for mammograms. They converted the filter selection to feature selection based on t -test statistics. To apply t-test in feature selection, they have first applied n Gabor filters with different orientations and frequencies, which have produced a feature vector of $2n$ dimensions. Then, t-test statistics has been performed on each feature. The feature with p-value of t-test less than 0.05 is significant, i.e., the feature can tell the difference between the normal and abnormal mammograms and so it was kept. Otherwise, the feature has been discarded. They have applied the approach used in [50] for feature weighting to differentiate the selected features. The approach not only alleviates the problem of low resolution and strong noise, but also reduces the feature space without affecting the classification accuracy.

A software program has been prepared in [51] to localize the abnormalities using information associated within the d ata f iles. T hen t hey h ave b een ab le t o p resent the st andardize

mammograms with the region of interest highlighted. The 32x32 pixel region of interest is then determined around the center of the abnormalities. Wavelet decomposition [52] is applied over these regions and the statistical features and wavelet coefficients vectors are then extracted. Also, 1st order statistics, median contrast and local binary partition features are calculated from ROI texture. These features are then presented to both the voting k-nearest and minimum distance classifiers. A scaling between (0-1) is made to judge the normality and abnormality of the imaged tissue. K-NN classifiers are much better than minimum distance classifier, also there are no considerable differences between K-NN classifiers using Euclidean distance (ED) and K-NN classifiers using sum of differences (SOD) with different values of k. Finally K-NN classifier using SOD and K= 1 has showed better results than other classifiers.

For classification part, Artificial intelligence based approaches, such as neural network (NN) and Fuzzy Logic, Support Vector Machine (SVM) and data mining rules can also be used.

Table2.4: Methods used for Mass Detection, Feature Extraction and Classification Sections.

Section	Methods
Mass Detection and Feature Extraction	Wavelet Transform [54]-[56], Non-linear Filter [57], Polygonal Modelling [58], Topographic Approach [59], Morphological Component Analysis [60], Concentric Morphology Model [61] Twin Support Vector system [62]
Classification	Neural Network [63]-[64], Fuzzy Neural Network [65], Genetic Algorithm, Decision Tree [66], Evolutionary Algorithm [67]

Chapter 3

A HYBRID FEATURE BASED METHOD FOR EARLY DETECTION AND CLASSIFICATION OF MAMMOGRAPHIC DENSITY IN BREAST CANCER

In t his c hapter t he m ethodology ha s be en di vided i nto f our parts: P re-processing, P ectoral Muscle S egmentation, D etection o f Ma mmographic D ensity an d C lassification. I n P ectoral Muscle Segmentation part, breast contour detection has been done in image pre-processing. The remaining works are focused on Detection of Mammographic Density and classification.

The proposed method described throughout this present work can be concisely given in the flow-chart of Fig. 3.1.
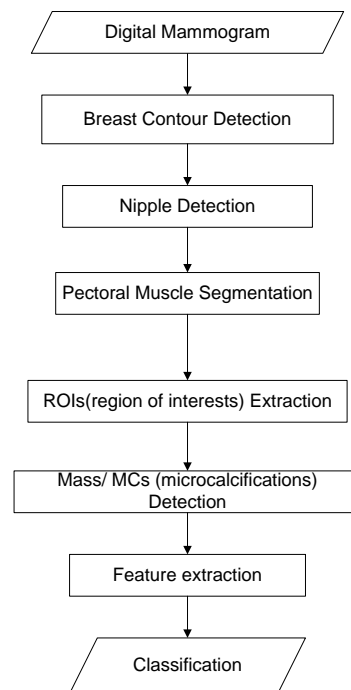
Fig. 3.1: Flow-chart of mammogram screening system design.

3.1 Pre-processing

3.1.1 Breast Contour Detection

Before t he actual p ectoral m uscle s egmentation p rocess b egins, i t i s n ecessary t o prepare t he mammograms for t hat by e liminating a ll the background o bjects, i.e., marking s tickers that a re not part of the breast. To achieve this objective, the whole mammogram has been converted to binary image from gray image using a very low threshold (0.1) and the binary image has been labeled. A s the ba ckground of t he m ammogram i s a lmost bl ack, this binary image c ontains all the c omponents i ncluding t he w hole br east, m arking s tickers e tc. O f a ll t he c omponents, t he breast h as the m aximum ar ea an d i t is easily d etectable b y co mputing the a reas o f t he components. T hus t he br east ha s be en s elected and t he i mage w indow has be en c ropped t o contain onl y t he b reast. T he or ientation of the i mage ha s been a djusted s o that t he pe ctoral muscle region is always positioned in upper left corner. Fig. 3.2(a)-(c) show these steps.

(a)   (b)   (c)
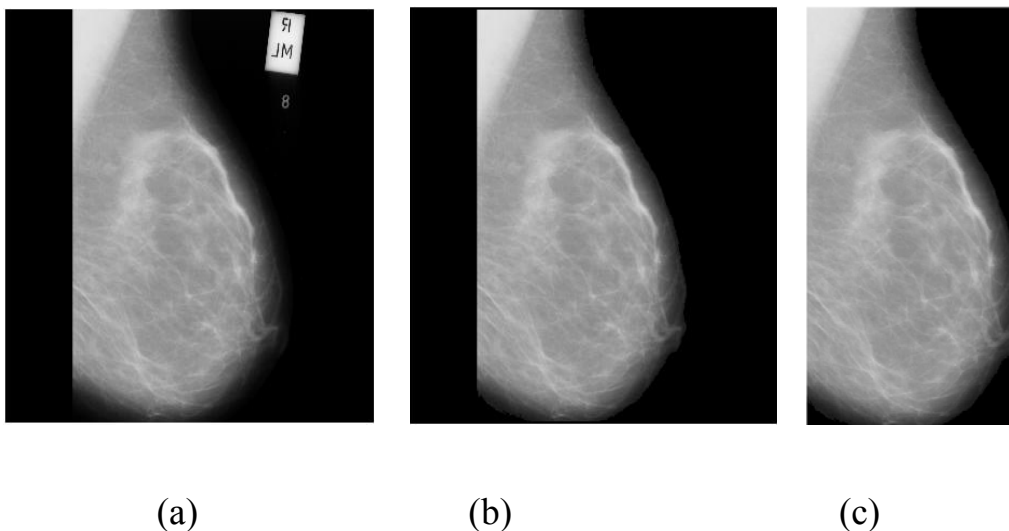
Fig. 3.2: Step-by-step procedure of delineation of the background and medical tags and other unnecessary parts from the mammogram [mdb020.pgm].

3.1.2 Pixel Value Mapping

The p ixel v alues o f t he s maller i mage t hat contains o nly t he b reast h as b een mapped t o an exponential function. L et *I(x, y)* is t he pi xel va lue of a ny c oordinate i s t he pi xel va lue of a ny coordinate *(x, y)*, then the output pixel value *I'(x, y)* the mapping would be:

$$I^{'}(x,y) = \frac{e^{I(x,y)}}{40} \qquad (3.1)$$

The pixel values of the image have been exponentially mapped according to equation (3.1). The high magnitude va lues ha ve be en c onverted t o t he va lues de picted i n F ig. 3.3(a). T he out put image, F ig. 3.3(b), has so me h omogeneous r egions w here al l the p ixel values ar e s ame. At t he same tim e it d ivides th e im age in to s everal re gions w ith d ifferent p ixel values. A s ingle re gion encapsulates ex actly a m ajor portion of the pectoral muscle and so metime ex tends to the breast. The regions within pectoral muscle have pixel values such that the region at the upper left corner has t he maximum pi xel va lue. T he ne xt r egion ha s a pi xel va lue l ower t han t hat and f or t he remaining regions the same rule can be applied. When the image is converted into binary image using the pixel value of pectoral muscle region as threshold, the output image has an object which has almost the same shape and position of the pectoral muscle. This is the most important feature of mapping. After some morphological processing we obtain the picture depicted in Fig. 3.3(c).
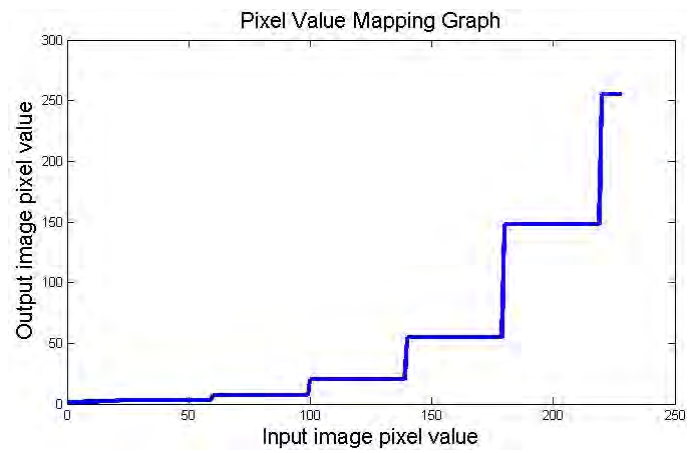
(a)
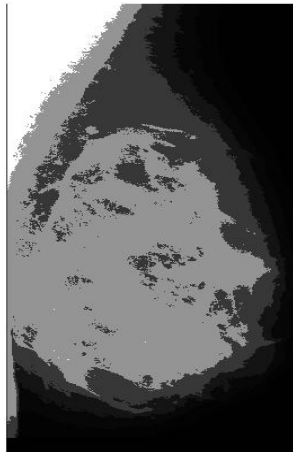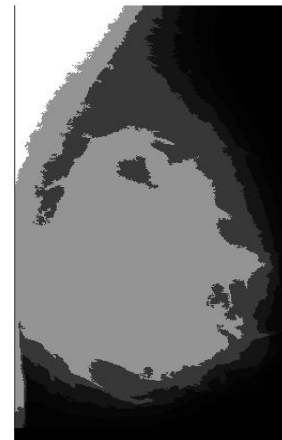
(b)   (C)

Fig. 3.3: (a) Graphical representation of the exponential mapping, (b) Mapping the image to get a

shape of the pectoral muscle, (c) Output image after the morphological operation [mdb311.pgm].

3.1.3 Morphological Processing and Pixel Value Equalization

As described earlier, masses have a multiple concentric layer characteristic, so the region grow method is effective in detecting mass. To take the leverage of this characteristic, some morphological operations have been performed on the mammogram to make the pixel values same for each layer to reduce computational complexities. Pixels in each layer are strongly connected. It is required to find out these strongly connected pixels to find out the different layers or regions. So the image has been morphologically opened with a disk shaped structural element. Let this image be denoted as I_{open}. With the same structural element the image, I_{open} has been morphologically closed. Let the image after closing be denoted as I_{close}. Mathematically the opening process can be written as:

$$A \cdot B = (A \odot B) \oplus B \qquad (3.2)$$

And the closing process can be written as:

$$A \cdot B = (A \oplus B) \odot B \qquad (3.3)$$

Where A =Mammogram, B =Structural element and $\odot$ and $\oplus$ represent erosion and dilation respectively. Dissimilar small objects in a layer are removed by opening while small holes in a layer are removed by closing operation. Morphological erosion operation has also been performed on the mammogram to remove local area boundaries. Let this image be denoted as $I_{erosion}$. Mathematically this process is:

$$A \odot B = \cap_{b \in B} A_{-b} \qquad (3.4)$$

The image, $I_{erosion}$ has been taken as marker image and the image, I_{CLAHE} has been taken as mask image. A new image has been reconstructed morphologically after repeated dilation of the marker image until the contour matches the mask image. Let the image after reconstruction be denoted as $I_{reconstruct1}$. The mathematical interpretation of morphological reconstruction is given below [71]:

If marker image is denoted by J and the mask image by I, both images are identical in size, and

$J \leq I$

The classical grayscale dilation of J with structuring element B is given by

$$\delta(J) = J \oplus B \tag{3.5}$$

The symbol $\oplus$ used for the dilation operation. The geodesic dilation of size 1 of the marker image J with respect to mask images I can be defined as:

$$\delta_1^1(J) = (J \oplus B) \wedge I \tag{3.6}$$

In this equation, $\wedge$ stands for the point-wise minimum between the dilated marker image and the mask image, $(J \oplus B)$ is the dilation of J with the elementary isotropic structuring element B. The geodesic dilation of size n of the marker image J with respect to a mask image I is obtained by performing n successive geodesic dilation of size 1 of J with respect to I

$$\delta_1^{(n)}(J) = \delta_1^1(J) \cdot \delta_1^1(J) \cdot ... \delta_1^1(J) \tag{3.7}$$

Equation (3.7) defines the morphological reconstruction by geodesic dilation of the mask I from marker the J. The desired reconstruction is achieved by carrying out geodesic dilations until stability is reached [72]

The image, $I_{reconstruct1}$ has been dilated with a disk shaped structuring element .Let this image be denoted as $I_{dilation}$. Mathematically this process is:

$$A \oplus B = \cap_{b \in B} A_b \tag{3.8}$$

Again a new image has been reconstructed with the complement of $I_{dilation}$ as marker image and the complement of $I_{erosion}$ as mask image .This new image has again been complemented to get the image whose layers have more distinct pixel value. Fig. 3.4 shows the image after pixel value equalization.
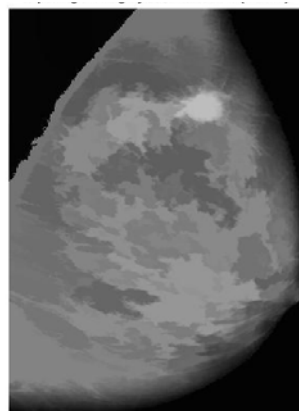
Fig. 3.4: Image after pixel Value Equalization.

3.2 Pectoral Muscle Segmentation

The pectoral muscle in mammograms acts as an additional complexity in automated analysis using CAD systems. The texture of the pectoral muscle may be similar to some abnormalities and thus may cause false positive results in detection of suspicious mammographic density [68]. Elimination of the pectoral muscle can improve automatic mammographic density identification and allow the radiologists to check for the development of cancers in the area overlying the pectoral muscle [36]. Hence, the elimination of pectoral muscle is very important in automated detection of mammographic density in mammographic images.

A novel and efficient pectoral muscle segmentation method has been proposed here using pixel value mapping to delineate pectoral muscle region accurately. The method achieved a high degree of accuracy in approximating the pectoral muscle in mammograms. This method has been found to be robust not only to large variations of size, shape and positions of pectoral muscle, but also to any kind of artifacts like medical tags.

In mass detection section, the emphasis has been given to decrease the false positive results in detecting the mammographic density. To obtain better result, medical tags and extra background have been eliminated and then pectoral muscle has been detected accurately. My proposed method is capable to eliminate the pectoral muscle with greater accuracy so that it can not interfere in the detection of exact mammographic mass/ density.

This methodology assumes following hypotheses,

     i)   The pectoral muscle is positioned in the upper left or upper right corner of the mammogram.

     ii)   The grayscale intensity of the pectoral muscle is higher than its surrounding tissues.

The methodology to segment out the pectoral tissue in digital mammogram can be divided into two sections:

i)   Finding proper threshold value for the preprocessed image using an iterative method and obtaining binary mask.

ii)   Processing the mask to be fitted for the intended pectoral tissue region and segmenting the pectoral tissue.

The flow chart of the algorithm is shown in the Fig. 3.5.

```
┌─────────────────────────────────────────┐
│          Digital mammogram               │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│        Mammogram preprocessing           │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  Pixel value mapping to exponential equation │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│        Morphological processing          │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│        Get proper threshold value        │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│            Get binary mask               │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│        Reshape the binary mask           │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│        Pectoral muscle region            │
└─────────────────────────────────────────┘
```
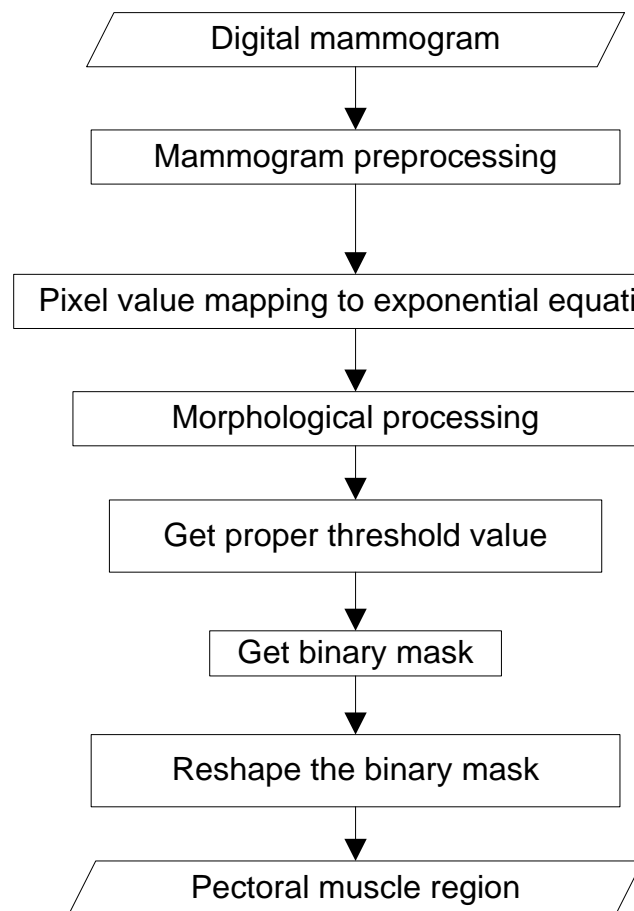
Fig. 3.5: Flow chart shows the steps for the algorithm of automatic pectoral muscle segmentation.

The pectoral margin is not a straight line as hypothesized by others ([69], [70], [32] and [42]) rather its margin that appears on the side of the nipple is mostly concave for upper portion and concave, or convex or a mixture of both for lower portion of the margin, moreover the variation of the margin from one mammogram to another becomes so complex that it cannot be generalized with any single geometrical or mathematical model. The other two sides of the pectoral muscle region are almost always straight lines with few exceptions. For the aforementioned reasons, a new approach to segment the pectoral muscle has been proposed here. The method searches the pectoral muscle by region growing method and verifies the region whether it encapsulates the pectoral muscle exactly or not. If not then it adjusts the region to encapsulate the desired region. The subsequent sections describe the methodology more elaborately.

3.2.1 Threshold Selection Using Region-Grow Method

As can be seen in Fig. 3.6(a)-3.6(c) there are only few pixel values that repeat in the whole image after pixel value mapping and morphological operation, but the boundary of the image is almost intact. At this point the difficulty to get a universal threshold to segment out the pectoral muscle region reduces a great deal. This is because of the existence of only five values as 255, 148, 55, 20 and 7, from which the proper threshold value has to be selected. Regions with other pixel values do not contain much information. Proper threshold value has been selected in two steps. In the first step two candidates for threshold have been selected and in second step final threshold has been selected from the two candidates.

It is assumed that the pectoral muscle region is in the upper left corner of the image. The regions within the pectoral muscle converted to binary image with its pixel value as threshold are of same shape and the areas of the regions are very close to each other. But when the region encapsulates a portion of the breast, the area increases rapidly compared to the area that lies within the pectoral muscle region. The rapid increase in area denotes the boundary of the

pectoral t issue an d t he breast t issue. T his r apid i ncrease i n b oundary may b e cau sed b y t wo reasons:

i) The area has covered exactly upper portion of the pectoral muscle and an area of br east a t the lower pa rt as in F ig. 3.6(a). I n this c ase, t he threshold va lue f or which this area has been obtained is the proper threshold value.

ii) The area completely outreaches the pectoral muscle region and there is no exact coverage for pectoral muscle as in Fig. 3.6(b). In this case, the 1st higher threshold value of t he t hreshold f or w hich t his a rea ha s been obt ained i s t he pr oper t hreshold value. In this case, the area obtained with 1st higher threshold value of the threshold for which this area has been obtained exactly covers the pectoral muscle.

(a)

(b)
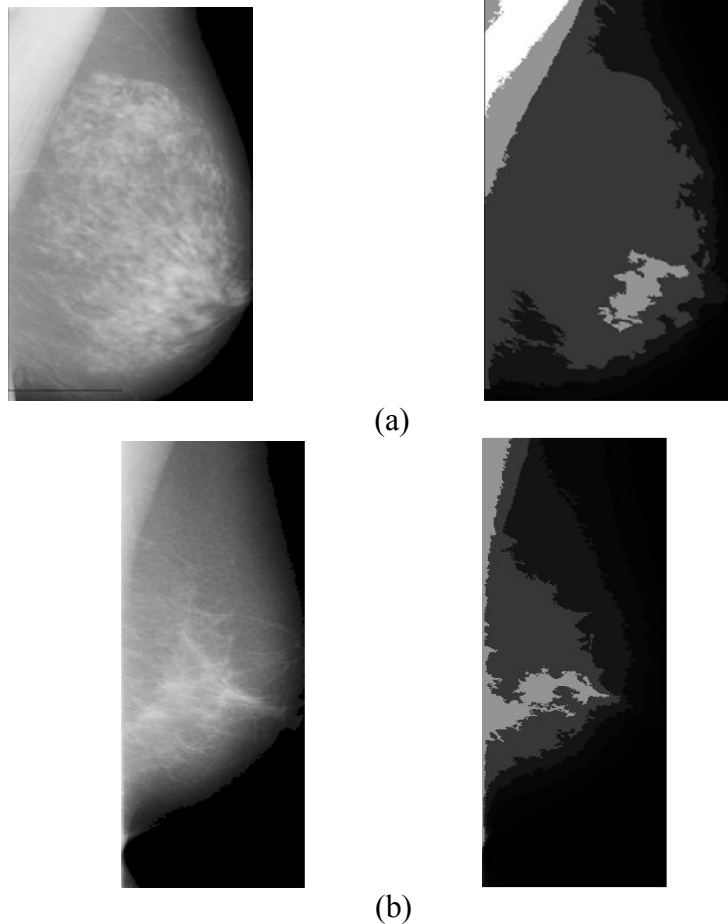
Fig. 3.6: (a) In case of mdb234.pgm current threshold is the proper threshold value; (b) In case of mdb201.pgm previous threshold is the proper threshold value.

Thus the threshold value for which the area the object in the binary image increases rapidly is a potential candidate for proper threshold.

And at the same time the previous threshold value is also a candidate for proper threshold value. The optimum value for the percentage of area increase of the objects in the converted binary image within the pectoral muscle region has been found to 200%.

(a)   (b)

(c)   (d)

(e)
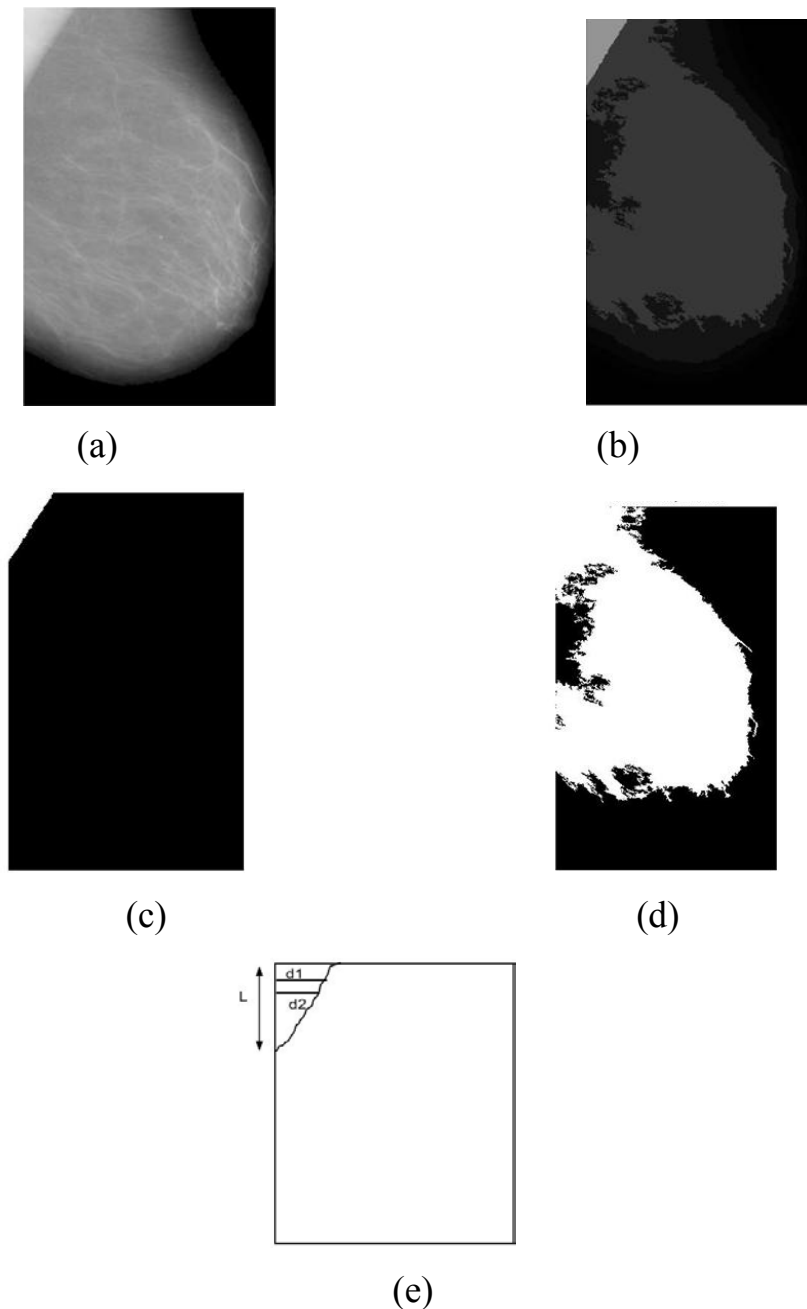
Fig. 3.7: (a)-(d). Step-by-step procedure to choose proper threshold value for using to mask the pectoral muscle region [mdb310.pgm], (e) Positional description of Pectoral Muscle.

The following algorithm has been developed to find the two potential candidates of proper threshold:

i) Get the binary image with the maximum pixel value in the image and get the area of the component appeared in the upper left corner.

ii) Get the binary image with the next maximum pixel value and get the area of the component appeared in the upper left corner.

iii) Compute the percentage of area increment of the areas obtained in step *i* and step *ii*.

iv) If percentage of area increment<200%, then maximum pixel value=next maximum pixel value, repeat steps *i* to *iv*. Otherwise go to step *v*.

v) First threshold candidate=maximum pixel value.

And, second threshold candidate= next maximum pixel value.

From these two threshold values the exact threshold value is selected using the observation that upper part of the pectoral muscle region margin facing the nipple is almost concave. Let *L* be the total length (in terms of pixel) the region obtained using first threshold candidate along the vertical axis of the picture. Then a window with vertical distance from upper leftmost point = 40% of *L* and the full horizontal axis has been chosen for the validation area of the second threshold candidate. This window has been divided into smaller windows of same horizontal distance but vertical distance = 5 pixels only. So number of smaller windows = L/5. In this grid, the region obtained using the second threshold value has been superimposed. The distance from the vertical axis to the region termination point has been determined in the upper and lower edges of the smaller windows. According to the assumption, within a single smaller window the upper edge distance (d1) should be greater than the lower edge distance (d2) if this region exactly matches the upper part of the pectoral muscle, otherwise the (d1-d2) will be negative or (d1-d2) will be randomly negative and positive for different smaller windows. The scenario can be clearly perceived in Fig. 3.7(a)-(d). Again there are local maxima and minima that may cause

some ne gative (d1-d2) for t rue pe ctoral muscle. T he d1' s and d2' s a re c omputed f or a ll t he smaller w indows an d the (d1-d2)'s compared for a ll the 322 mammographic pictures of MIAS mini database. The optimum acc eptable tolerance of negative (d1-d2)'s has been found 45% of M. T his m eans, for M number of s maller w indows, i f t he t otal num ber of ne gative (d1-d2) is 0.45xM, then the second threshold candidate is the proper threshold; otherwise the first threshold candidate is the proper threshold.

3.2.2 Binary Musk Creation

The gray image obtained after the morphological processing has been converted to binary image using the proper threshold obtained in the previous section. This binary image has been labeled and t he upper l eft corner co mponent h as b een sel ected as b inary m ask f or p ectoral m uscle segmentation. B ut t his mask may cover so me g landular tissue o f b reast as d escribed i n t he previous s ection. T o m ake t his m ask of t he s ame s hape a s of t he pe ctoral m uscle r egion, t he following steps have been adopted.

L has been recomputed for the region obtained using the proper threshold value. And (d1-d2) are computed for the whole length of L. If (d1-d2) is found to be positive, the region for that smaller window has been kept intact. Otherwise last positive (d1-d2) has been deducted from the current d2, e.g., new d2= old d2 – last positive (d1-d2). A new window has been selected that lays at the right s ide a nd l ower s ide of ne w d2. I f a ny por tion of t he r egion i s f ound t o b e w ithin th is window, that portion has been omitted. The process has been applied for the whole length of L. Fig. 3.8(a)-(e) shows the process.

(a)   (b)   (c)

(d)   (e)
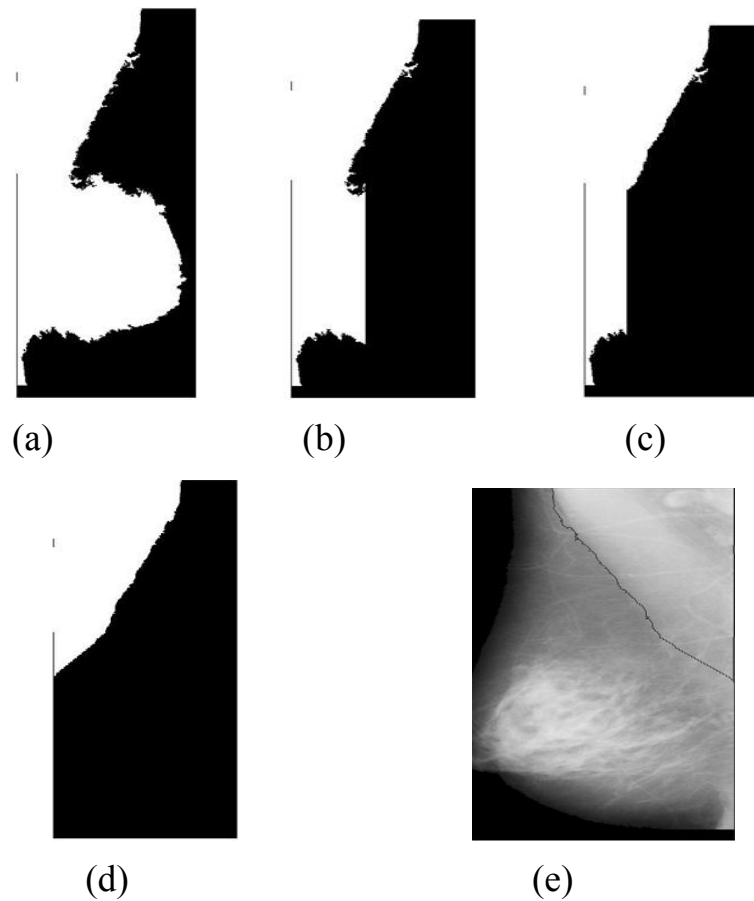
Fig. 3.8: (a)-(e). Step-by-step procedure to create binary mask.

The binary mask obtained here is of the same size and shape of the pectoral muscle region. This mask has been applied in the image obtained in the preprocessing section to segment out the pectoral muscle.

3.3 Mammographic Density Detection

3.3.1 Image Enhancement of ROI

After image preprocessing phase there remains only the breast region as this is the region of interest. Figure 3.7(a) shows the breast tissues only .Before going further, this image needs to be enhanced in order to get better output in the mammographic density/mass detection process. Experiments have been done with many image enhancement techniques and it has been found

that c ontrast lim ited a daptive h istogram e qualization (CLAHE) serves d esired p urpose v ery well. Let t he i mage af ter C LAHE operation b e d enoted as I_{CLAHE} .CLAHE d ivides th e e ntire picture into many small tiles and performs histogram equalization on each of the tiles .After that all t he e nhanced t iles a re j oined t ogether by bi linear i nterpolation t o e xclude f alsely i nduced boundary l ines .This p rocess suppresses t he noi se a nd t he or iginal f oreground o f t he pi cture becomes enhanced .While performing the CLAHE operation few parameters have been tuned so that th e mammographic de nsity/mass en hance m ore t han t he n ormal t issue .As t he mammographic de nsity/mass i s b righter t han normal b reast tissue, the h istogram sh ape distribution parameter has been set to exponential so that brighter sections of the mammogram get m ore en hanced .The out put i mage da ta r ange pa rameter ha s be en s et t o or iginal t o a void overspreading. F ig. 3.9(b) is the e nhanced image a fter CLAHE ope ration of the image in Fig. 3.9(a).

(a)   (b)
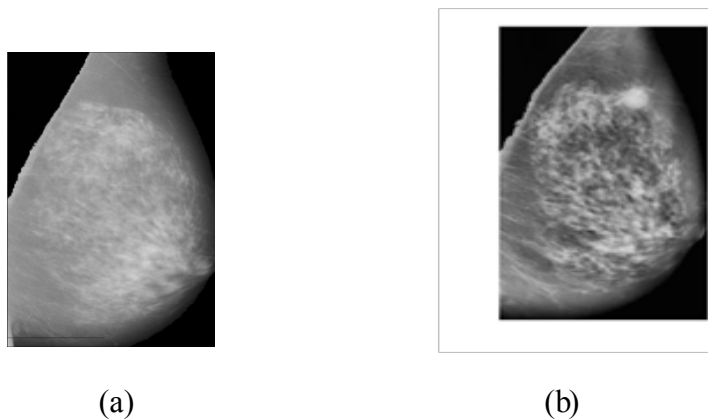
Fig. 3.9: (a) Pectoral Muscle Segmented Image, (b) Image after CLAHE operation.

3.3.2 Suspected Mammographic Density Area Detection in ROI

A careful examination of the mammograms, after removing all noises, reveals that there are four categories of objects in one mammogram:

The background: This section has very low pixel values. In section 3.1.1, the breast area has been fully detected leaving only the background which can be set to complete black by putting pixel values equal to zero for whole background

Pectoral muscle region: A novel and efficient method of successfully detection and removing of this part has been described in the section 3.2.

Breast tissue region: This is the region that contains the normal breast tissue including fat and blood vassals. The textural features and intensity are different in different parts of this region though they have some patterns in common for which this can be considered as a single region for the simplicity of image processing. Breast tissue region is easily separable from background by comparing pixel value.

Mammographic density: If there is any mammographic density/mass in the concerned mammogram, textural and intensity properties of this area will be different compared to the breast tissue region. But the distinction between normal tissue and mammographic density/mass in terms of textural and intensity properties is not sharp, hence special classification algorithm has to be used to identify the Mammographic density/mass.

From the above discussion, it is clear that to detect mass area, the classification algorithm has to find three distinct area in the mammogram as pectoral muscle region has already been detected and removed and these three areas are:

➢ The background

➢ Breast tissue region

➢ Mammographic density

The smallest area that encircles a Mammographic density/mass in all of the mammograms in MIAS database was found to be (16×16) pixels. So the whole mammogram has been divided in small areas of (16×16) pixels for each of which gray-level co-occurrence matrix (GLCM) features has been calculated and fed to the classifier to determine its class i.e. background, breast tissue region and mass area.

The gray-level co-occurrence matrix (GLCM) is a method of calculating statistical information which examines textural features of an image considering spatial relationship of pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix.

The GLCM is created by calculating the frequency of occurrence of any intensity value i with respect to specific spatial relationship to another intensity value j. In this case, the spatial relationship of the pixel of interest and the second pixel to its immediate above (vertically second) has been taken to calculate GLCM though other spatial relationship can also be considered. Each element (i,j) in GLCM specifies the number of times that the pixel with value i occurred horizontally adjacent to a pixel with value j. The number of gray levels in the image determines the size of the GLCM.

To illustrate, the following figure shows how the first three values in a GLCM of the 4-by-5 image I are calculated. Element (1,1) in the GLCM contains the value 1 because there is only one instance in the image where two, horizontally adjacent pixels have the values 1 and 1. Element (1,2) in the GLCM contains the value 2 because there are two instances in the image where two, horizontally adjacent pixels have the values 1 and 2.

GLCM	1	2	3	4	5	6	7	8
1	1	2	0	0	1	0	0	0
2	0	0	1	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	1	2	0
6	0	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

I:

1	1	5	6	8
2	3	5	7	1
4	5	7	1	2
8	5	1	2	5

Fig. 3.10: GLCM calculation demonstration of a 4 X 5 image.

But before the statistical information could be calculated, the GLCM is to be make symmetrical. Another step is still required for texture calculation which is to make the symmetrical GLCM normalized. The normalizing process is done by the following equation:

$$\mathbf{P}_{(i,j)} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}} \tag{3.9}$$

Some statistical information have been calculated from the symmetrical normalized GLCM, i.e. Contrast, Correlation, Energy and Homogeneity.

Contrast: Measures the local variations in the gray-level co-occurrence matrix. This is a measurement of the intensity contrast between a pixel and its neighbor over the whole image. The formulae for calculating Contrast is:

$$\sum_{i,j=0}^{N-1} P_{i,j} (i-j)^2 \tag{3.10}$$

Correlation: Returns a measure of how correlated a pixel is to its neighbor over the whole image.

$$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j} \tag{3.11}$$

Homogeneity: Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Formulae for Homogeneity calculation:

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{3.12}$$

Energy: Energy uses each $P_{(I,j)}$ as a weight for itself. High value of Energy occurs when the window is very orderly. Formulae:

$$\sum_{i,j} p(i,j)$$

Based on GLCM and its variants, pectoral muscle segmented image is divided into three different areas using K-mean clustering. The area with highest pixel value is detected as the desired area of mammographic density.

Fig. 3.11 shows the detected classified area after K-mean clustering.

Area 1 (Background)   Area 2 (Suspected mass area)   Area 3 (breast tissue)
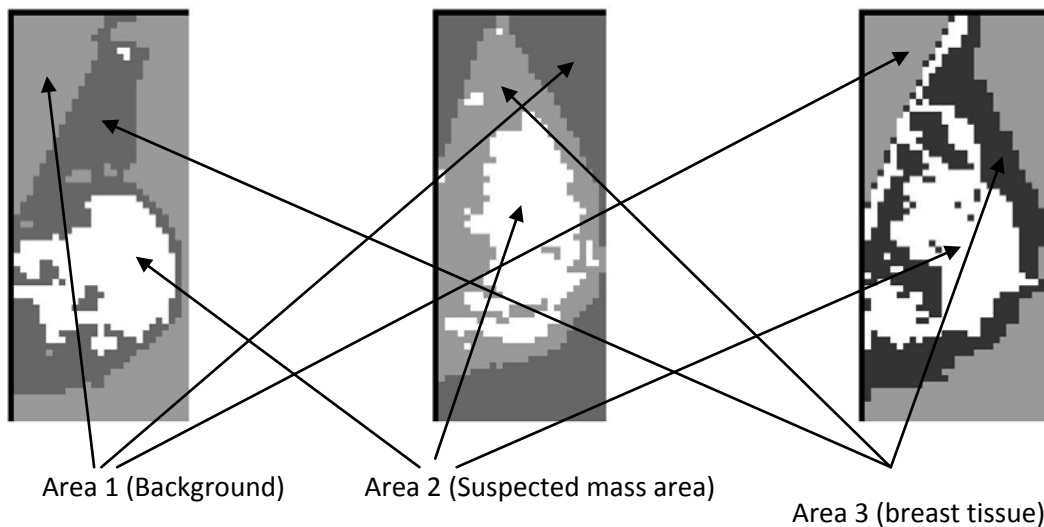
Fig. 3.11: Image after pixel Value Equalization.

In order to make the automated mass detection algorithm presented in this Thesis more accurate and robust, classification is done in two steps. The first step only determines the area which has a high potential of being mass, which is called here *suspected mass area*, thus narrowing the target space for more detail classification described in the following section. This two step classification gives the liberty of ignoring the exact class membership in the first classification process and K-mean clustering has been used as *unsupervised learning* mode as the first classification process.

The goal of *k*-means clustering is to partition the data into *k* groups such that the within-group sum-of-squares is minimized. k means treats each observation in data as an object having a location in space. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. kmeans uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters. This algorithm moves objects between clusters until the sum can not be decreased further. The result is a set of clusters that are as compact and well-separated as possible.

3.3 Feature Selection

Mammographic densities are a particular challenge to detect for radiologists and CAD systems . The fact is that there is no predefined size, shape or contrast of breast mammographic density . Mammographic densities of breast that are found near breast skin line, the chest wall, and generally those appearing in dense breasts are very difficult to find out .In order to overcome this critical situation, all the mammograms in MIAS database have been examined in search of a common characteristic of mammographic densities but it did not result in a fruitful finding rather it has been found that mammographic densities share some common features with remaining portion of the ROI in the mammogram with respect to textural and intensity with subtle differences which cannot be differentiated using threshold value. But these features bear the clue to mammographic density detection.

The main difficulty of this work is to distinguish breast area and Mammographic density. The distinguishing features of these two areas are both textural and intensity based. The features that will be used for the classifier must encompass both of these parameters. Thus we proposed to use both intensity and area based features as described here.

A. *Intensity based Features*

Gray-Level Co-Occurrence Matrix (GLCM) and its variants of the mammographic density area already computed previously have been selected as intensity features for the classifier.

B. *Area based Features*

Following *Area Based Features* were selected as input features to the classifier algorithms:

i) Area: The actual number of pixels in the region.

ii) Eccentricity: This value specifies the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1. (0 and 1 are degenerate cases; an ellipse whose eccentricity is 0 is actually a circle, while an ellipse whose eccentricity is 1 is a line segment.)

iii) Diameter of equal area circle: Scalar that specifies the diameter of a circle with the same area as the region. Computed as : $\sqrt{(4 \times Area)/\pi}$

iv) Euler number: Scalar that specifies the number of objects in the region minus the number of holes in those objects.

v) Filled Area: Scalar specifying the number of on pixels in *Filled Image*.

vi) Major Axis Length: Scalar specifying the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

vii) Minor Axis Length: Scalar; the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

viii) Orientation: Scalar; the angle (in degrees ranging from -90 to 90 degrees) between the *x*-axis and the major axis of the ellipse that has the same second-moments as the region.

ix) Perimeter: Scalar; the distance around the boundary of the region. This parameter was computed by measuring the perimeter by calculating the distance between each adjoining pair of pixels around the border of the region.

x) Solidity: Scalar specifying the proportion of the pixels in the convex hull that are also in the region. Computed as Area/Convex Area; convex area specifies the number of pixels in 'Convex Image' which is a binary image (logical) that specifies the convex hull or smallest convex polygon that can contain the region, with all pixels within the hull filled in (i.e., set to on).

Here t he o utput d ata of i ntensity b ased an d ar ea b ased f eatures a re us ed a s i nput t o t he classification al gorithm an d cl assification i s d one f or ar ea-based a nd i ntensity-based f eatures individually and then for combination of all the features.

3.4 Classification

Classification i s done using t wo different t ypes of algorithm: 2 -layer N eural Network a nd Support Vector Machine. Here the output data of intensity based and area based features are used as input to the classification algorithm and classification is done for area-based and intensity-based features individually and then for combination of all the features.

a) Support Vector Machine

A S upport V ector M achine (SVM) pe rforms c lassification by c onstructing a n *N*-dimensional hyperplane that optimally separates the data into two categories. A SVM model using a sigmoid kernel function is equivalent to a two-layer neural network.

In of SVM literature, a predictor variable is called an attribute, and a transformed attribute that is used t o d efine t he h yperplane i s c alled a feature. T he t ask of c hoosing t he m ost s uitable representation i s know n a s feature sel ection. A set o f f eatures that describes o ne i s cal led a vector. So the goal of SVM modeling is to find the optimal hyperplane that separates clusters of vector in such a w ay that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the support vectors. The Fig. 3.12 presents an overview of the SVM process.
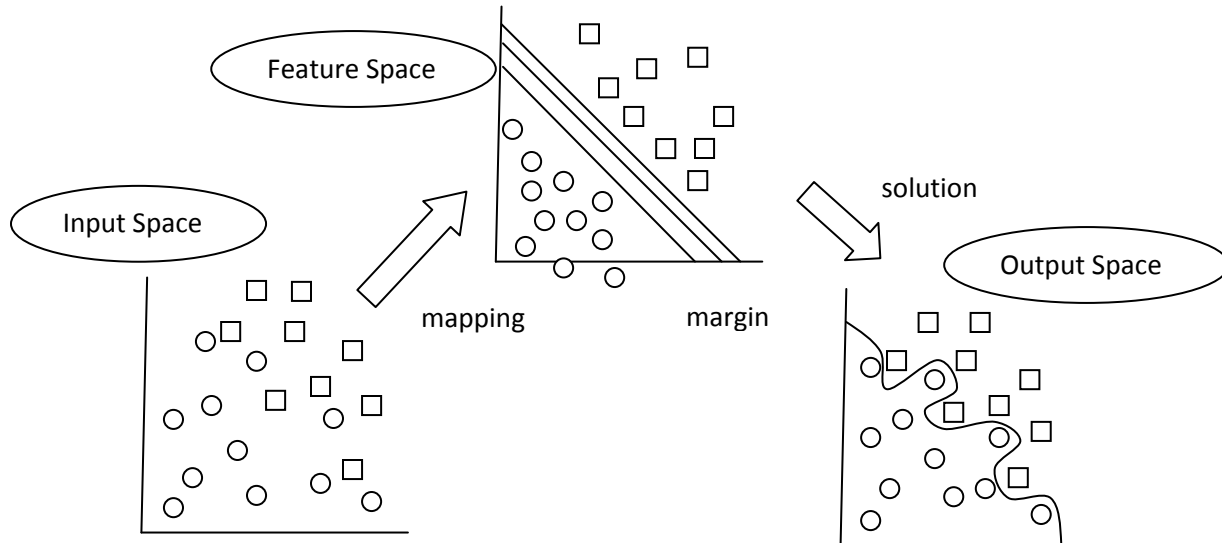
Fig. 3.12: An example of SVM.

The accuracy of an SVM model is largely dependent on the selection of the model parameters. Two methods for finding optimal parameter values, a grid search and a pattern search. A grid search tries values of each parameter across the specified search range using geometric steps. A pattern search starts at the center of the search range and makes trial steps in each direction for each parameter. If the fit of the model improves, the search center moves to the new point and the process is repeated. If no improvement is found, the step size is reduced and the search is tried again. The pattern search stops when the search step size is reduced to a specified tolerance.

Grid searches are computationally expensive because the model must be evaluated at many points within the grid for each parameter. For example, if a grid search is used with 10 search intervals and an RBF kernel function is used with two parameters (C and Gamma), then the model must be evaluated at 10*10 = 100 grid points. An Epsilon-SVR analysis has three parameters (C, Gamma and P) so a grid search with 10 intervals would require 10*10*10 = 1000

model evaluations. If cross-validation is used for each model evaluation, the number of actual SVM calculations would be further multiplied by the number of cross-validation folds (typically 4 to 10). For large models, this approach may be computationally infeasible.

A pattern search generally requires far fewer evaluations of the model than a grid search. Beginning at the geometric center of the search range, a pattern search makes trial steps with positive and negative step values for each parameter. If a step is found that improves the model, the center of the search is moved to that point. If no step improves the model, the step size is reduced and the process is repeated. The search terminates when the step size is reduced to a specified tolerance. The weakness of a pattern search is that it may find a local rather than global optimal point for the parameters.

In case of combination of two methods, the grid search is performed first. Once the grid search finishes, a pattern search is performed over a narrow search range surrounding the best point found by the grid search. Hopefully, the grid search will find a region near the global optimum point and the pattern search will then find the global optimum by starting in the right region.

The idea of using a hyperplane to separate the feature vectors into two groups works well when there are only two target categories. But SVM handles this existing case where the target variable has more than two categories. Several approaches have been suggested, but two are the most popular: (1) "one against many" where each category is split out and all of the other categories are merged; and, (2) "one against one" where $k(k-1)/2$ models are constructed where k is the number of categories. Here "one against many" approach is used to classify the detected suspected areas into 3 different classes: benign, malignant and normal.

b) 2-layer Neural Network

A 2 -layer f eed f orward N eural N etwork i s u sed t o cl assify t he su spected m ass ar eas i n 3 different classes. To calculate the weights for minimum error, back propagation process is used. Neural n etworks a re co mposed o f si mple el ements o perating i n p arallel. T hese el ements ar e inspired by biological nervous systems. As in nature, the connections between elements largely determine the network function. A neural network can be trained to perform a particular function by adjusting the values of the connections (weights) between elements.

Typically, neural networks are adjusted, or trained, so that a particular input leads to a sp ecific target output. The next figure illustrates such a situation. There, the network is adjusted, based on a c omparison of t he out put a nd t he t arget, unt il t he n etwork out put m atches t he t arget. Typically, many such input/target pairs are needed to train a network.

A feed forward neural network begins with an input layer. The input layer may be connected to a hidden layer or directly to the output layer. In case of its connection to a hidden layer, the hidden layer can then be connected to another hidden layer or directly to the output layer. There can be any num ber of hi dden l ayers, as l ong a s t here i s a t least one hi dden l ayer or o utput l ayer provided. Commonly most neural networks will have one hidden layer, and it is very rare for a neural network to have more than two hidden layers.

Fig. 3.13 illustrates a typical feed forward neural network with a single hidden layer.
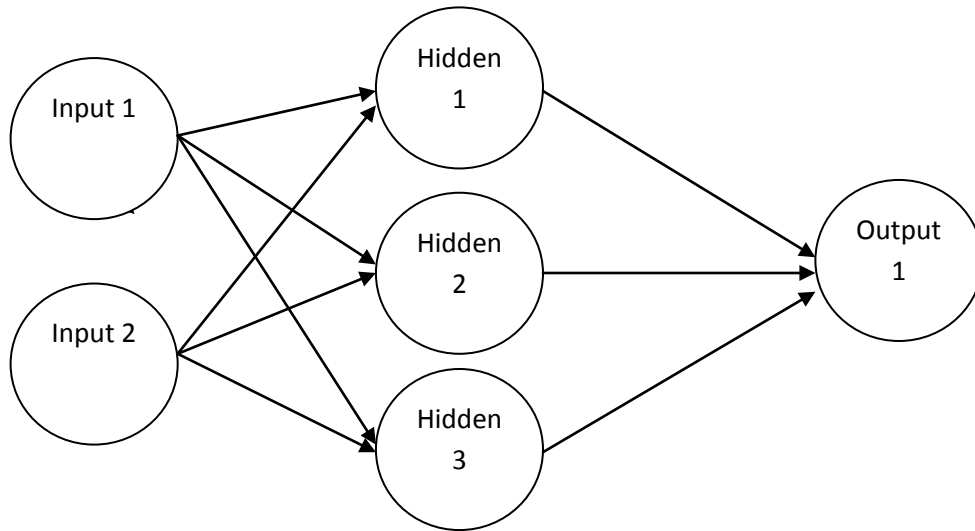
Fig. 3.13: An example of Feed Forward NN.

Equation 3.13 describes how the output of a single neuron can be calculated.

$$\textbf{Output} = \sum_{i=0}^{n-1} x_i w_i + w_n \qquad (3.13)$$

The above equation takes input values named **x**, and multiplies them by the weight **w.** Here, the

last value in the weight matrix is the threshold. This threshold is w_n .

To perform the above operation with matrix mathematics, the input is used to populate a matrix

and a row is added, the elements of which are all ones. This value will be multiplied against the

threshold value. For example, if the input were 1, 2, and 3, the following input matrix would be

produced.

$$[1\ 2\ 3\ 1]$$

The dot pr oduct w ould t hen be t aken be tween t he i nput matrix a nd t he w eight matrix. T his

number would then be fed to the activation function to produce the output from the neuron.

The va lues c ontained in t he w eight a nd t hreshold m atrix w ere de termined us ing t he ba ck

propagation algorithm. This is a very useful algorithm for training neural networks. In the back

propagation a lgorithm, t he ne ural network i s presented w ith t raining da ta. A s e ach i tem of

training data is presented to the neural network, the error is calculated between the actual output of the neural network and the output that was expected. The weights and threshold are then modified, so there is a greater chance of the network returning the correct result when the network is next presented with the same input.

Back propagation is a very common method for training multilayered feed forward networks. Back propagation can be used with any feed forward network that uses an activation function that is differentiable. This derivative function is used during training. For one of the common activation functions, the activation function derivative is available from a chart.

To train the neural network, a method must be determined to calculate the error. As the neural network is trained, the network is presented with samples from the training set. The result obtained from the neural network is then compared with the anticipated result that is part of the training set. The degree to which the output from the neural network differs from this anticipated output is the error.

To train the neural network, the important point is to minimize this error. To minimize the error, the neuron connection weights and thresholds must be modified. A function must be defined at this stage to calculate the rate of error of the neural network. This error function must be mathematically differentiable. Because the network uses a differentiable activation function, the activations of the output neurons can be thought of as differentiable functions of the input, weights, and thresholds. If the error function is also a differentiable function, such as the sum of square error function, the error function itself is a differentiable function of these weights. This leads to evaluate the derivative of the error using the weights. Then, using these derivatives, weights and thresholds are found in order to minimize the error function.

There are several ways to find weights that will minimize the error function. The most popular approach is to use the gradient descent method. The algorithm that evaluates the derivative of the error function is known as back propagation, because it propagates the errors backward through the network.

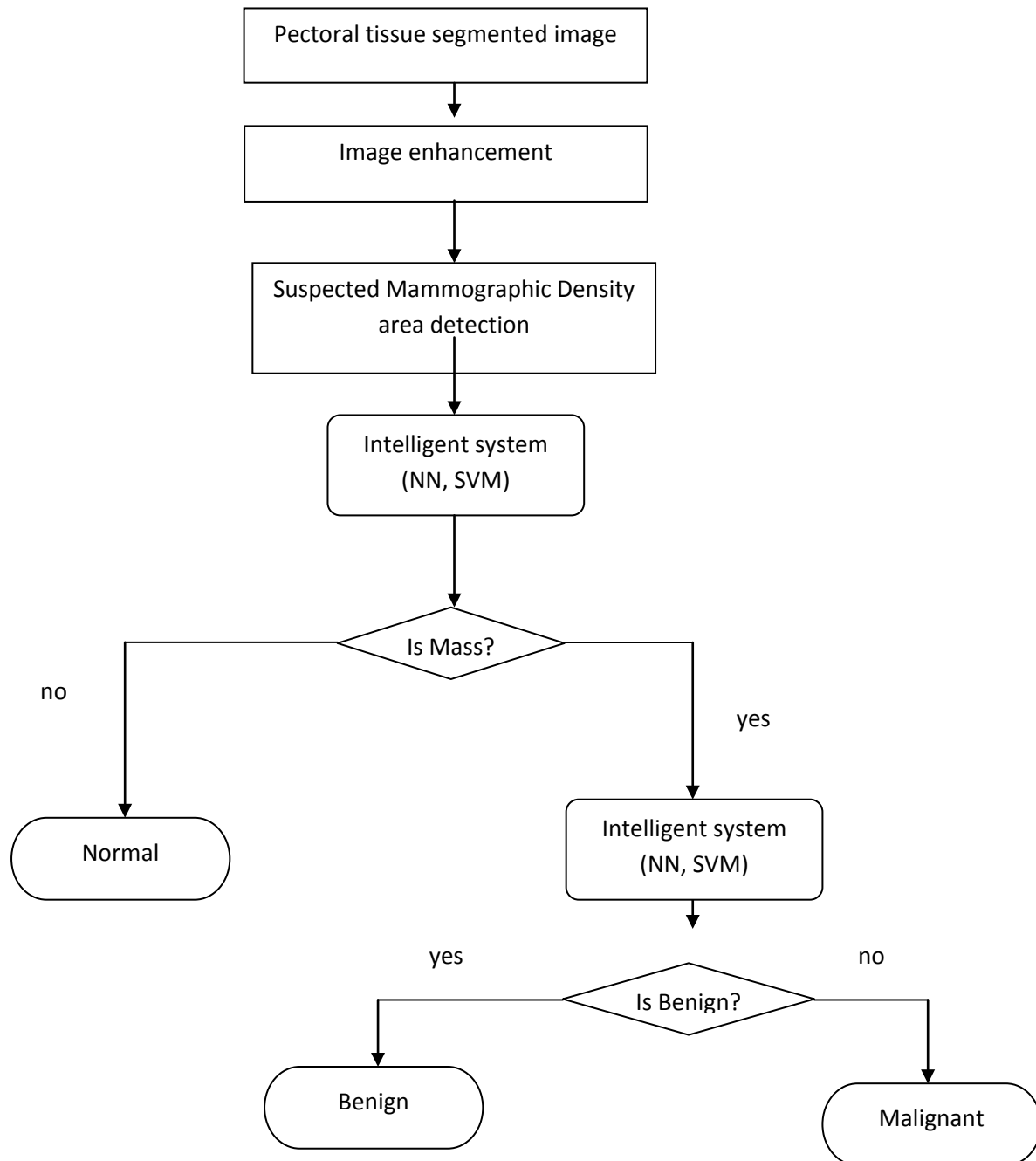An algorithm given in Fig. 3.14 has been developed to accomplish this task.

Fig. 3.14: Flow-chart for Detection and Classification of Masses.

Chapter 4

SIMULATION RESULTS AND DISCUSSION

4.1 Simulation Conditions

MATLAB is used for simulating the proposed method. It is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows formulating solutions to many technical computing problems, especially those involving matrix representations, in a fraction of the time. On the other hand it would require writing a program in a scalar non-interactive language like C/C++.

The Image Processing Toolbox software is a collection of functions that extend the capability of the MATLAB numeric computing environment. The toolbox supports a wide range of image processing operations, including

• Spatial image transformations

• Morphological operations

• Neighborhood and block operations

• Linear filtering and filter design

• Transforms

• Image analysis and enhancement

• Image registration

• Deblurring

• Region of interest operations

Many of the toolbox functions are MATLAB files with a series of MATLAB statements that implement specialized image processing algorithms. You can view the MATLAB code for these functions using the statement

```
type function_name
```

One can extend the capabilities of the toolbox by writing your own files, or by using the toolbox in combination with other toolboxes, such as the Signal Processing Toolbox™ software and the Wavelet Toolbox software.

4.2 Database

The Mammographic Image Analysis Society (MIAS) is an organisation of UK research groups interested in the understanding of mammograms and has generated a database of digital mammograms. Films taken from the UK National Breast Screening Programme have been digitised to 50 micron pixel edge with a Joyce-Loebl scanning microdensitometer, a device linear in the optical density range 0-3.2 and representing each pixel with an 8-bit word. The database contains 322 digitised films and is available on 2.3GB 8mm (ExaByte) tape. It also includes radiologist's "truth"-markings on the locations of any abnormalities that may be present. The database has been reduced to a 200 micron pixel edge and padded/clipped so that all the images are 1024x1024. Mammographic images are available via the Pilot European Image Processing Archive (PEIPA) at the University of Essex. The algorithm for pectoral muscle segmentation and mass detection presented here is applied on these 322 mammographic images of MIAS mini database [73].

By popular request, the original MIAS Database (digitised at 50 micron pixel edge) has been reduced to 200 micron pixel edge and clipped/padded so that every image is 1024 × 1024 pixels.

The follow list gives the films in the MIAS database and provides appropriate details as follows (http://peipa.essex.ac.uk/info/mias.html):

1st column:
MIAS database reference number.

2nd column:
Character of background tissue:
 F Fatty
 G Fatty-glandular
 D Dense-glandular

 3rd column:
Class of abnormality present:
 CALC Calcification
 CIRC Well-defined/circumscribed masses
 SPIC Spiculated masses
 MISC Other, ill-defined masses
 ARCH Architectural distortion
 ASYM Asymmetry
 NORM Normal

4th column:
Severity of abnormality;
 B Benign
 M Malignant

5th, 6th columns:
x,y image-coordinates of centre of abnormality.

7th column:
Approximate radius (in pixels) of a circle enclosing the abnormality.

There are also several things should be note:

> The list is arranged in pairs of films, where each pair represents the left (even filename numbers) and right mammograms (odd filename numbers) of a single patient.

> The size of *all* the images is 1024 pixels x 1024 pixels. The images have been centered in the matrix.

> When calcifications are present, centre locations and radii apply to clusters rather than individual calcifications. Coordinate system origin is the bottom-left corner.

> In some cases calcifications are widely distributed throughout the image rather than concentrated at a single site. In these cases centre locations and radii are inappropriate and have been omitted.

4.3 Protocol for Subjective Evaluation

Two expert radiologists were invited to evaluate the performance of this pectoral muscle segmentation methodology. First they were informed about the purpose and objective of this methodology. Five point assessment scales, used by Kwok et al., 2004, was used as evaluation scale. Table 4.1 shows the scale with score description. The segmentation results were displayed on a computer screen. And radiologists were asked to determine whether the output was acceptable or not for each image. After that the output images were displayed again, this time the radiologists marked them according to the five point assessment score and at the same time outputs were binary classified, i.e., true positive (TP), true negative (TN), false positive (FT) and false negative (FN).

Table 4.1: Protocol for Evaluation.

Score	Meaning	Description
1	Exact	The segmented line delineates the pectoral margin exactly. Any deviations from the visually perceived margin are imperceptible or insignificant.
2	Optimal	The segmented line delineates the pectoral margin exactly for at least half its length and adequately for the other half.
3	Adequate	The segmented line delineates the pectoral margin inexactly but with sufficient accuracy for localizing the pectoral margin for the intended purpose.
4	Sub-optimal	The segmented line delineates the pectoral margin inadequately for at least half its length.
5	Inadequate	The segmented line is either absent or is inaccurate as to be inadequate for localizing the pectoral margin for the intended purpose.

4.4 Results on Subjective Evaluation of Pectoral Muscle Segmentation

Radiologist 1 marked 265 images as true positive, 4 images as true negative, 25 images as false positive and 28 images as false negative. Radiologist 2 marked 295 images as true positive, 7 images as true negative, 8 images as false positive and 20 images as false negative. Fig. 4.1 describes these assessments. Table 4.2 shows the confusion matrix of the binary classification classified by the two radiologists. Sensitivity or true positive rate (TPR) was determined to be

0.9 by radiologist 1 and 0.96 by radiologist 2. And false positive rate (FPR) was determined to be 0.86 by radiologist 1 and .53 by radiologist 2.
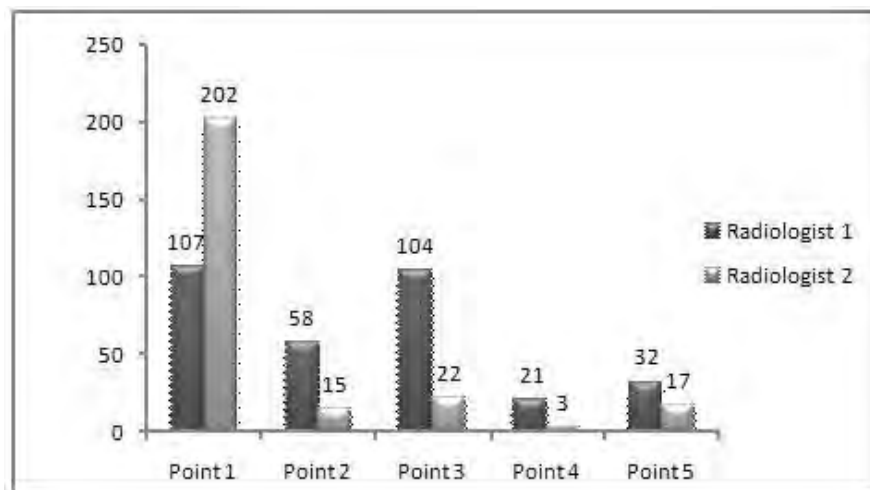

Fig. 4.1: Chart containing the assessment by 2 radiologists.

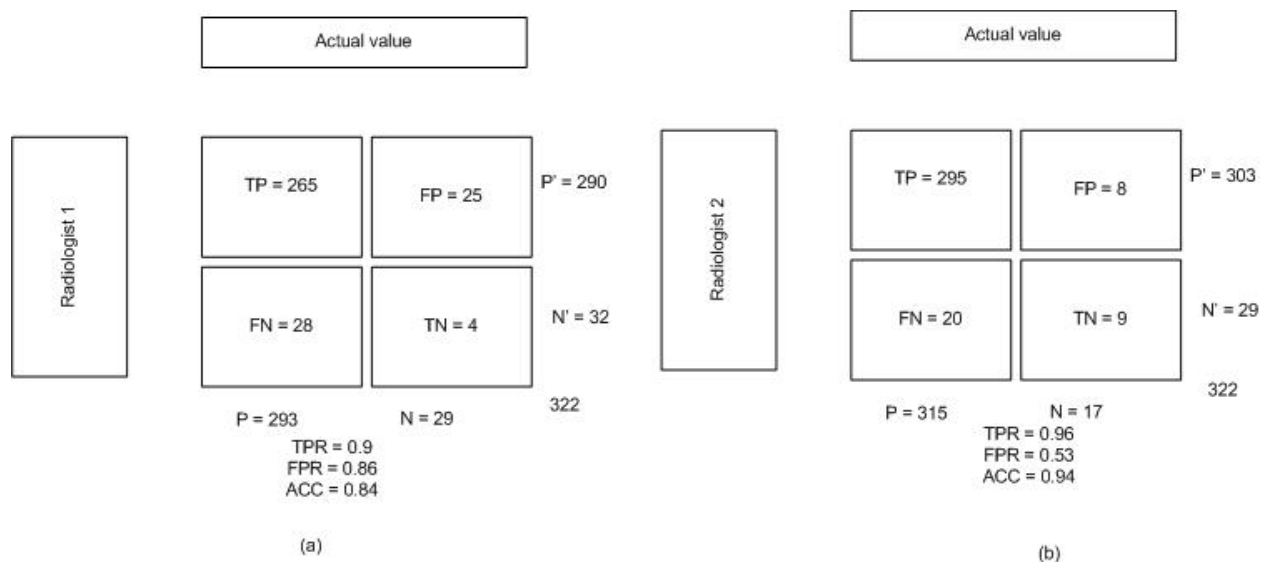Table 4.2: Confusion Matrix.

Fig. 4.2 describes the percentage of acceptance of the assessed results by 2 expert radiologists.

Accuracy of the algorithm was deemed to be 84% by radiologist 1 and 94% by radiologist 2.
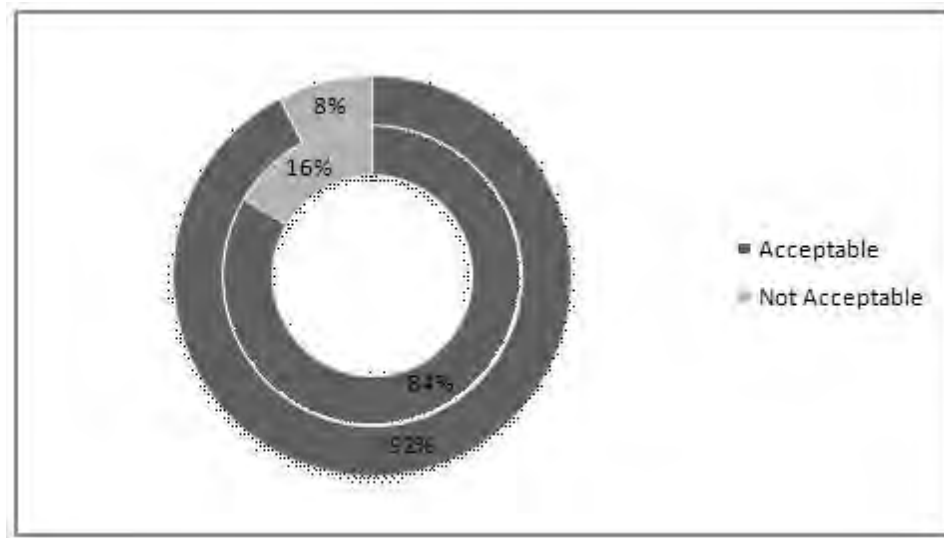
Fig. 4.2: Percentage of Acceptance of the results.

4.5 Performance Analysis on Pectoral Muscle Segmentation

The novel method is capable of segmenting pectoral muscle of a broad range of size, shape and position. Fig. 4.3 shows some of the outputs. The algorithm successfully delineates small pectoral tissue as in Fig. 4.3(a), 4.3(c) and 4.3(j) as well as large pectoral tissue as in Fig. 4.3(b) and 4.3(e). In Fig. 4.3(d), 4.3(f) and 4.3(l), the pectoral tissue in the lower part is merged with the fibro-glandular tissue, making it difficult for segmentation. But the algorithm is also effective in this scenario. In Fig. 4.3(g), there are several layers within the pectoral tissue region. In this case, the correct layer was selected by the proposed methodology accurately. In Fig. 4.3(n), the pectoral muscle is not easily distinguishable from the other breast tissues. The method achieved a high degree of accuracy in approximating the pectoral muscle in this figure. In Fig. 4.3(o), the method again proved its effectiveness by exactly delineating the pectoral tissue while there is a sticky tag in the pectoral tissue. In Fig. 4.3(m), there is no pectoral tissue and so is the output.
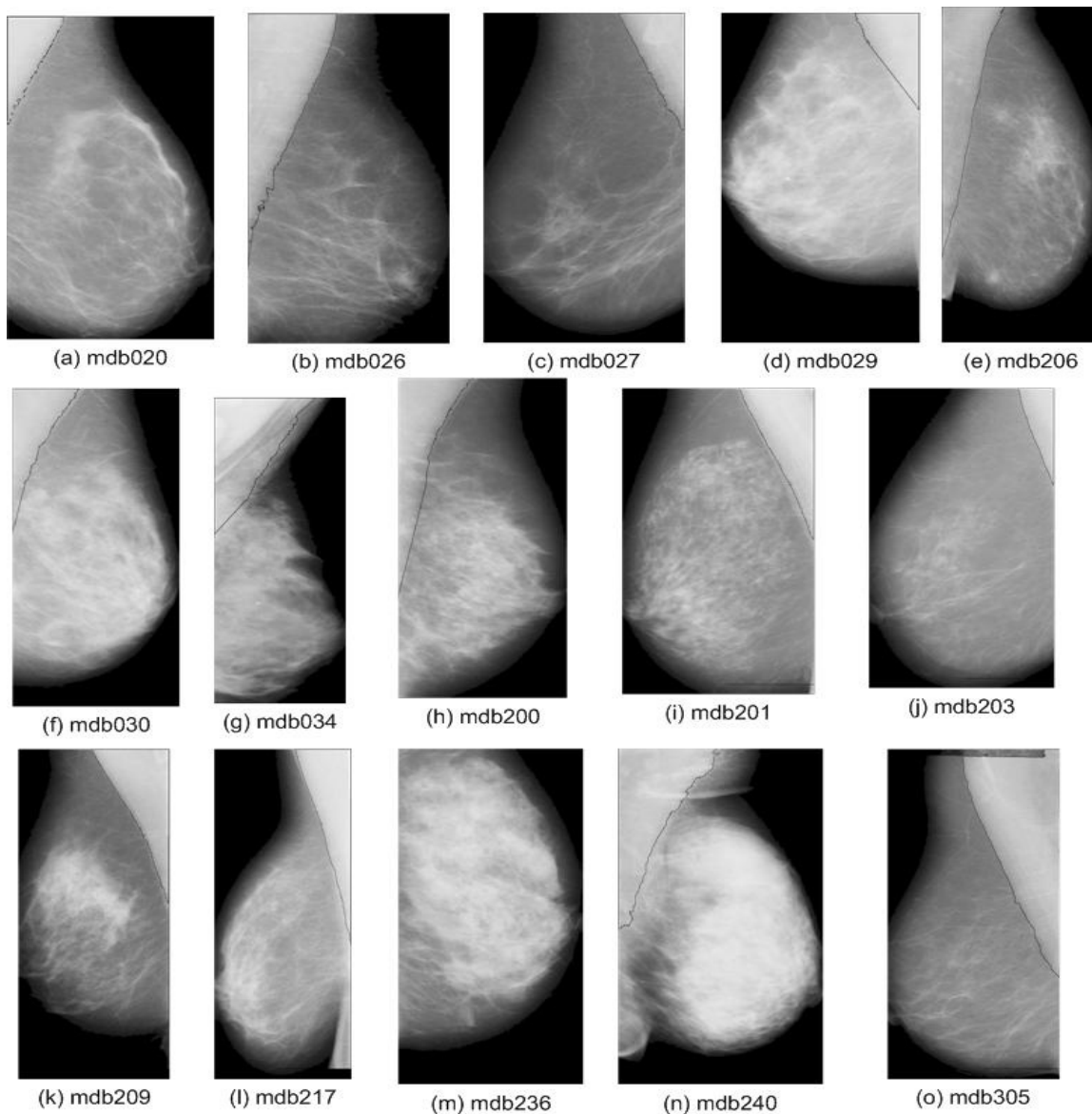
Fig. 4.3: Some obtained outputs.

A R eceiver o perating c haracteristic (ROC) cu rve w as d rawn u sing t he T PR an d FPR v alues determined by the radiologists. The ROC curve is depicted in Fig. 4.4. Both the points obtained from radiologist 1 & 2 are above the "line of no-discrimination" which points to the fact that the new methodology yi elded a goo d r esult. F urthermore, t o t est the qua lity o f t he bi nary

classification by the two radiologists, Matthews's correlation coefficient (MCC) was determined. It was done by the following formula:

$$MCC= \frac{TP \; X \; TN - FP \; X \; FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

For the classification by radiologist 1, MCC was found to be 0.04 and for the classification by radiologist 2, MCC was found to be 0.38. Both of the values of MCC are above zero, which means that the methodology segments the pectoral muscle with a greater degree of good prediction.

Fig. 4.4: Receiver Operating Characteristic (ROC) curve of the assessment by 2 radiologists.

In addition to the method's ability to segment out a wide range of size, shape and position of pectoral tissue, the method offers a higher statistical confidence as described in result section. The ROC in Fig. 4.4 and t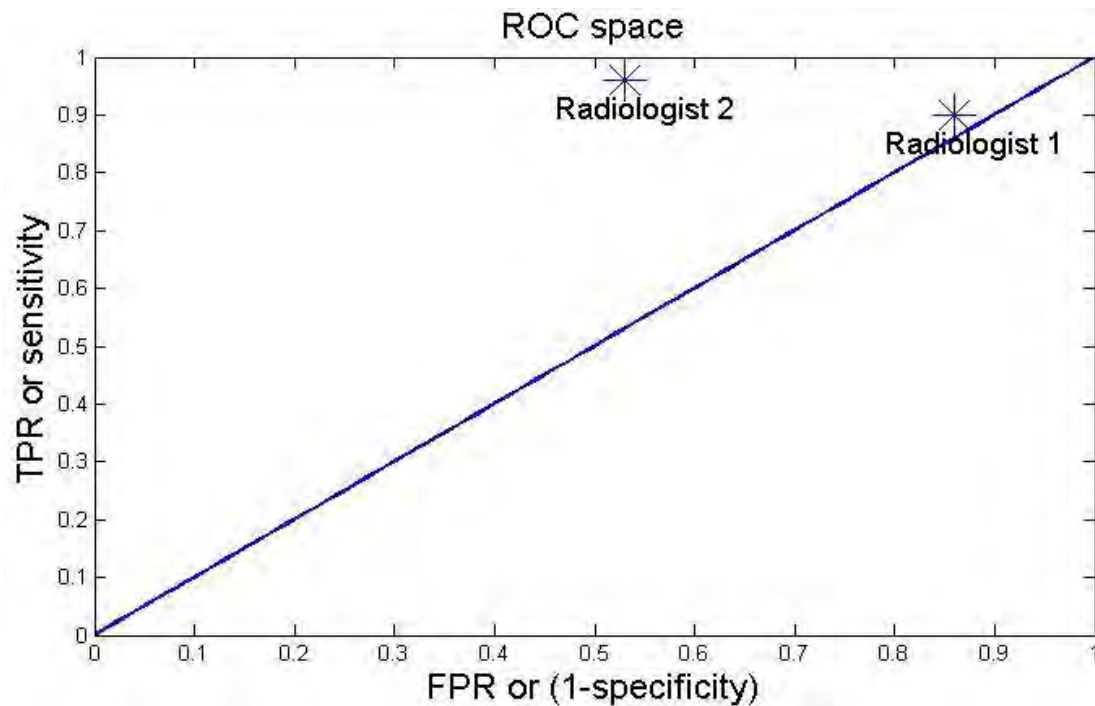he value of MCC clearly indicate that, the result obtained with this method was not happened by chance. In [73], the analysis part has been described up to these statistical methods.

Now, Table 4.3 depicts the comparison study of accuracy rate of the proposed method with existing methods. Though the position of the proposed method is in 2nd position, there are some points to be cleared. The 1st positioned method has some drawbacks of assumptions used in [33]. Compared to existing methods, the proposed method is quite good to do the specified task.

Table 4.3: Comparison Study of Proposed Method with Existing Methods.

Method	Percentage of Accuracy
Proposed Method	**88%**
Wavelet Decomposition [29]	85%
Curve fitting Method [36]	83.9%

4.6 Results on Objective Evaluation of Suspected Area Detection

Kmean clustering was used to segregate the area having high potential of being mass. To get an idea of how well the clustering was performed silhouette plot is presented in Fig. 4.5 for the clustering of image mdb120.
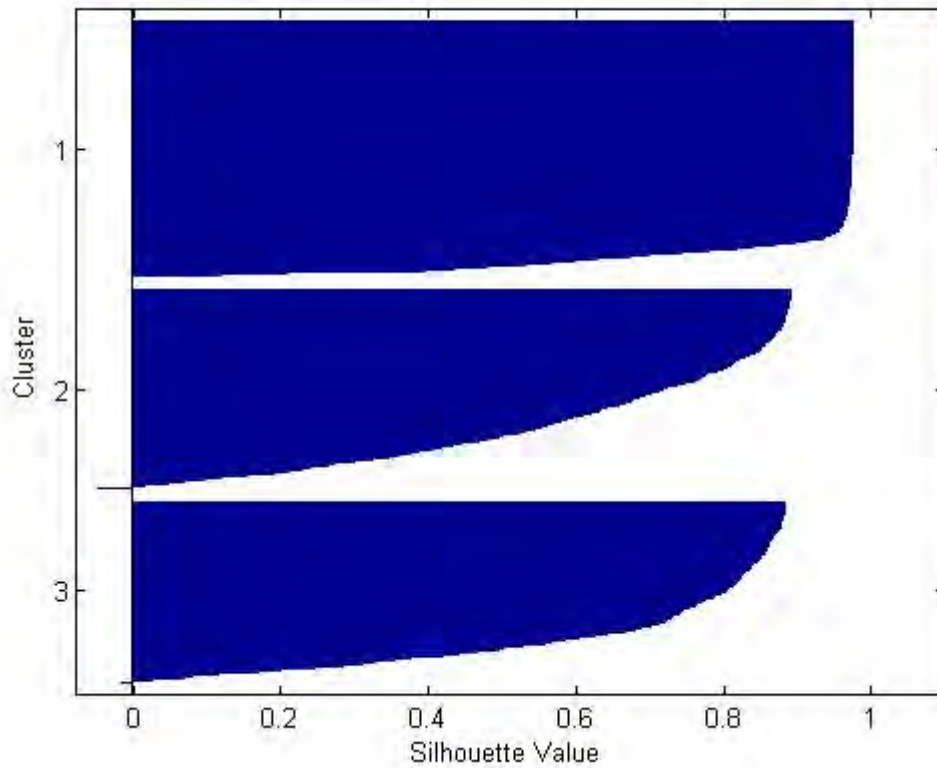
Fig. 4.5: Clustering of Image mdb120.

4.7 Performance Analysis on Suspected Area Detection

In this part, suspected area is detected by clustering the image using 16-by-16 blocks and taking GLCM of the blocks.

Silhouette plot provides a graphical representation of how close each point in one cluster is to points in the neighboring clusters. The range is from -1 to +1. A value of +1 indicates that the point is very distant from neighboring clusters, through 0, indicating the point lies in the middle of neighboring clusters, to -1 which indicates the point is wrongly clustered. From the simulation it was found that the clustering was perfectly done as there are very few points having negative value whereas most of the points lie in the positive side with high positive value

indicating that those points are well separated. However there are some points having low values

indicating that they are nearby to neighboring clusters.
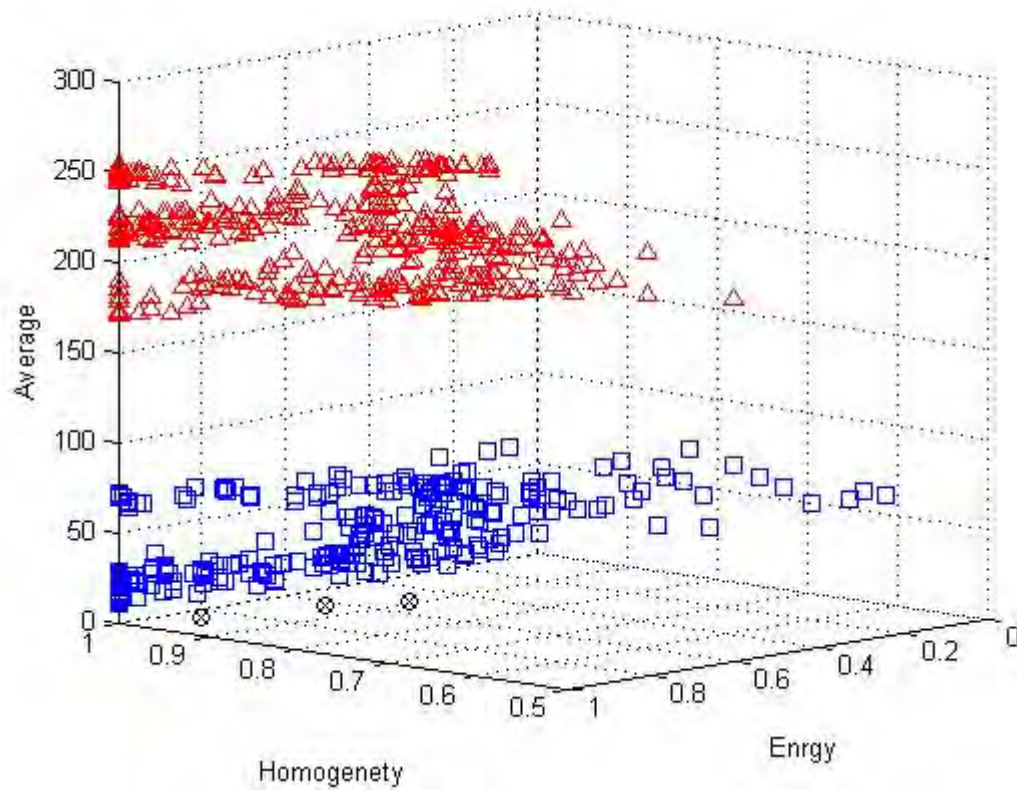
Fig. 4.6: 3D visualization of K-means clustering.

The 3 -D pl ot gi ves a good pr esentation of ho w w ell t he points are s eparated a nd e ffects of

different features in clustering.

4.8 Results on Objective Evaluation of Feature Extraction

One set of 17 features have been extracted from the detected suspected area of the images. Here

two types of features are used to classify the suspected area with greater accuracy. One type is of

area-based features and another type is of intensity-based features.

4.9 Performance Analysis on Feature Extraction

Features of mammograms have been extracted to classify them in 3 different classes. Here, area-based and intensity-based features are used to do further procedure of the total method. Area-based features are extracted depending on the shape, size and region of the suspected area. On the other hand, intensity-based features are extracted from the pixel value of the suspected area. These two types of extracted features are then used to classify the suspected area in 3 different classes- benign, malignant or normal. To manipulate with only significant features, different features are extracted and classification is done depending on them. The main effect of these features can be decided statistically only when the classification part is accomplished. So to discuss the accuracy of extracted features further explanation should be on the classification part. This concludes that the accuracy of classification implies indirectly the accuracy of detected suspected area and extracted features.

4.10 Results on Objective Evaluation of Classification

Classification of detected suspected area was done by two different methods:

  a) Support Vector machine
  b) 2-layer Neural Network

The outputs of these two methods are described below:

  a) Support Vector Machine:

Classification is done depending on area-based and intensity-based features. Table 4.4 below shows classification of detected suspecting areas depending on two types of features individually and also on both type of features.

Table 4.4: Classification by Support Vector Machine.

Feature set	(% of accuracy)			Sensitivity	Specificity
	Benign	**Malignant**	**Normal**		
Area Based	0.8385	0.8199	0.7950	0.9063	**0.5758**
Intensity Based	0.7323	0.7199	0.8696	**0.6667**	0.9901
Area Based+ Intensity Based	0.8571	0.8261	0.8882	**0.9766**	**0.9406**

b) 2-stage Neural Network

By using 2-stage Neural Network, classification of suspected areas was done very efficiently. Like earlier, to classify the suspected areas using this method two types of feature set have been used. The output of classification by Neural Network can be described by the help of following confusion matrices.

Confusion Matrices

Table 4.5(a) shows the confusion matrix for classification depending on area-based feature set. Here, training is done by 194 data, testing is done by 64 data; validation is done by rest 64 data. From the Table, overall TP, FP, TN and FN values are 199, 13, 3 and 107 respectively. These values lead to overall accuracy of 95%.

Table 4.5(a): Confusion Matrix for area-based feature set (2-stage NN).

		Target			
		Normal	Benign	Malignant	
Output	Normal'	199	4	2	205
	Benign'	3	60	5	71
	Malignant'	0	2	47	49
		202	66	54	322

Normal=98.51%

Benign=90.91%

Malignant=87.03%

ACC=95.03%

Table 4.5(b) shows the confusion matrix for classification depending on intensity-based feature set. Here, training is done by 194 data, testing is done by 64 data; validation is done by rest 64 data From the Table, overall TP, FP, FN and TN values are 199, 19, 3 and 101 respectively. These values lead to overall accuracy of 93.2%. Here FP value (19) is a bit higher than that for area-based feature set which leads to lower TN value (101).

Table 4.5(b): Confusion Matrix for intensity-based feature set (2-stage NN).

		Target			
		Normal	Benign	Malignant	
Output	Normal'	199	5	4	208
	Benign'	3	59	8	70
	Malignant'	0	2	42	44
		202	66	54	322

Normal=98.51%

Benign=89.39%

Malignant=77.77%

ACC=93.2%

Table 4.4(c) s hows t he c onfusion m atrix f or c lassification de pending on c ombination of bot h types of feature set. Like earlier examples, training is done by 194 da ta, testing is done by 64 data; validation is done by rest 64 data From the Table, overall TP, FP, FN and TN values are 197, 15, 5 a nd 105 r espectively. These va lues lead t o ove rall accuracy of 93.8%. Here overall

efficiency is a bit higher than that for intensity-based feature set, but a bit lower than that for area-based feature set.

Table 4.5(c): Confusion Matrix for area-based+ intensity-based feature set (2-stage NN).

		Target			
		Normal	Benign	Malignant	
Output	Normal'	197	3	3	203
	Benign'	3	62	8	73
	Malignant'	2	1	43	46
		202	66	54	322

Normal=97.52%

Benign=93.94%

Malignant=79.63%

ACC=93.79%

4.11 Performance Analysis on Classification

(a) Support Vector Machine

Specificity and sensitivity are very important statistical performance measurement parameter of a binary classification. **Sensitivity** (also called **recall rate** in some fields) measures the proportion of actual positives which are correctly identified while **Specificity** measures the proportion of negatives which are correctly identified. For area based feature set, specificity of Support Vector Machine method is only 0.5758 whereas for intensity based feature set sensitivity is only 0.6667. This indicates that area based feature set produces a handsome result of sensitivity (0.9063) but a poor value of specificity (0.5758). On the other hand, intensity based feature set leads to a poor value of sensitivity (0.6667) and a better value of specificity (0.9901). Finally by using combination of both types of features set, optimization of sensitivity and specificity can be obtained. This conclusion can be given from obtained sensitivity and specificity values which are 0.9766 and 0.9406 respectively, for combination of both types of features.

The average accuracy of the methodology was 85.71% for benign masses, 82.61% for malignant masses and 88.82% for normal condition. From Table 5.2, it is clear that for area-based feature set classification of Normal condition as unaffected one is done with lower accuracy (79.50%), while for intensity-based feature set classification of benign and malignant masses as affected ones is done with lower accuracy (73.23% and 71.99% respectively)than that by area-based feature set. And combination of two types of feature set results in optimized accuracy for normal condition as unaffected one (88.82%) and benign/ malignant masses as affected ones (85.71% and 82.61% respectively).

(b) 2-layer Neural Network
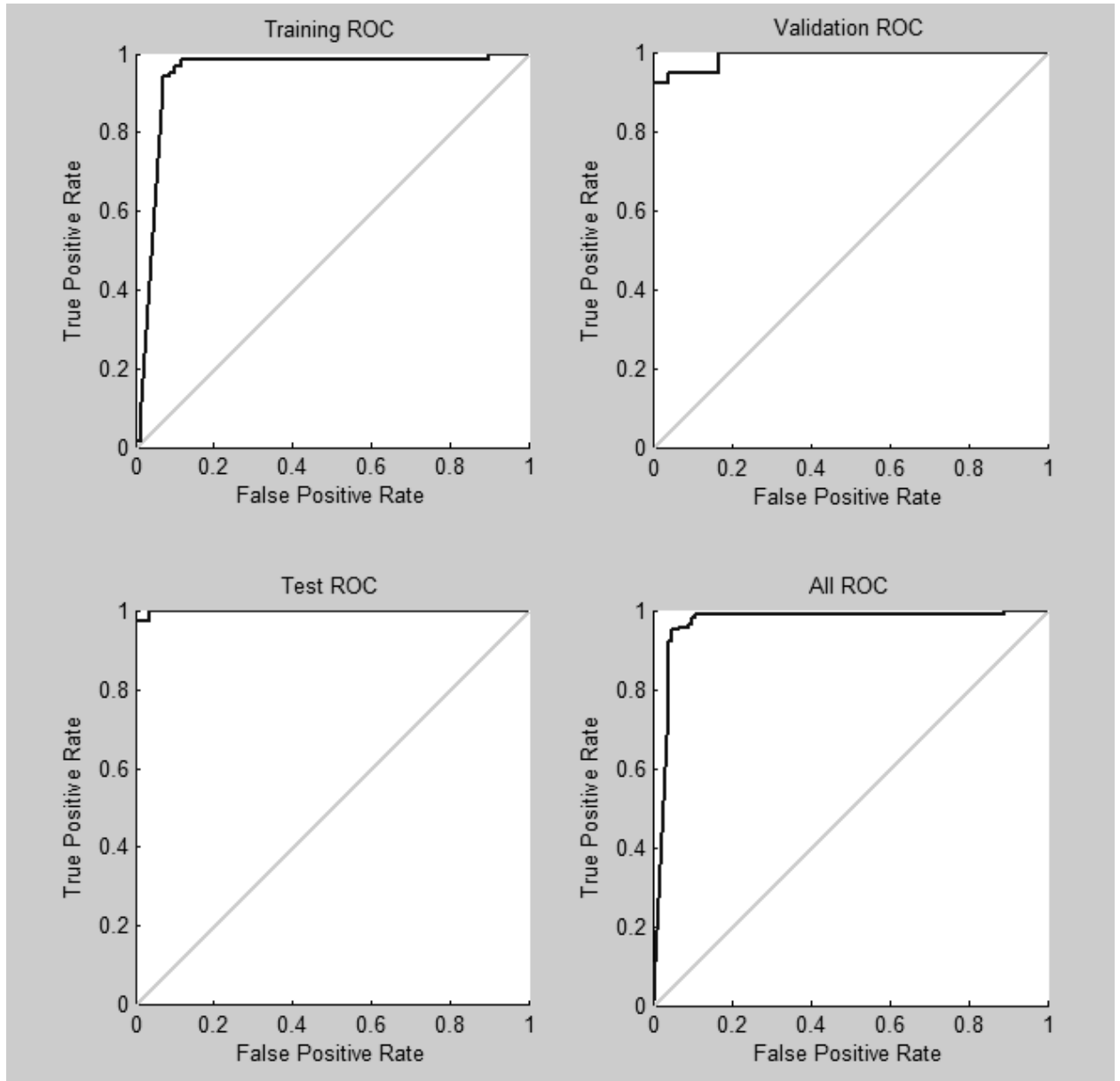
Fig. 4.7(a): Receiver Operating Characteristic (ROC) curve for area-based feature set (2-layer NN classification).
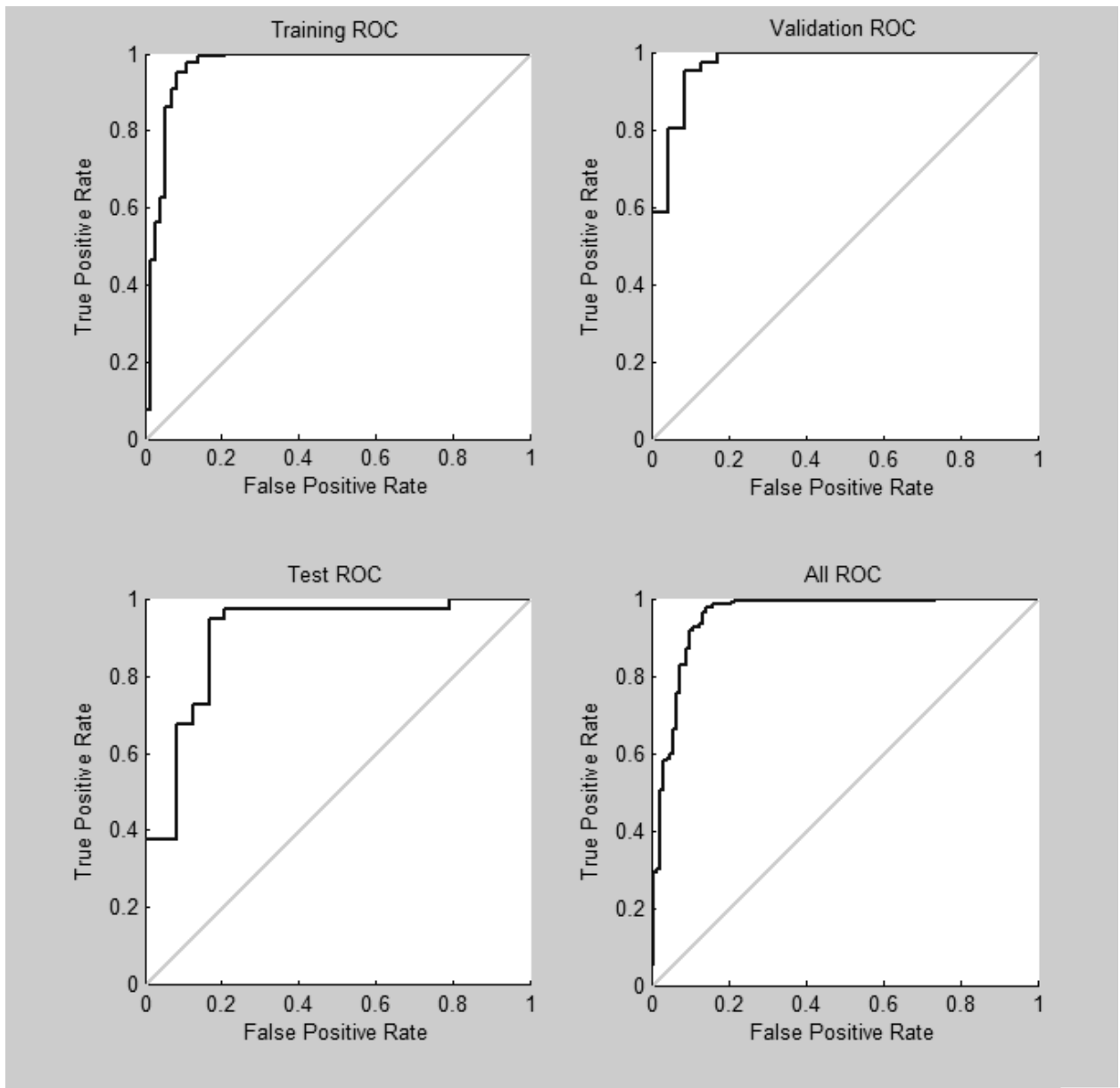
Fig. 4.7(b): Receiver Operating Characteristic (ROC) curve for intensity-based feature set (2-layer NN classification).
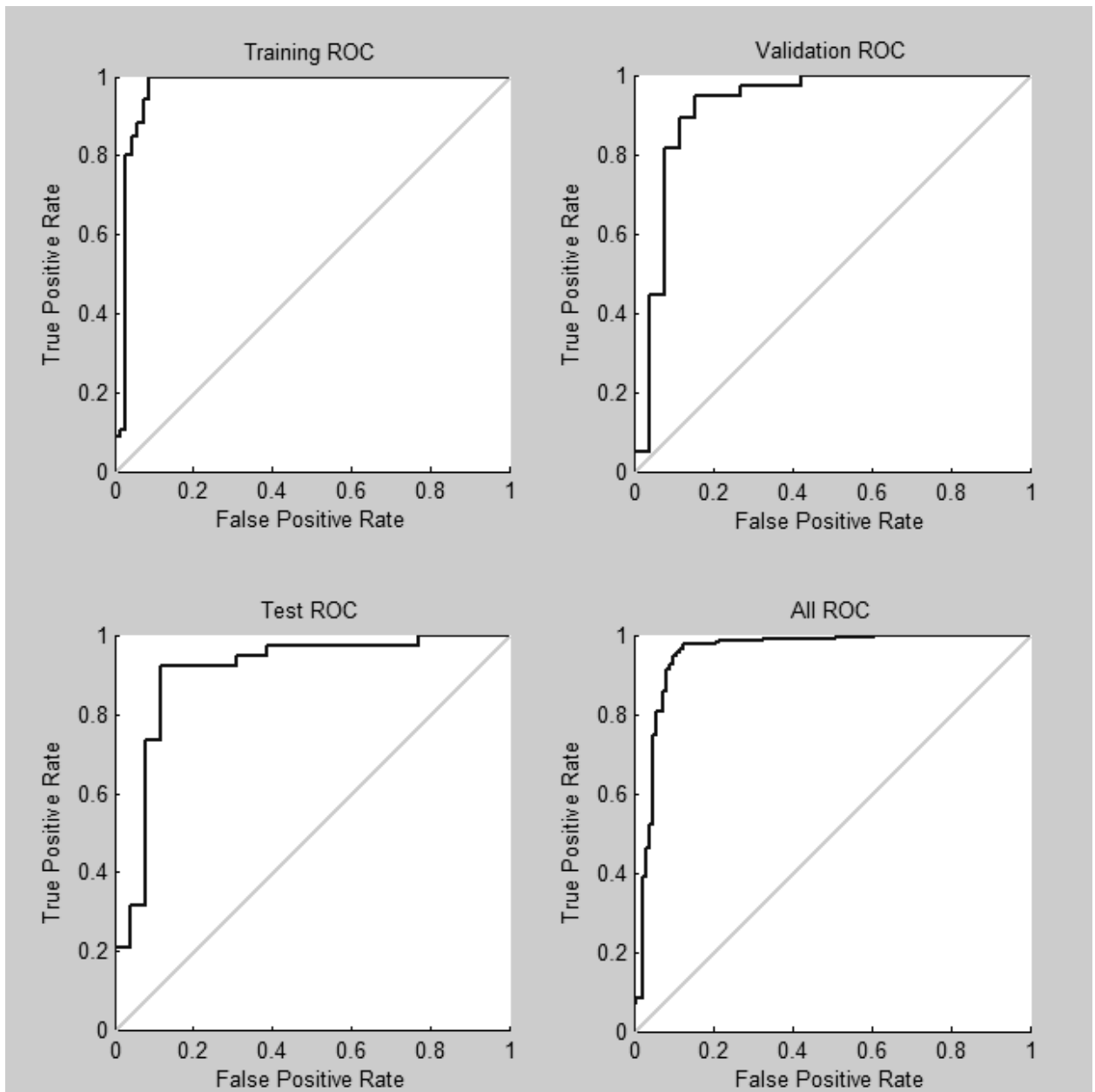
Fig. 4.7(c): Receiver Operating Characteristic (ROC) curve for area-based+ intensity-based feature set (2-layer NN classification).

Observing Fig. 4.7(a)-(c), it is clear that classification by 2-layer NN gives better result for area-based feature set than that by intensity-based or combination of both sets. To evaluate the quality

of this method, Matthews's correlation coefficient (MCC) was also calculated. MCC value for this method is found 0.89, 0.85 and 0.87 for area-based, intensity-based and combinations of both feature sets respectively. All the three values are greater than zero, which depicts perfect prediction level of classification.

By analyzing the performance of these two algorithms it is clear that it has high accuracy as well as precision for 2-layer NN classification depending on area-based feature set. But before classification, suspected areas need to be segmented from the mammographic image using GLCM, which is consisting of intensity-based features. Above two points lead to the fact that to classify the mammograms more accurately intensity-based calculation is needed to detect the suspected area which can be used to calculate area-based features. So, it can be concluded that to get optimized result area-based and intensity-based both the feature sets carry unavoidable importance in designing the proposed method.

Chapter 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this report, an overall picture of my present work is presented. All the background study that has been completed to have a clear concept about the topic has been described here. Best effort has be en t aken t o go t hrough r elated pr evious w orks t o d ecide a bout t he m ost a ccurate a nd efficient method. M ost of t he time of f irst three m onths ha s be en s pent t o ga ther background knowledge related to t his pr esent work. A ll t he e xisting methods a nd a pproaches ha ve be en explored to compare the strong as well as the weak points of those methods.

And t hen, a ccording to t he plan d iscussed ea rlier t he m ain p art o f t his w ork h as b een st arted. Simulation ha s be en w orked out a nd t he i mplementation of pe ctoral muscle s egmentation ha s been completed. After completion of this part, the obtained results of proposed method have also been a nalyzed a nd a ssessed by 2 e xpert r adiologists. T hen t he out puts ha ve be en a nalyzed statistically to calculate the ability of the method in segmenting the pectoral muscle accurately. Up to th is l evel o f this p resent w ork, tw o p apers related t o th e im plemented p art h ave b een prepared and submitted to two different conferences. They have been successfully accepted and published.

Next, the core part of this present work has been started. The mammographic density has been detected f rom t he seg mented r egion of i nterest of t he mammograms. F eatures ha ve be en extracted f rom su spected m ammographic d ensity a rea. Classification p art h as b een d one

depending on the extracted features using two different artificial intelligence methods: Support Vector Machine and Neural Network. The algorithms have been used on all 322 mammograms of mini MIAS database and results have also been analyzed to measure the accuracy level.

In summary, it can be claimed that this work is very important in breast cancer detection process. This early detection of breast cancer will play an important role to decrease the mortality rate related to breast cancer. With the help of this method, the radiologists can predict the disease in its early stage and also they can save their time. In addition to this, false diagnosis due to human fatigue can also be reduced to a great extent.

5.2 Future Work

This is the final report of this present work. Pre-processing of the mammograms, pectoral muscle segmentation, mammographic density detection, feature extraction and classification have already been done successfully. Then the detected abnormalities have been classified as benign or malignant or normal using two methods, Neural Network and Support Vector Machine. The outputs of these two methods have been compared with each other.

As future work, improvisation of the methods and increasing the percentage of accuracy to classify masses can be considered. Other artificial intelligence methods (like Adaptive Neuro Fuzzy Intelligent System (ANFIS) method) can be used to do the classification part of present work and best methodology can be pointed out for this specific job. As it is the core job of the present work, it will make it happen to detect mammographic density in early stage with greater accuracy.

Another important point is about micro-calcification detection. This part can also be added in further extension of the method. This part will make the method able to do the task of detection of breast cancer with greater accuracy and consequently, development in medical world will reach to a new era.

In this report, a method for breast cancer detection is discussed. At the end a continuous process has been prepared to detect mammographic density and predict breast cancer in its early stage. This early detection of breast cancer will play an important role to decrease the mortality rate related to breast cancer and serve the nations all over the world.

Bibliography

[1] http://www.banglapedia.org/httpdocs/HT/C_0033.HTM

[2]Bankman, I.N. (eds.). Handbook of Medical Imaging. San Diego, USA:Academic Press. pp. 323-325. 2000.

[3] D. Roder, N. Houssami, G. Farshid, G. Gill, P. Downey, K. Beckmann, P. Iosifidis, L. Grieve, and L. Williamson, "Population screening and intensity of screening are associated with reduced breast cancer mortality: evidence of efficacy of mammography screening in Australia," *Breast Cancer Research and Treatment*, vol. 108, no. 3, pp. 409-416, 2008.

[4] D. Miglioretti, R. Smith-Bindman, L. Abraham, J. Brenner, E. Aiello, M. Buist, P. Carney, and J. Elmore, "Radiologist characteristics associated with interpretive performance of diagnostic mammography," *Journal of the National Cancer Institute*, vol. 99, no. 24, pp. 1854-63, 2007.

[5] R. Brem, "Clinical versus research approach to breast cancer detection with CAD: where are we now?," *American Journal of Roentology*, vol. 188, pp. 234-235, 2007.

[6] J. Dean and V. Ilvento, "Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers." *American Journal of Roentology*, vol. 187, pp. 20-28, 2006.

[7] M. Morton, D. Whaley, K. Brandt, and M. Amrami, "Screening mammograms: interpretation with computer-aided detection - prospective evaluation," *Radiology*, vol. 239, pp. 204-212, 2006.

[8] R. Birdwell, P. Bandodkar, and D. Ikeda, "Computer-aided detection with screening mammography in university hospital settings," *Radiology*, vol. 236, pp. 451-457, 2005.

[9] H. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in Mammograms: a survey," *Pattern Recognition*, vol. 36, pp. 2967-2991, 2003.

[10] S. Halkiotis, T. Botsis, and M. Rangoussi, "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks," *Signal Processing*, vol. 87, pp. 1559-1568, 2007.

[11] X.H. Wang, B. Zheng, W.F. Good, W.F. Chang, and Y.H. Chang, "Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network," *International Journal Medical Informatics*, vol. 54, pp. 115-126, 1999.

[12] http://www.breastcancer.org/pictures/breast_anatomy/image_1.jsp

[13](Merck Manual Online, Breast Cancer).

[14] J.N. Wolfe, "Breast patterns as an index of risk of developing breast cancer", *American Journal of Roentgenology,* vol. 126, pp. 1130-1139, 1976.

[15] J.N. Wolfe, J.N. (1976). "Risk for breast cancer development determined by mammographic parenchymal pattern," *Cancer,* vol. 37, pp. 2486-2492, 1976.

[16] B. Verma and J. Zakos,(2001). " A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques," *IEEE Transaction on Information Technology in Biomedicine,* vol. 5, no. 1, pp. 46-54, 2001.

[17] R. Chandrasekhar and Y. Attikiouzel, Y. (2001). "Automatic breast border segmentation by background modeling and subtraction", *5th International Workshop on Digital Mammography,* 2001, pp.560-565.

[18] R. Chandrasekhar and Y. Attikiouzel, (1997). "A simple method for automatically locating the nipple on mammograms," *IEEE Transactions on Medical Imaging,* vol. 16, pp. 483-494, 1997.

[19] D. Raba, A. Oliver, J. Marti, M. Peracaula, and J. Espunya (2005). "Breast segmentation with pectoral muscle suppression on digital mammograms," *IbPRIA*, 2005, pp. 471-478.

[20] U. Bick and M. Giger , "Automated Segmentation of Digitized Mammograms," *Academic Radiology,* vol. 2, pp. 1-9, 1995.

[21] R.J. Ferrari and R.M. Rangayyan, "Identification of the breast boundary in mammograms using active contour models," *Medical and Biological Engineering and Computing,* vol. 42, pp. 201-208, 2004.

[22] H. Mirzaalian, M.R. Ahmadzadeh, and F. Kolahdoozan,(2006). "Breast contour detection on digital mammograms," *Information and Communication Technologies,* vol.1, pp. 1804-1808, 2006.

[23] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transaction Pattern Analysis Machine Intelligence,* vol. 15, pp. 850-863, 1993.

[24] B.K.M. Shahedi, R. Amirfattahi, F.T. Azar, and S. Sadri, "Accurate breast region detection in digital mammograms using a local adaptive thresholding method," *8th International Conference on Image Analysis for Multimedia Interactive Services,* 2007, pp. 26-26.

[25] M. Mustra, J. Bozek, and M. Grgic, " Nipple detection in craniocaudal digital mammograms," *51st International Symposium*, 2009, pp. 15-18.

[26] C. Zhou, H.P. Chan, C. Paramagul, M.A. Roubidoux, B. Sahiner, L.M. Hadjiiski, and M. Petrick, " Computerized nipple identification for multiple image analysis in computer-aided diagnosis," *Medical Physics,* vol. 31, no. 10, pp. 2871-2882, 2004.

[27] M. Karnan and K. Thangavel, "Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of

microcalcifications," *Computer Methods and Programs in Biomedicine,* vol. 87, no. 1, pp. 12-20, 2007.

[28] S .K. Kinoshita, P.M. A zevedo-Marques, R .R. P ereira J r, J .A.H. R odrigues, a nd R .M. Rangayyan (2008). " Radon-domain de tection of t he ni pple a nd the pe ctoral muscle i n mammograms", *Journal of Digital Imaging,* vol. 21, no. 1, pp. 37-49, 2001.

[29] M. Mustra, J. Bozek, and M. Grgic, "Breast border extraction and pectoral muscle detection using wavelet decomposition", *EUROCON ,*2009 , pp. 1426-1433.

[30] M . M asek, R . C handrasekhar, C .J.S. D esilva, an d Y . Attikiouzel, "S patially b ased application of the minimum cross-entropy thresholding algorithm to segment the pectoral muscle in m ammogram," *Seventh Australian and New Zealand Intelligent Information Systems Conference*, 2001, pp. 101-107.

[31] X. Weidong and X. Shunrena, "A model based algorithm to segment the pectoral muscle in mammograms", *IEEE Internatonal Conference on Neural Networks & Signal Processing,* 2003, pp. 1163-1169.

[32] R.J. F errari, R.M. Ranagayyan, J.E.L. Desautels, R.A. Borges, and A.F. Frere, "Automatic identification of the pectoral muscle in mammograms," *IEEE Transactions on Medical Imaging,* vol. 23, no. 2, pp. 232-245,2004.

[33] S .M. Kwok, R . Chandrasekhar, a nd Y . A ttikiouzel, " Automatic pe ctoral m uscle segmentation on mammograms b y s traight line e stimation a nd c liff de tection," *Seventh Australian and New Zealand Intelligent Information Systems Conference,* 2001, pp. 67-72.

[34] L. Tabar and P.B. Dean (1985). *Teaching Atlas of Mammography,* 2nd revised ed. Thieme-Stratton, New York**.**

[35] R. Chandrasekhar an d Y . A ttikiouzel, "Mammogram- attribute d atabase: A to ol f or mammogram segmentation and analysis" *Proceedings of the IASTED International Conference: Signal Processing, Pattern Recognition, and Applications*, 2001, pp. 143-148.

[36] S.M. Kwok, R . C handrasekhar, Y . A ttikiouzel, a nd M .R. R ickard, " Automatic P ectoral Muscle S egmentation o n Med iolateral O blique V iew Mam mograms," *IEEE Transactions on Medical Imaging,* vol. 23, no. 9, pp. 1129-1140, 2004.

[37] P.F**.** Felzenszwalb and D.P. Huttenlocher, "Efficient graph-based image segmentation," *Int. Jour. Computer Vision,* vol. 59, no. 2, pp. 167-181, 2004.

[38] M. Bajger, F. Ma, and M.J. Bottema, .(2005). "Minimum spanning trees and active contours for i dentification of pe ctoral m uscle i n s creening mammograms," *Digital Image Computing: Techniques and Applications,* 2005, pp. 323-329.

[39] H. Mirzaalian, M.R. A hmadzadeh, a nd S. Sadri, "Pectoral m uscle segmentation on di gital mammograms by nonl inear diffusion filtering," *IEEE Conference on Communication, Computer and Signal Processing,* 2007, pp. 581-584.

[40] J . S uckling, J. P arker, D .R. Dance, S . Astely, I. Hutt, C .R.M. B oggis, I . R icketts, E . Stamakis, N. Cerneaz, S.L. Kok, P. Taylor, D. Betal, and J. Savage, "The mammographic image analysis so ciety d igital m ammogram d atabase," *Digital Mammography: Proc. of the 2nd International Workshop on Digital Mammography*, 1994, pp. 375-378.

[41] D .P. H uttenlocher, G .A. K landerman, a nd W.J. Rucklidge, "Comparing i mages us ing t he hausdorff distance," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 15, pp. 850-863, 1993.

[42] N . Karssemeijer, "A utomated cl assification o f p arenchymal p atterns in m ammograms," *Phys. Med. Biol.,* vol. 43, no. 2, pp. 365-378, 1998.

[43] F errari, R .J., R anagayyan, R .M., D esautels, J .E.L., B orges, R .A. and F rere, A.F. (2004). "Automatic i dentification of t he pectoral m uscle i n m ammograms", *IEEE Transactions on Medical Imaging, vol. 23(2),* pp. 232-245.

[43] W. Xu, L. Li, and W. Liu, (2007). "A novel pectoral muscle segmentation algorithm based on pol yline f itting a nd e lastic t hread a pproaching," *The 1st International Conference on Bioinformatics and Biomedical Engineering,* 2007, pp. 837-840.

[44] M. Lyra, S. Lyra, B. Kostakis, S. Drosos, C. Georgosopoulos, and K. Skouroliakou,(2008). "Digital m ammography t exture a nalysis by c omputer a ssisted i mage pr ocessing", *IEEE International Workshop on Imaging Systems and Techniques,* 2008, pp. 73-76.

[45] C.B.R. F erreira a nd D .L. B orges, " Automated m ammogram classification us ing a multiresolution pattern recognition approach," *XIV Brazilian Symposium on Computer Graphics and Image Processing,* 2001, pp. 76-83.

[46] R. Jahanbin, M.P. Sampat, G.S. Muralidhar, G.J. Whitman, A.C. Bovik, and M.K. Markey, "Automated R egion of i nterest de tection of s peculated m asses on di gital mammograms," *IEEE Southwest Symposium on Image Analysis and Interpretation,* 2008, pp. 129-132.

[47] A . D ong a nd B . Wang, " Feature s election a nd a nalysis on m ammogram c lassification," *IEEE Conference on Communication, Computer and Signal Processing,* 2009, pp. 731-735.

[48] B.S. Manjunath and W.Y. Ma, "Texture features for browsing and retrieval of image data", *IEEE Transactions on Pattern Analysis and Machine Intelligence,"* vol. 18, no. 8 , pp. 837-842, 1996.

[49] A .C. B ovik, M . C lark, a nd W.S. G eisler, "Multichannel t exture a nalysis us ing l ocalized spatial filters," *IEEE Transactions on Patern Analysis and Machine Intelligence,* vol. 12, no. 1 , pp.55-73, 1990.

[50] C .H. Wei, Y . L i, a nd C .T. Li, " Effective e xtraction of ga bor f eatures f or a daptive mammogram re trieval," *IEEE International Conference on Multimedia and Expo,* 2007, pp. 1503-1506.

[51] M.E. Osman, M.A. Wahed, A.S Mohamed, and Y.M. Kadah, "Computer aided diagnosis system for classification o f microcalcifications in d igital m ammograms," *26th National Radio Science Conference,* 2009, pp. 1-6.

[52] R. Mousa, Q. Munib, and A. Moussa, "Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural," *Expert systems with application*, vol. 28, pp. 713-723, 2005.

[53] N. Nicolaou, S. Petroudi, J. Georgiou, M. Polycarpou and M. Brady, " Digital mammography: towards pectoral muscle removal via independent component analysis," *4th IET International Conference on Advances in Medical, Signal and Information Processing*, 2008, pp. 1-4.

[54] I. Faye, B.B. Samir, and M.M. Eltoukhy, " Digital mammograms classification using a wavelet based feature extraction method," *Computer and Electrical Engineering,* vol. 2, pp. 318-322, 2009.

[55] T.A. Docusse, J.R. Furlani, R.P. Romano, R.C. Guido, Shi-Huang Chen, N. Marranghello, and A.S. Pereira, " Microcalcification enhancement and classification on mammograms usingt the wavelet transform," *IEEE International Joint Conference on Neural Networks,* 2008, pp. 3181- 3186.

[56] C. Juarez, M.E. Castillo, and V. Ponomaryov, 2007) . " Classification of masses in Mammography images using wavelet transform and neural networks," *The Sixth International Kharkov Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves and Workshop on Terahertz Technologies,* vol. 2, pp. 956- 958, 2007.

[57] J. Quintanilla-Dominguez, M.G. Cortina-Januchs, A. Jevtic, D. Andina, J.M. Barron-Adame, and A. Vega-Corona, " Combination of nonlinear filters and ANN for detection of microcalcifications in digitized mammography," *IEEE Conference on Systems, Man and Cybernetics,* 2009, pp. 1516-1520.

[58] D. Guliato, R.M. Rangayyan, J.D. Carvalho, and S.A. Santiago, (2008). " Polygonal modeling of contours of breast tumors with the preservation of spicules," *IEEE Transaction on Biomedical Engineering,* vol. 55, no. 1, pp. 14-20, 2008.

[59] B.W. Hong and B.S. Sohn, " Segmentation of region of interest in mammograms in a topographic approach," *IEEE Transaction on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 129-139, 2010.

[60] X. Gao, Y. Wang, X. Li, and D. Tao, 2010) . "On combining morphological component analysis and concentric morphology model for mammographic mass detection," *IEEE Transactions on Information Technology in Biomedicine,* vol. 14, no. 2, pp. 266-273, 2010.

[61] N.H. Eltonsy, G.D. Tourassi, and A.S. Elmaghraby,(2007). " A concentric morphology model for the detection of masses in mammography," *IEEE Transactions on Medical Imaging,* vol. 26, no. 6, pp. 880-889, 2007.

[62] X. Si and L. Jing, " Mass detection in digital mammograms using twin support vector machine-based CAD system," *WASE International Conference on Information Engineering,* 2009, pp. 240-243.

[63] Xu, W., Li, L. and Xu, P.(2007). "A new ANN-based detection algorithm of the masses in digital mammograms," *IEEE International Conference on Integration Technology, 2007(ICIT'07),* pp. 26-30.

[64] G. Rezai-rad and S. Jamarani, "Detecting microcalcification clusters in digital ammograms using c ombination of w avelet a nd ne ural n etworks," *International Conference on Computer Graphics, Imaging and Vision: New Trends,* 2005, pp. 197-201.

[65] B . V erma a nd J . Z akos, 2001). " A c omputer-aided d iagnosis s ystem f or di gital mammograms ba sed on f uzzy-neural an d f eature ex traction t echniques," *IEEE Transaction on Information Technology in Biomedicine,* vol. 5, no. 1, pp. 46-54, 2001.

[66] L. V ibha, G.M. H arshavardhan, K . P ranaw, P . D eepa S henoy, K.R. V enugopal, and L .M. Patnaik, "C lassification o f mammograms u sing d ecision trees," *10ᵗʰ International Database Engineering and Applications Symposium,* 2006, pp. 263-266.

[67] D .C. H ope, E . M unday, a nd S.L. S mith, "Evolutionary a lgorithms i n the c lassification of mammograms," *IEEE Symposium on Computational Intelligence in Image and Signal Processing,* 2007, pp. 258- 265.

[68] R . G upta and P .E. U drill, "T he u se o f t exture analysis to d elineate su spicious masses i n mammography", *Phys. Med. Biol.,* vol. 40, pp. 835-855, 1995.

[69] M. Yam, M. Brady, R. Highnam, C. Behrenbruch, R. English, and Y. Kita, (2001). "Three-dimensional reconstruction of microcalcification clusters from two mammographic views," *IEEE Trans. Med. Imag.,* vol. 20, pp. 479-489, 2001.

[70] F. Georgsson, "Algorithms and techniques for computer aided mammographic screening," Ph.D. dissertation, Umeå Univ., Dept. Comput. Sci., Umeå, Sweden, 2001.

[71] H. Arefi and M. Hahn, "A morphological reconstruction algorithm for separating off-terrain points from terrain points in laser scanning da ta," *ISPRS WG III/3, III/4, V/3 Workshop"Laser scanning 2005, Enschede, the Netherlands, September 12-14, 2005.*

[72] L . V incent, "Morphological g rayscale r econstruction i n i mage a nalysis: applications a nd efficient algorithms," *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 176-201, 1993.

[73] http://peipa.essex.ac.uk/ipa/pix/mias/

[74] N . S altanat, M .A. H ossain, and M .S. A lam, (2010), " Pectoral muscle seg mentation o n mediolateral oblique view mammograms," *IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications, Liverpool, UK,September 8-10,2010.*