

CANKAYA UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE
COMPUTER ENGINEERING

MASTER THESIS

AUTOMATIC DETECTION OF BREAST CANCER
IN MAMMOGRAM IMAGES

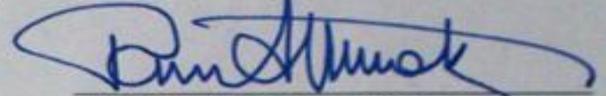
FAOZIA ALI SAIF ALSARORI

SEPTEMBER 2013

Title of Thesis: **Automatic Detection of Breast Cancer In Mammogram Images**

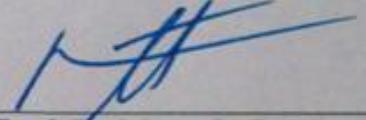
Submitted By **FAOZIA ALI SAIF ALSARORI**

Approval of the Graduate School of Natural And Applied Science, Çankaya University.



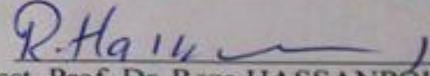
Prof. Dr. Taner ALTUNOK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Computer Engineering.



Asst. Prof. Dr. Murat SARAN
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis of the degree Master of Engineering (M.S.) in Computer Engineering.



Asst. Prof. Dr. Reza HASSANPOUR
Supervisor

Examination Date: 5 / 9 / 2013

Examining Committee Members

Asst. Prof. Dr. Reza HASSANPOUR (Cankaya University)



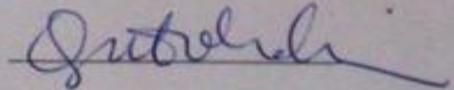
Dr. Ersin ELBAŞI

(TÜBİTAK)



Dr. Gül TOKDEMİR

(Cankaya University)

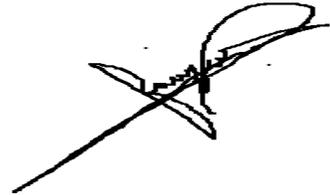


STATEMENT OF NON PLAGIARISM

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Faozia Ali S. ALSARORI

Signature :

A handwritten signature in black ink, appearing to be 'Faozia Ali S. ALSARORI', written in a cursive style.

Date : 10 / 10 / 2013

ACKNOWLEDGEMENT

No one can achieve his/her dreams alone or pass through all the difficulties while walking on the journey of life, and this project wouldn't be accomplished without support of a lot of people.

In the beginning, I would like to express my appreciation to my great supervisor, Prof. Dr. Reza HASSANPOUR who gave me unlimited supporting and valuable guidances, there is no enough words to express thanks and gratitude for you sir.

My parents, the most precious thing I have , who have given me so much, thanks for your faith on me , for teaching me that I should never surrender, all my love for you to inspire me to be successful in accomplishing all my dreams.,“I think I got one more”.

Hisham, Amna, Bushra'a, Eiman, Ezz, Rama, Osama, Raheem, Rawa, Ariam, I'm so lucky to have you lovers, thank you for everything I would not be able to make it without your love and encouragement.

Special respect to fantastic and lovely 'struggle friends', really I appreciate your continued support I won't ever forget your love and your standing beside of me in difficult and tough moments.

In addition, I would like to thank my sponsor Islamic Development Bank /scholarships office for giving me a big chance to achieve one of my dreams, thanks a lot for the unlimited and continued support during my study period, respect and appreciation for you sirs.

First and foremost, the great thank for my life brace: God.

ABSTRACT

AUTOMATIC DETECTION OF BREAST CANCER IN MAMMOGRAM IMAGES

ALSARORI, FAOZIA ALI SAIF

M.S. Department of Computer Engineering

Supervisor: Asst. Prof. Dr. Reza HASSANPOUR

September 5, 140 Pages

The aim of this thesis research is to harness the great potential of image processing techniques which have evolved significantly in the last years, to build an automatic system to detect and diagnose breast cancer in the digital mammographic images in order to help those interested people in this domain, such as radiologists and specialists in oncology and to improve their performance by reducing error rates of breast cancer diagnosis. Since segmentation and extracting the effective features of mammograms play a major role to isolate and classify suspicious regions which can be subject to cancer, in this work, we focus on abnormality detection using seed-based region growing(SBRG) and Multi-thresholding OTSU's method to segment the region of interest (ROI). Then the texture features of the segmented ROI are extracted which are used to classify the ROI as normal or abnormal tissue by using an Artificial Neural Network (ANN). This system can correctly classify the tested region by a rate of 93.80%.

Keywords: Automatic Breast Cancer Detection, Computer Aided Diagnosis, Mammogram Images, Artificial Neural Networks.

ÖZET
AUTOMATIC DETECTION OF BREAST CANCER
IN MAMMOGRAM IMAGES

ALSARORI , FAOZIA ALI SAIF

M.S. Department of Computer Engineering

Supervisor: Asst. Prof. Dr.Reza HASSANPOUR

September 5, 140 Pages

Bu tezin amacı son yıllarda belirgin ölçüde gelişen görüntü işleme tekniklerinin kullanarak, onkolojide radyolog ve uzmanlar gibi bu alanda bulunan ilgili insanlara yardımcı olmak amacıyla ve göğüs kanseri teşhis hatası oranını düşürerek performansını arttırmaya yönelik dijital mamografi görüntülerinde meme kanseri tespit ve teşhis etmek için otomatik bir sistemi oluşturmaktır.Kanser ile ilişkili olabilecek şüpheli bölgelerin izole edilmesi ve sınıflandırılmasında mamogramların vetkin özelliklerinin segmentasyonu ve ayıklanmasının önemli bir rol oynamasından dolayı bu çalışmamızda anormalliklerin teşhisinde tohum temelli bölgesel büyümeye (SBRG) odaklanarak ilgili bölgeyi (ROI) segmente etmek için OTSU yönteminin çoklu eşikleme sine odaklandık. Daha sonra da ROI'nin segmente edilen doku özellikleri ayıklanarak, ROI'nin normal ya da anormal doku olduğunu tanımlamak bir Yapay Sinir Ağı (ANN)'nı kullandık.Bu sistem test edilen bölgeyi % 93.80'lik bir oranla doğru olarak sınıflandırabilmektedir.

Anahtar sözcükler: Otomatik Göğüs Kanseri Tespiti, Bilgisayar destekli teşhis, Mamogram Görüntüleri, Yapay Sinir Ağları.

LIST OF CONTENTS

| | |
|---------------------------------------|------|
| ABSTRACT | v |
| ÖZET | vi |
| LIST OF CONTENTS | vii |
| LIST OF TABLES | xii |
| LIST OF FIGURES | xiii |
| LIST OF SYMBOLS / ABBREVIATIONS | xvi |

CHAPTER I

INTRODUCTION

| | |
|----------------------------------------|---|
| 1.1 Motivations | 3 |
| 1.2 Thesis Challenges | 4 |
| 1.3 Thesis Goal: Automatic System..... | 4 |
| 1.4 Thesis Outlines | 6 |

CHAPTER II

MEDICAL AND TECHNICAL BACKGROUND

| | |
|------------------------------|---|
| 2.1 Medical Background | 9 |
| 2.1.1 Breast anatomy | 9 |

| | |
|---------------------------------------------------------|----|
| 2.1.2 Breast cancer | 11 |
| 2.1.3 Benign and malignant tumors | 14 |
| 2.1.4 Breast cancer lesions..... | 16 |
| 2.1.4.1 Calcifications | 16 |
| 2.1.4.2 Masses | 18 |
| 2.1.5 The importance of detect breast cancer early..... | 21 |
| 2.2 Technical Background..... | 21 |
| 2.2.1 Advanced of mammography imaging..... | 21 |
| 2.2.1.1 Digital mammography (DM)..... | 22 |
| 2.2.1.2 Breast tomosynthesis | 25 |
| 2.2.2 Computer-aided detection (CAD) | 26 |

CHAPTER III

REVIEW OF IMAGE PROCESSING METHODS

| | |
|-----------------------------------------------------------------|----|
| 3.1 Image Enhancement | 28 |
| 3.1.1 Spacial domain methods | 29 |
| 3.1.1.1 Simple intensity transformations | 29 |
| 3.1.1.1.1 Contrast stretching | 29 |
| 3.1.1.1.2 Grayscale adjustments | 29 |
| 3.1.1.1.3 Image nignive | 29 |
| 3.1.1.2 Histogram processing | 30 |
| 3.1.1.2.1 Global histogram equalization(GHE) | 30 |
| 3.1.1.2.2.Adaptive histogram equalization (AHE or LHE) | 31 |
| 3.1.2 Frequency special domain methods..... | 32 |
| 3.1.2.1 Low pass filter (LPF)..... | 33 |
| 3.1.2.1.1 Mean filter | 33 |
| 3.1.2.1.2 Median filter | 33 |

| | |
|--------------------------------------------------------------|----|
| 3.1.2.2 High pass filter (HPF)..... | 34 |
| 3.1.2.2.1 Gradient operators | 34 |
| 3.2 Image Segmentation..... | 35 |
| 3.2.1 Classical techniques..... | 36 |
| 3.2.1.1 OTSU method | 36 |
| 3.2.1.1.1 Thresholding and multithresholding by OTSU..... | 36 |
| 3.2.2 Multiscale approach (Discrete Wavelet Transform) | 39 |
| 3.2.2.1 Daubechies | 39 |
| 3.2.2.2 Coiflets..... | 40 |
| 3.2.2.3 Symlets..... | 40 |
| 3.3 Fetures Extraction And Selections | 41 |
| 3.3.1 Intensity features..... | 41 |
| 3.3.2 Shape features..... | 41 |
| 3.3.3 Textural features | 41 |
| 3.4 Classifications..... | 42 |
| 3.4.1 Artificial neural network | 42 |
| 3.4.2 K-nearest neighborhood..... | 42 |
| 3.4.3 Support vector machine..... | 43 |

CHAPTER IV

LITERATURE SURVEY

| | |
|--------------------------------------|----|
| 4.1 Enhancement..... | 44 |
| 4.1.1 Spatial domain methods..... | 44 |
| 4.1.2 Frequency domain methods | 47 |
| 4.1.3 Mixed methods..... | 49 |
| 4.2 Segmentation | 50 |
| 4.2.1 Thresholding techniques | 50 |
| 4.2.2 Region based techniques..... | 51 |

| | |
|---------------------------------------------------------|----|
| 4.2.3 Edge detection techniques..... | 52 |
| 4.2.4 Hybrid techniques..... | 53 |
| 4.3 Feature Extraction and Selection | 54 |
| 4.4 Classification | 56 |
| 4.4.1 Neural networks approache..... | 56 |
| 4.4.2 Linear and non-linear discriminant analysis | 75 |
| 4.4.3 Support vector machine | 58 |
| 4.4.4 Other methods | 59 |
| 4.5 Proposed Detection Methods..... | 60 |

CHAPTER V

APPLIED METHODS AND PROPOSED SYSTEM

| | |
|----------------------------------------------------|----|
| 5.1 Introduction | 64 |
| 5.2 Proposed System..... | 64 |
| 5.2.1 Preprocessing | 66 |
| 5.2.2 Enhancement | 68 |
| 5.2.2.1 The main enhancement techniques | 68 |
| 5.2.2.1.1 Unsharp masking..... | 68 |
| 5.2.2.1.2 Histogram modification | 70 |
| 5.2.2.1.2.1 Exponential modification approach..... | 70 |
| 5.2.2.1.2.2 Linear modification approach..... | 71 |
| 5.2.2.2 Enhancement techniques manually used | 71 |
| 5.2.3 Segmentation..... | 72 |
| 5.3.3.1 Seed-based region growing (SBRG)..... | 72 |
| 5.3.3.2 Multi thresholding OTSU's method..... | 72 |
| 5.2.4 Feature extraction | 73 |
| 5.2.4.1 Photometric features | 74 |
| 5.2.4.2 Haralick texture features | 75 |

| | |
|------------------------------|----|
| 5.2.5 Feature selection..... | 78 |
| 5.2.6 Classification..... | 81 |

CHAPTER VI
EXPERMENTS AND RESULTS

| | |
|----------------------------------------------|----|
| 6.1 Mammography Dataset and Used Tools | 83 |
| 6.2 Experiments | 84 |
| 6.3 Classification | 96 |
| 6.4 Evaluation..... | 97 |

CHAPTER VII
CONCLUSION AND FUTURE WORK

| | |
|-------------------------|-----------|
| 7.1 Conclusions | 103 |
| 7.2 Future Works..... | 104 |
| REFERENCES | R1 |

LIST OF TABLES

| | |
|-------------------------------------------------------------------------------------------------------------------|-----|
| Table 2.1 Difference between a Benign and Malignant Tumor | 15 |
| Table 2.2 Classification of Mass According to The Tensity | 20 |
| Table 2.3 Classification of Mass According to The MArgin | 20 |
| Table 2.4 Classification of Mass According to The Shape | 21 |
| Table 4.1 Comparison of Various Parameters | 46 |
| Table 6.1 Part of The Results After Classifying a Divided Image | 97 |
| Table 6.2 (a) Statistical Values of The System Performance | 99 |
| Table 6.2 (b) Statistical Values of The System Performance..... | 99 |
| Table 6.3 Statistical Values of The Sensitivity and FP at Threshold Point..... | 100 |
| Table 6.4 Comprehensive Summary of the Breast Cancer Detection Systems Using Different Techniques | 102 |

LIST OF FIGUERS

| | |
|--------------------------------------------------------------------------------------------------------------------|----|
| Figure 1.1 Block Diagram for Automated Detection System | 6 |
| Figure 2.1 Human Breast Anatomy..... | 10 |
| Figure 2.2 Normal versus Cancerous Cell Growth..... | 11 |
| Figure 2.3 Axillary lymph nodes near a breast with cancer | 13 |
| Figure 2.4 Rates of Breast Cancer Incidence In the world..... | 14 |
| Figure 2.5 (a) Benign Tumor, (b) Malignant Tumor | 15 |
| Figure 2.6 Calcifications Distributions | 17 |
| Figure 2.7 Classification Tumor Type According To Cs Form..... | 17 |
| Figure 2.8 Breast Mass..... | 18 |
| Figure 2.9 The shape and margin of a mass are strong signs of their Malignancy / Benignancy degree | 19 |
| Figure 2.10 Mammography Equipment | 22 |
| Figure 2.11 Views Taken In Screening Mammography | 24 |
| Figure 2.12 Views Taken In Diagnostic Mammography | 24 |
| Figure 2.13 Tomosynthesis: Working Mechanism | 25 |
| Figure 2.14 CT vs Tomosythesis | 26 |
| Figure 3.1 Gray-Scale Adjustments | 29 |
| Figure 3.2 Image After and Before Applying GHE and The Histograms | 30 |
| Figure 3.3 Image After and Before Applying (AHE) | 31 |
| Figure 3.4 Image After and Before Applying (CLAHE)..... | 32 |
| Figure 3.5 Low Pass Filter..... | 33 |
| Figure 3.6 High Pass Filter..... | 34 |
| Figure 3.7 Filtered Image By Using an adaptive un-sharp filter and the Original one..... | 38 |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 3.8 Result of Multiple OTSU Method | 38 |
| Figure 3.9 Daubechies Families | 39 |
| Figure 3.10 Coiflets Families | 40 |
| Figure 3.11 Symlet Families..... | 40 |
| Figure 4.1 (a) Original Image, (b) The Horizontal GVF, (c) Enhanced image through adaptive histogram equalization, (d) Horizontal GVF, (e) Generated mass mask. | 53 |
| Figure 4.2 (a) Original Image, (b) Segmented Image , (c) Enhanced image, (d) Segmented Image Based On Enhanced Image | 54 |
| Figure 5.1 Flowchart Digram of The System..... | 65 |
| Figure 5.2 Sub steps of Preprocessing Process..... | 66 |
| Figure 5.3 Mammographic Image with Light Problem | 67 |
| Figure 5.4 Block Diagram of Features Extraction and Selection phase..... | 74 |
| Figure 5.5 The Probability Density Function Of a Normal Distribution for Smoothness of Abormal Tissue | 79 |
| Figure 5.6 The Probability Density Function Of a Normal Distribution for Correlation of Abnormal Tissue | 79 |
| Figure 5.7 The Probability Density Function Of a Normal Distribution for Correlation of Normal Tissue..... | 80 |
| Figure 5.8 The Probability Density Function Of a Normal Distribution for Smoothness Of Normal Tissue | 80 |
| Figure 5.9 How NN works | 81 |
| Figure 5.10 Snapshot of ROC for Training Phase | 82 |
| Figure 5.11 Snapshot of Confusion for Training Phase | 82 |
| Figure 6.1 SnapshotShows The Original Image Before and After Applying 2D Medan Filter | 84 |
| Figure 6.2 Original Image, Binary Image After applying Global Threshold, Result Sfter keeping the Object with largest number of pixels and Final Result..... | 85 |

| | |
|---------------------------------------------------------------------------------------------------------|-----|
| Figure 6.3 Snapshot shows the result After Extracting Breast Profile | 86 |
| Figure 6.4 Mammographic Image Before and After Enhancing The Contrast | 87 |
| Figure 6.5 Binary Image ,the Segmented part of the pectoral muscle and The final result | 87 |
| Figure 6.6 Snapshot Shows The Image Before and After Removing The Pectoral Muscle | 88 |
| Figure 6.7 Image Before and After By Using Modifying The Contrast | 89 |
| Figure 6.8 Image Before and After Modifying the contrast | 89 |
| Figure 6.9 Image Before and After Histogram Equalization..... | 90 |
| Figure 6.10 Image Before and After Image Enhancement By Using Median Filter..... | 90 |
| Figure 6.11 Snapshot shows the segmented region after Applying SBRG..... | 91 |
| Figure 6.12 (a) Snapshot shows the segmented region after Applying OTSU's Method | 92 |
| Figure 6.12 (b) Snapshot shows the segmented region after Applying OTSU's Method | 93 |
| Figure 6.13 Snapshot of the Output Result of The Feature Extraction | 94 |
| Figure 6.14 Features Dataset | 95 |
| Figure 6.15 Calss Attribute..... | 95 |
| Figure 6.16 Snapshot of Sub-images (mdb025) | 96 |
| Figure 6.17 Performance Evaluation | 98 |
| Figure 6.18 ROC curve | 101 |

LIST OF SYMBOLS / ABBREVIATIONS

| | |
|-------|--------------------------------------------------|
| 1D | One Dimensional |
| 2D | Two Dimensional |
| 3D | Three Dimensional |
| ACR | American College of Radiology |
| ACS | American Cancer Society |
| AHE | Adaptive Histogram Equalization |
| AMA | American Medical Association |
| ANCE | Adaptive Neighborhood Contrast Enhancement |
| ANN | Artificial Neural Network |
| AUC | Area Under the Curve |
| BC | Breast Cancer |
| BLS | Biological Learning Systems |
| CADx | Computer Aided Diagnosis |
| CADe | Computer Aided Detection |
| CC | Cranial-Caudal |
| CLAHE | Contrast-Limited Adaptive Histogram Equalization |
| CLHE | Constrained Local Histogram Equalization |
| CM | Completeness |
| CR | Correctness |
| CRG | Constraint Region Growing |
| Cs | Calcifications |
| CT | Conventional mammography |
| DICOM | Digital Imaging and Communications in Medicine |

| | |
|------|---------------------------------------------|
| DM | Digital Mammography |
| DWT | Discrete Wavelet Transform |
| FAHE | Fast Adaptive Histogram Equalization |
| FN | False Negative |
| FP | False Positive |
| GCM | Gray Co-occurrence Matrices |
| GCE | Global Contrast Equalization |
| GHE | Global Histogram Equalization |
| GLDS | Gray Level Co-occurrence Matrices |
| GLN | Gray Level Number |
| HE | Histogram Equalization |
| HHS | Health and Human Services |
| HPF | High Pass Filter |
| KNN | K-Nearest Neighbor |
| LGF | Laplacian of Gaussian Filter |
| LHE | Local Histogram Equalization |
| LM | Llatero-Medial |
| LPF | Low Pass Filter |
| MAHE | Multi-scale Adaptive Histogram Equalization |
| MIAS | Mammographic Image Analysis Society |
| MC | Missed Cancers |
| MCs | Microcalcifications |
| ML | Medio-Lateral |
| MLO | MedioLateral-Oblique |
| NCI | National Cancer Institute |
| NN | Neural Network |
| PDF | Probability Density Function |
| RLS | Run Length Statistics |

| | |
|------|-------------------------------|
| ROI | Region Of Interest |
| SBRG | Seed-based region growing |
| SFM | Screen-Film Mammography |
| SVM | Support Vector Machines |
| SGLD | Spatial Gray Level Dependence |
| SNN | Simulated Neural Network |
| SPC | Specificity |
| TDLU | Terminal Duct Lobular Units |
| TN | True Negative |
| TNR | True Negative Rate |
| TP | True Positive |
| TPR | True Positive Rate |
| WHO | World Health Organization |

CHAPTER I

INTRODUCTION

Among the various kinds of cancer disease, breast cancer (BC) is one of the most dangerous cancer, also it is considered the second leading cause of death among women aged between 45-55 years, according to the American College of Radiology (ACR) statistics 1 out of 9 women might develop breast cancer during her lifetime[1].

Almost (269,000) of breast cancer deaths occur in those countries with low and middle incomes due to that disease is diagnosed at late stages which decrease the survival chance (for lack of awareness of early detection and limitations in health services); however, during 1989-1995 the BC mortality rates was declined by 1.4% per year and by 3.2% afterwards [1]. These declines have been attributed, in large part, to early detection [2]. Early tumor detection is the main cause of reducing mortality and the most effective treatment method to redress the disease without resorting to massive surgeries.

For early detection, mammography imaging is one of the most effective tools because of its ability to detect abnormalities more than 2 years before they are obvious. In fact, a tumor at the early stage appears as tiny shiny spots in the mammography images these spots are calcium sediments its scientific label is Microcalcifications (MCs) and the most important signs of a breast cancer. MCs have a subtle nature, in most cases unclear in the images and difficult to identify them where this constitutes a big challenge for radiologists due to the nature of human vision system.

Conversely, the second type of breast cancer called Masses may be easier to detect due to the nature of the size, shape and color contrast and at the same time some types of masses can be difficult to detect the result of tissue nature, which may look similar with normal breast tissue (parenchyma).

In order to ameliorate the radiologists' performance which has been leading to increase the treatment efficiency and the survival chance in the same time, reducing the misdiagnosis, sought a lot of specialists to other technologies to help the radiologists in reading mammography images and identify suspicious regions in the breast and provide a second opinion for them to detect microcalcifications and masses earlier and to diagnose if those MCs or mass benign or malignant. These technologies are Computer Aided Systems which supply better detection and diagnosis techniques (Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADx)).

This study proposes an automatic system to detect and diagnose abnormalities in digital mammography images.

The proposed system works as the following:

- 1- Preparing the image by using morphological operations and mathematical methods to remove the extra image's elements such labels and stamps.
- 2- Using some of image processing enhancement algorithms to enhance the breast's edges in order to segment region of interest (ROI) from the original mammographic image. To remove the pectoral muscle geometrical operations were used.
- 3- Segmenting the (ROI).
- 4- Importing (ROI) which was segmented to detect the suspicious area, region growing algorithm was used.
- 5- Extracting and selecting the features from (ROI) and also from normal tissue to be used as a training dataset in the next step, an Artificial Neural Network (ANN) classifier was used.
- 6- Classification process was divided into two phases :
 - i. Training phase where the training database was used for classifier training to have an ability to distinguish between the normal and abnormal tissues.

ii. Testing phase to determine if the suspicious region in the test mammogram image normal or abnormal which has been performed by using the trained classifier.

Building this system was implemented by using MATLAB® R2010b/2013a and evaluated by using the famous mammographic dataset (MIAS) Mammographic Image Analysis Society.

1.1 Motivations

Although breast radiological imaging and scanning has allowed for more accurate diagnose at earlier stages of the disease development, between 10% and 30% of malicious cancer cases (biopsy proven cancerous) are undetectable for different causes such as technical still there are problems with imaging procedure, abnormalities that are not observable, and abnormalities that are misinterpreted [3] generically, Missed Cancer (MC) point to collect of “non-detected” cancers. Evidence indicates that somewhere between 7-20% of mammograms with abnormalities currently detected also show signs in the previous mammogram when viewed in retrospect, which may be considered as false negative (FN) errors [3][4]. Since it is much more better to error on the side of safety, about 65-80% of breast biopsies result in benign diagnosis, which may be considered as false positive (FP) biopsies [3]. In addition, relieve emotional stress linked to the False Positive (FP) reading as well as the cost of FP misconstruction is vast.

So, the detection and diagnosis errors previously discussed form the basis of presented thesis work. That is, we believe that computer aided decision methods can improve both false positive (FP) and false negative (FN) diagnosis rates. Although there are a few commercially available automated detection systems (discussed below) that are used as imaging checking systems in conjunction with the radiologist interpretation, the idea of automated classification has not been used clinically as of yet to any extent [5].

In considering computer technologies we find that there is so important differences between suspicious area detection and classification of it. Where detection process must always be done before classification and it can be performed by radiologists by using traditional means or any automatic method, once discovering any abnormalities, by any method, it must be classified in that truly class, this can be achieved by human estimation with helping from automated systems.

Here, the work has been presented, which may be seen as groundwork for an overall automatic detection and diagnosis system. This system facilitates the radiologist's task and helps them to detect and classify the abnormality whatever would be. Briefly the system is complement to radiologists assessment.

1.2 Thesis Challenges

Analyzing mammographic images is not an easy matter at all as it is a complicated work due to following influential factors:

- 1- Large contrast of lesions appearance.
- 2- Lesions often tend to be hidden, and density of breast tissue makes detection more difficult.
- 3- CADx systems can have some serious implications.

1.3 Thesis Goal : Automatic System

The proposed system will acquires the image that will go through the following processes:

- 1- PREPROCESSING
 - a. Suppressing all additional contents labels, light problems and noise.
 - b. Removing a chest muscle.
 - c. Extracting the breast profile.

2- ENHANCEMENT

- a. Enhancing the preprocessed image contrast to increase the gray level values of the region of interest(ROI).

3- SEGMENTATION

- a. Determining the exact place of (ROI) in the enhanced image as accurate as possible.
- b. Separating (ROI) from the naturalistic breast tissues.

4- FEATURE EXTRACTION AND SELECTION

- a. Extracting 15 photometric and texture features of the (ROI).
- b. Selecting the most useful ones in order to design a classifier with high accuracy .

5- CLASSIFIER

- a. In this step, training the classifier is done then the tested image area is classified.

The Figure 1.1 below represents the main steps of the proposed automated detection system of breast cancer in the mammographic images scheme.

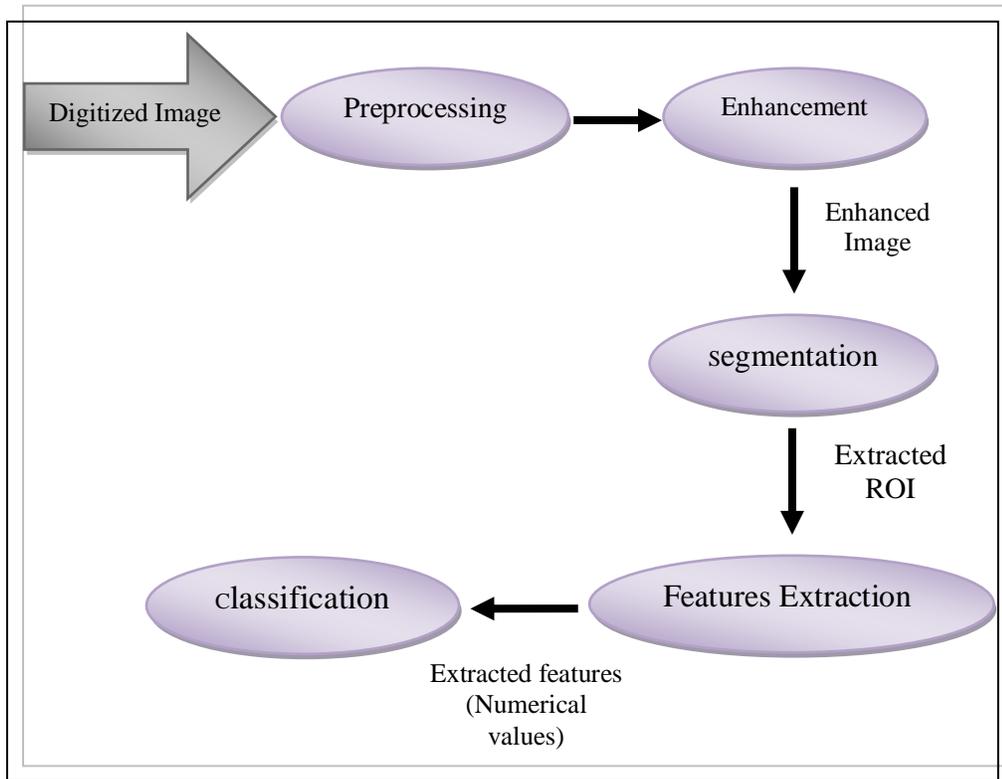


Figure 1.1 Block Diagram for Automated Detection System

1.4 Thesis Outline

The chapters of this research have been arranged as follows:

Chapter One – Introduction : This chapter shows the motivations of the thesis project and its main goal, the objectives and thesis outline.

Chapter Two – Background: This chapter has separated into the following sections:

1. Medical background

This section provides an overview of cancers and highlights on breast cancer where it defines what is it, its types and what are the areas that may be susceptible to cancer by clarifying the medical anatomy of the breast; In addition, this section shows the types of breast cancer lesions masses or calcifications and whether it is malignant or benign, also shows the difference between benign and malignant tumors.

2. Mammography imaging techniques, CADe and CADx systems:

This section concentrates on mammography imaging as one of the most common imaging techniques in medical imaging and advanced forms, also this section define Computer Aided Diagnosis and shows a short history of CADx and the used techniques in CADx systems.

Chapter Three – Review of Image Processing Methods: Image processing techniques and methods applied in the detection and diagnosis systems are discussed in detail.

Chapter Four - Literature Survey: In this chapter a survey of previous studies in image processing for mass and calcifications detection is introduced, and a detailed description for the algorithms in each detection step is given as well as some of the proposed detection systems.

Chapter Five – Applied Methods And Proposed System: The proposed algorithms methods that is implemented in this work have been discussed as well as this chapter presents statistical comparisons between the current work and the work done by other researchers that were displayed in the fourth chapter.

Chapter Six – Experiments and Results: Experiment’s steps have explained clearly with snapshots of the proposed system besides with final results after evaluating the work were presented in this chapter.

Chapter Seven – Conclusions and Future Work: The last chapter gives the conclusion which has been drawn from the research and puts some overtures for future works.

CHAPTER II

MEDICAL AND TECHNICAL BACKGROUND

This chapter has been divided into two sections, the first one provides an overview of cancer and highlights specifically on breast cancer through the detailed anatomy of the breast and determines most vulnerable regions to injury; moreover, this part lists in detail the lesion types (Calcifications and Masses) that differentiate between benign and malignant tumors according to their characteristics. Finally, it mentions the importance of detecting breast cancer at its early stages. Second section focuses on the technical side that starts with a detailed overview of mammography imaging as one of the most common imaging techniques in medical imaging besides the advanced forms of used systems that play an important role during detection and diagnosis applications. Since CADe and CADx most commonly used technologies, Has been allocated part of this chapter to show the history of the development of CAD systems.

2.1 Medical Background

2.1.1 Breast anatomy

Breasts are located in the upper ventral region in human at both sides (left, right) of the stem and each includes part of frontal area of a human body from the beginning of the second rib to sixth which contain the mammary gland. After giving the baby birth , mostly, the glands excrete milk with stimulation. These glands exist in both male in only primitive form and female with some exceptions.

The adult female breast consists of gland tissue, fibrous tissue, fatty tissue, blood vessels, nerves and ducts. The breast has numerous lobes, usually 15 to 20, which are composed of lobules. These consist of alveoli and lactiferous ducts. These lactiferous ducts enlarge to form a small lactiferous sinus, which accumulates milk during lactation. The milk leaves the breast through some holes in the nipple. (See Figure 2.1), The fibrous tissue lies on the entire surface of the breast and connects the lobes together. The fatty tissue covers the surface of the gland, except for the areola, and is located between the lobes. Usually, this tissue is abundant and determines the form and size of the gland [6][7].

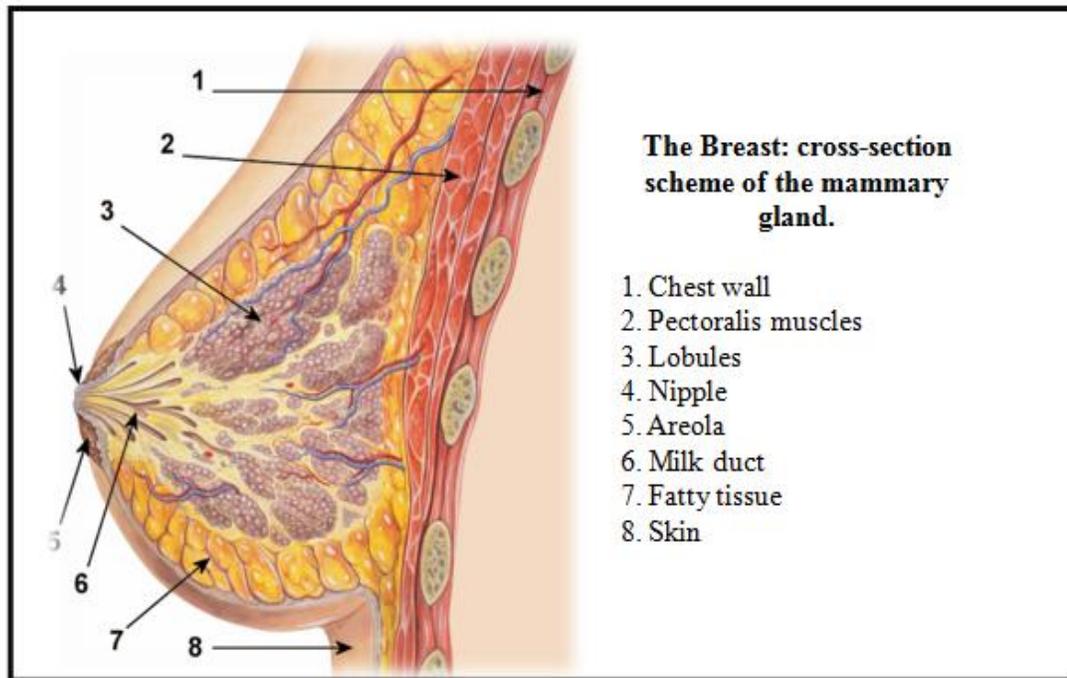


Figure 2.1 Human Breast Anatomy[8].

2.1.2 Breast cancer

In normal case body's cells are growing and dividing for a specific duration time and then stop while these processes continue producing cells and dividing them unregulated this action caused disease known *CANCER*. (See Figure 2.2)

Medically, cancer is defined as a malignant tumefaction and it is possible to move to distant parts of the body by lymphatic system and blood circulation actions.

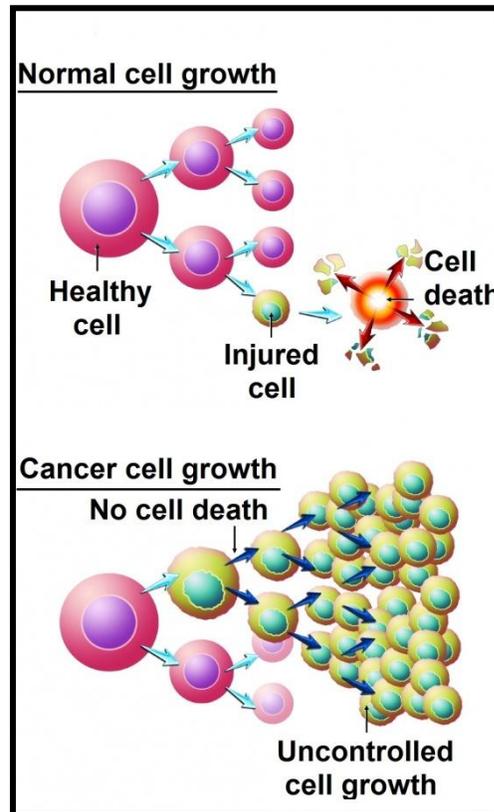


Figure 2.2 Normal versus Cancerous Cell Growth [8]

As all tumors are not cancerous there are 2 types of neoplasm benign and malignant. The cells in benign neoplasm are not spreading uncontrollably which makes removing it easy and also growing back after removing it , to some extent rare.

Malignant neoplasms, on the other hand, are growing quickly and spreading from the premier tumor site to normal parts of the body and this action is known as a *METASTASIS*. These two types will be discussed in more details later.

Determining the a specific cause of the cancer is somehow complex because there are many known factors that increase the risk of this disease such as smoking, obesity, environmental pollution, some infections, genetic causes and lack of physical activity. These factors can cause the disease by changing the activity of body cells or damage genes indirectly or directly. There are more than 190 different known of the cancer.

One of the riskiest types of cancers which is considered the most common reason of death between the women worldwide is a *Breast Cancer* (BC), according to World Health Organization (WHO) that 1.6 million new cases is recorded annually in the world, Figure 2.4 shows rates of Incidence of breast cancer in the world. BC is a cancer that infects the tissues of the breast, ordinarily the tubes that carry the milk on to a nipple which medically known as ducts and glands that responsible to make the milk lobules ,and and occurs in both genders men and women, but rarely among males [9].

As mentioned previously, each breast includes between 15 to 20 lobes that are linked together to the nipple via a complex structure of Interrelated ducts. Each lobule has between 10 to 100 Terminal Duct Lobular Units (TDLU), the areas where breast cancer originates (Figure 2.3). When the tumor has no spread actions via the basal membrane but is totally consist of the lobules of the ducts, cancer is called in situ and when the cancer has broken through the basal membrane it is called *INVASIVE*, and chances on metastases increase sharply [10].

In 2012, it is estimated that there were more than 2.9 million women living in the US with a history of invasive breast cancer as of January 1, and an additional 226,870 women are expected to be diagnosed. The median age at the time of breast cancer diagnosis is 61. About 20% of breast cancers occur among women younger than age 50 and about 40% occur in those older than 65 years[11].

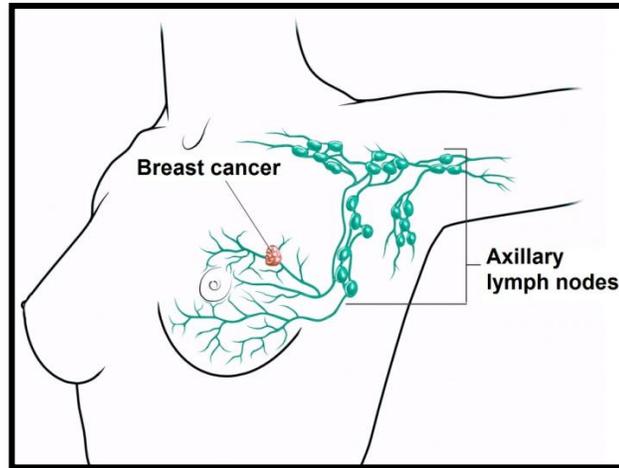


Figure 2.3 Axillary lymph nodes near a breast with cancer

In fact, each type of cancers has some factors called risk factors. These factors affects the chance of having a cancer. In breast cancer there are several factors: lack of procreation or breastfeed, disorders in hormones levels, regimen and corpulence, life style such as smoking or lack of physical activity, mutations and inheritance, etc. But not every time existence of these factors mean that he woman will get cancer or not and this fully applies to all cancer types. For example, there are incidences among women without having risk factors except being woman and being older.

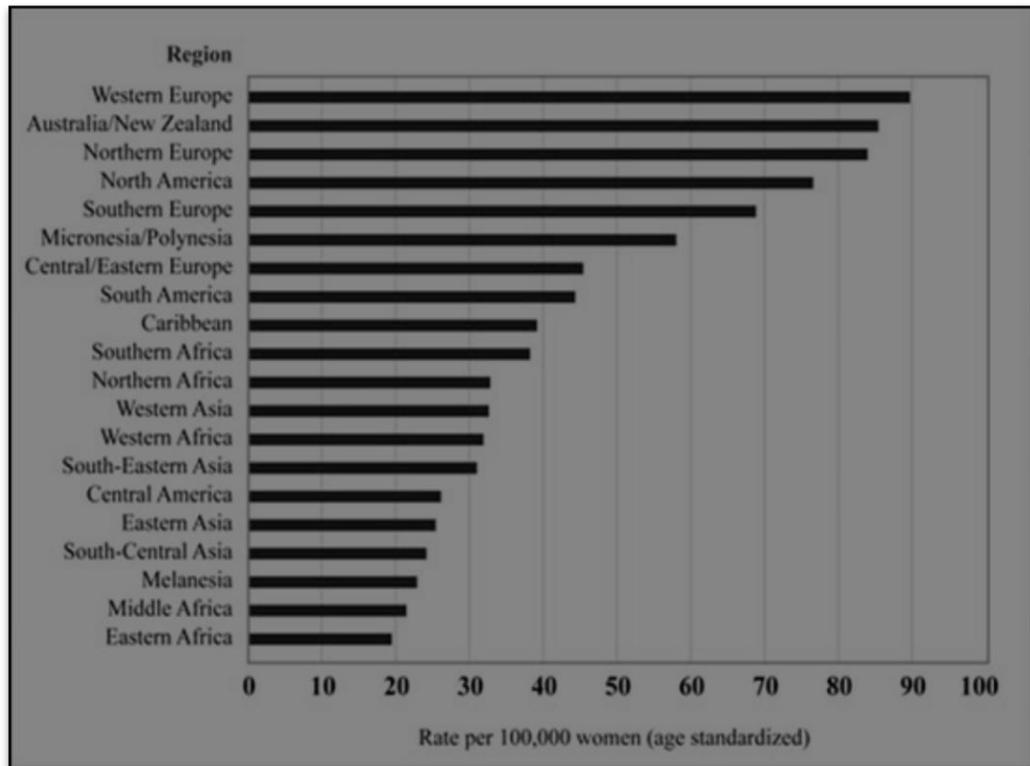


Figure 2.4 Rates of Breast Cancer Incidence In the World

2.1.3 Benign and malignant tumors

Tumors have attributes, in fact, some of these attributes such as position or tumor's size are easier to distinguish specially by knowing the probability distributions. But some of them like texture and shape are more difficult to distinguish. The two main types of cancerous tumors are either malignant or benign, the characteristics of each play a main role to know which type the tumor belongs to. For instance malignant tumor tends to be in unorganized shape with unclear borders (speculation at the boundaries), while benign frequently shows in regular shape (spherical, ovoid) with sharp borders.

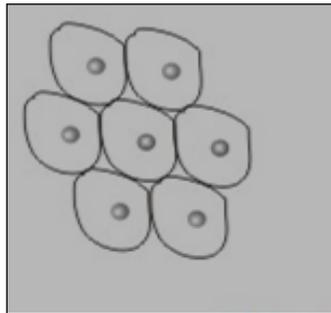
On the other hand, the calcifications that are created breast cancer may appear in three forms :

- 1- granular or crushed stone,
- 2- casting or linear,
- 3- powderish or amorphous

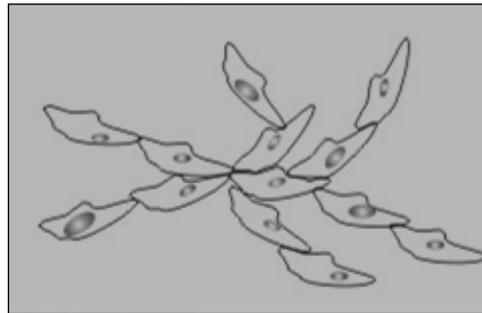
The following table (Table 2.1) compares between malignant and benign tumors.

Table 2.1 Difference between a Benign and Malignant Tumor [17]

| Benign Neoplasms | Malignant Neoplasms |
|------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| Moving mass | Fixed mass |
| Soft and clear round with besetment fibrous capsule. | Irregular shaped with no capsule |
| Cells multiply slowly. | Cells multiply rapidly. |
| The tumor is growing by expanding and pushing away and against surrounding tissue. | The tumor growth by invading and destroying surrounding tissue. |
| Mass is movning and not linked with surrounding tissue. | Mass is settled and linked with surrounding tissue and fixed in surrounding tissue pointedly |
| Never diffuses to remote parts | Always diffuses to other remote parts |
| Remove it easy and does not recur again | Remove it more difficult and may recur again |



(a)



(b)

Figure 2.5 (a) Benign Tumors, (b) Malignant Tumor

2.1.4 Breast cancer lesions

2.1.4.1 Calcifications

One of the most important abnormalities which can be seen on mammography images are called (Cs) *Calcifications*, and they have been shown in the soft tissue of the breast as white dots of calcium and they are very small dots. Although, existence of the calcification does not mean existing of breast cancer and not always form a danger sign. But calcification sometimes is a very important sign that may indicate a cancer in its primary stages which increases the chance of cure. There are two types of calcifications:

- **Macrocalcifications:** Big dots of calcium often do not mean breast cancer.
- **Microcalcifications:** They are so small dots of calcium and may appear in (groups) clusters, or in forms such as lines or circles. This type is strongly related to breast cells' activities.

Commonly, excessive cell growth isn't a cancer, where occasionally clusters of microcalcifications can be a sign of breast cancer in its prime stages, specially, scattered microcalcifications which indicate a benign tumor. Calcifications are usually found by chance on a mammogram (breast x-ray) which is taken for another problem, or during routine breast screening, where they appear as small white dots. The results of breast imaging are checked independently by one or two radiologists, who look carefully at the shape, size and pattern of the calcifications. Then they write a report in which the calcifications are usually categorized as being benign, indeterminate (uncertain) or suspicion of being cancerous. Types of calcifications distribution :

- **Diffuse or Scattered:** Diffuse calcifications may be scattered calcifications or multiple similar appearing clusters of calcifications throughout the whole breast.
- **Regional:** Scattered in a larger volume (> 2 cc) of breast tissue and not in the expected ductal distribution.

- Clustered: At least 5 calcifications occupy a small volume of tissue (Linear: calcifications arrayed in a line, which suggests deposits in a duct.
- Segmental: Calcium deposits in ducts and branches of a segment or lobe[12].

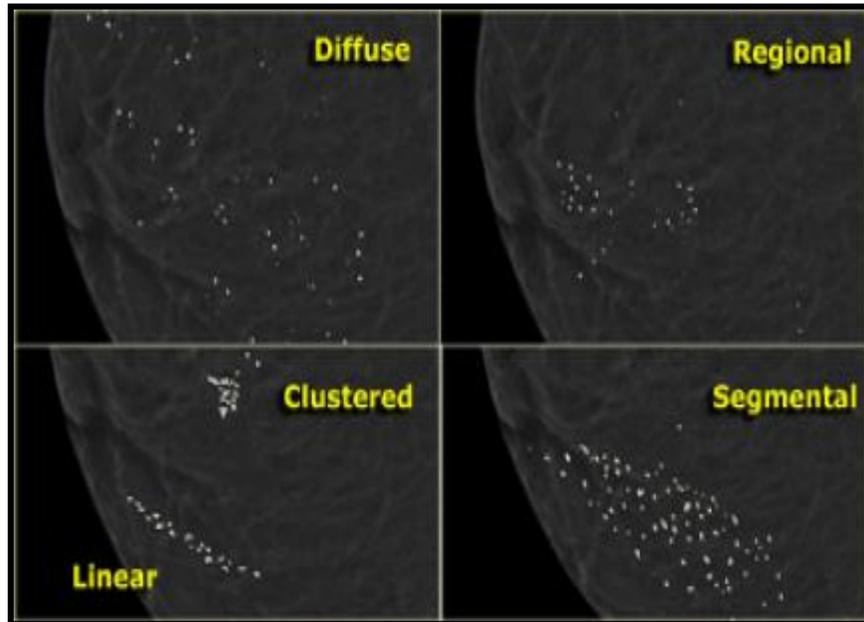


Figure 2.6 Calcifications Distributions

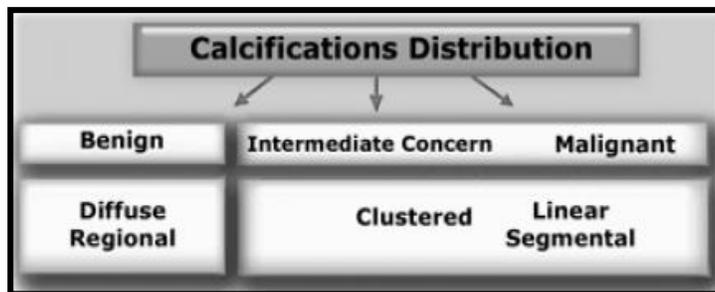


Figure 2.7 Classification Tumor Type According to Cs Form

2.1.4.2 Masses

The definition of mass is an affected area which includes the suspicious region where radiologists can characterize by masses' attributes such as shape, boundaries and margins. Often times, masses are present in breast cancer with or without the existence of calcifications; however, a cyst is muscular bag just contains fluids and it is not malignant area. It might appear as a mass in ultrasound films but the difference can be distinguished by fine needle aspirations.

Due to the similarity the intensities and the morphology between the masses and the normal breast tissues, masses detection is more difficult comparing with calcifications in the breast [14].

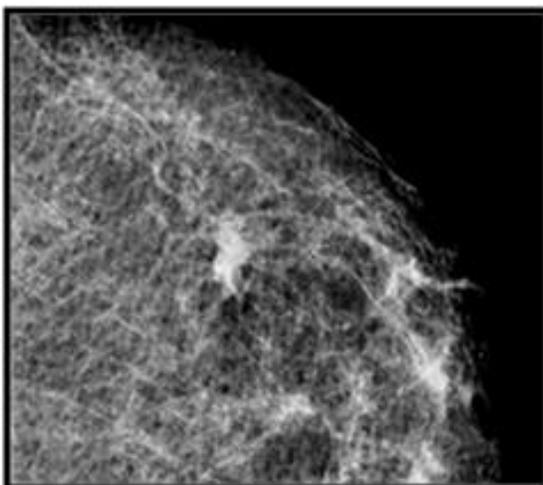


Figure 2.8 Breast Mass

To evaluate the probability of breast cancer by the radiologist there are helpful parameters such as location, margins, density, and shape of the mass.

Benign masses are compact, and nearly circularity or oval but in some cases benign lesions might be in a speculated form or unclear borders. On the other hand, malignant masses often have an irregular shape, unclear boundary, and are in many cases surrounded by a shiny pattern of linear thistles[15].

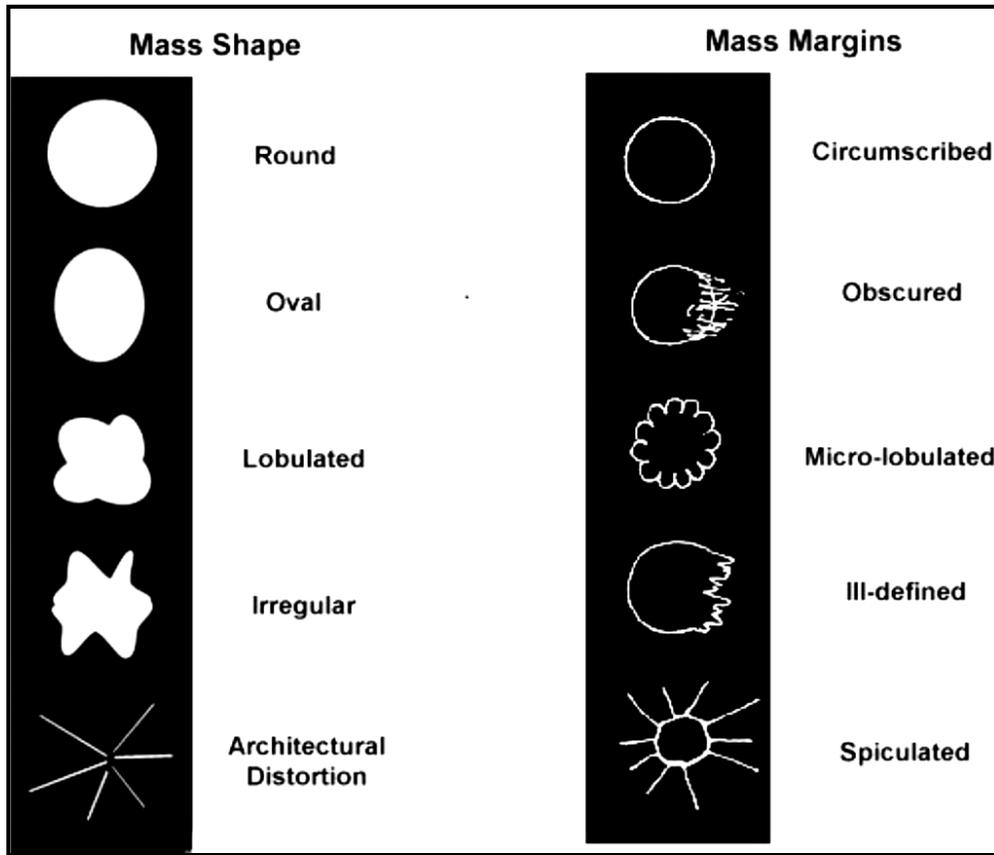


Figure 2.9 The Shape and Margin of a Mass is Strong signs of them Malignancy / Benignancy Degree [18]

A breast mass is occupying an area and usually it has a shape tends to be with a convex outside edges and more density in the middle than the edges. Often, fat cells no permeate the masses as much with the normal breast tissue, Breast masses might be in a shape out of four point assessments. These assessments are oval, round, lobular or Irregular shape. The masses can also described according to the amount of fat cells of suspicious cells which known as density, in addition the characteristics of mass edges (mass margins)[16]. Tables (Table 2.2, Table 2.3 and Table 2.4) give a summary of masses characteristics. Table 2.2 describes that density of breast mass is described according to the amount of fatty tissue of the breast, where mass density be more suspicious for malignant breast cancer when the mass density is high.

Table 2.2 Classification of Mass According to The Density[16].

| Density | Benign | Suspicious | Malignant |
|----------------|--------|------------|-----------|
| Completely fat | √ | | |
| Low | √ | | |
| Medium | √ | √ | |
| High | | | √ |

Table 2.3 shows the types of breast mass margin and illustrates that when the mass margin is ill-defined or spiculated be more suspicious for malignant cancer in the breast. Table 2.4 shows that the masses appearing with an irregular, fuzziness shape is a clear sign for malignant cells in the breast, when the breast mass has shape looks more 'lobule' than an obvious round or oval , then it is somewhat a sign of the presence of malignant cancer.

Table 2.3 Classification of Mass According to The Margin[16].

| Margin | Benign | Suspicious | Malignant |
|------------------------------|--------|------------|-----------|
| Well-defined | √ | | |
| Obscured/ 75% hidden or more | | √ | |
| Microlubulated | | √ | |
| Indistinct and ill-defined | | √ | √ |
| Speculated | | | √ |

Table 2.4 Classification of Mass According to The Shape[16].

| Shape | Benign | Suspicious | Malignant |
|-----------|--------|------------|-----------|
| Round | √ | | |
| Oval | √ | | |
| Lobular | √ | √ | |
| Irregular | | | √ |

2.1.5 The Importance of Detect Breast Cancer Early

Screening exams point out to make some tests to find disease instance, breast cancer, in people who do not have clear disease's symptoms , as well as aim to find it in advance before symptoms appear. Detection of breast cancer in early stages by screening mammography gives a big chance for a woman to survive by using a process that lets diagnose breast cancer earlier than might have occurred, Actually classifying and determining a kind of tumor in order to decide if it is benign or malignant is not an easy process while there are many obstacles facing the radiologists, In some cases, it may detect tumors with 5 mm volume and surrounded by greasy tissue easily.

2.2 Technical Background

2.2.1 Advanced of mammography imaging

Digital mammography and Breast Tomosynthesis are two recent sophisticated forms of mammography. In 2000, Food and Drug Administration (FFDM) has approved to sale the first digital mammography unit, as any new technology there is a need to emulate its efficiency by comparing with known standard. Screen-film mammography (SFM) the standard for breast detection which known as "gold standard".

SFM includes well established protocols and high quality for breast cancer detection and diagnosis. It is also has a well decision to detect microcalcifications which are considered one of the clear signs of early breast cancer. These detectors are similar to those found in digital cameras. To get breast images which can be shown on the screen it should use the electrical signal, these image also can be printed on film similar to that used in conventional mammograms. From the patient's point of view, having a digital

mammogram is essentially the same as having a conventional film mammogram [25].

2.2.1.1 Digital mammography (DM)

One of latest technique in the medical imaging field is mammography which has been made in so high contrast, working with high resolution film which makes it very suitable and useful to examine the breast. In fact, this technique basically an X-ray, uses a low- energy X-ray system usually around 30 kVp (an X-ray technique is still used frequently in medical imaging and it is considered as the oldest technique). X-ray mammography can be used in two types: Screening and Diagnosis.



Figure 2.10 Mammography Equipment

1) Screening Mammography

Mammography plays a very critical role in breast cancer detection at its early stages due to the changes which can be shown in the breast. these changes may be occur in the breast up to two years before an Infected or physician can feel them.

Current guidelines from the United State Department of Health and Human Services , American Cancer Society, American Medical Association and American College of Radiology are supporting the women to conduct an annual screening mammography, with age 40 years. Researchs have shown that annual mammograms help to detect breast cancers earlier which increases the chance of survival and making the treatment process more easily.

The National Cancer Institute (NCI) advises women less than 40 years old who have a genetic history or suffered from breast cancer should start screening program because the risk rate of injury is high. By taking two x-ray images for both breasts; typically cranial-caudal view, mediolateral oblique (MLO) and (CC); figure 2.11 shows these directions in taking the images.

2) Diagnostic Mammography

Diagnostic Mammography is an x-ray examination of the breast in a woman who is symptomatic and it is more involved and time-consuming than screening mammography, also it may done after a screening process to evaluate the interested area, and can be used to evaluate breast abnormalities (tumor, distortions) of clinical results for a patient which have been noted by the affected woman , regular screening, doctors or nipple discharge.

The goal of diagnostic mammography is to pinpoint the size and location of the breast abnormality and to image the surrounding tissue and lymph nodes or to rule-out the suspicious findings.

Typical views for diagnostic mammograms include lateromedial (LM) and mediolateral view (ML) along with the CC and MLO views as defined on the previous page. For specific problems additional special views such as exaggerated cranial-caudal, spot compression, and magnified may be taken. (Spot compression, magnification views often to evaluate micro calcifications and Ductogram / Galactogram for imaging the Breast Ducts are some of the special mammography views as shown in Figure 2.12) [21].

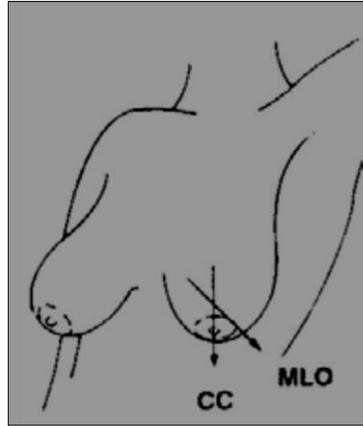
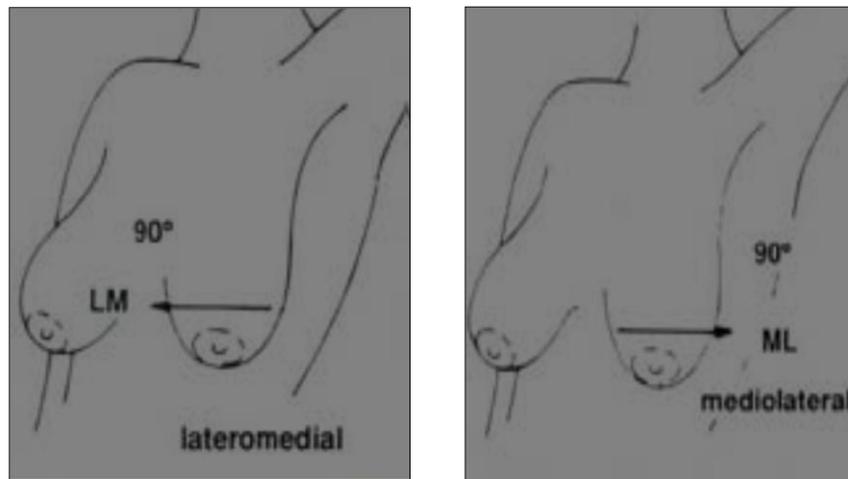


Figure 2.11 Views Taken in Screening Mammography



(a)

(b)

Figure 2.12 Views Taken in Diagnostic Mammography

(a) LateroMedial (LM) Mammographic View (Left)

(b) MedioLateral (ML) Mammographic View (Right)

The rate of a false negative or misdiagnosis in mammography reach around 10% and this is due to slightly intensive tissues stow the breast cancer. In fact, the nature of the emergence of normal breast tissues has a great superposition with the nature of the emergence of breast cancer in mammogram image. A meta-analysis survey of programs in countries with regular screening found 52 percent over-diagnosis [22].

2.2.1.2 Breast tomosynthesis

Breast Tomosynthesis also known as Digital Breast Tomosynthesis (DBT) is a 3D mammography and the latest technology in breast mammography imaging that performs high resolution to take images. Technically, tomosynthesis machine able to take 15 digital images of the breast sequentially at a partly several angles in few seconds where an x-ray tube moves in an arc around the breast (See Figure 2.13).

Collected images produces 3D images for individual breast tissue layers which helps the radiologist to find out good internal details at breast structure layer by layer clearly, examine the breast tissue 1 mm layer at a time, and reduce overlapping breast tissue.

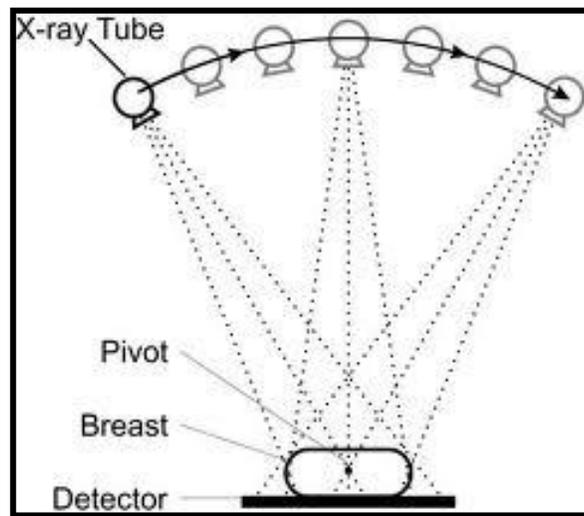
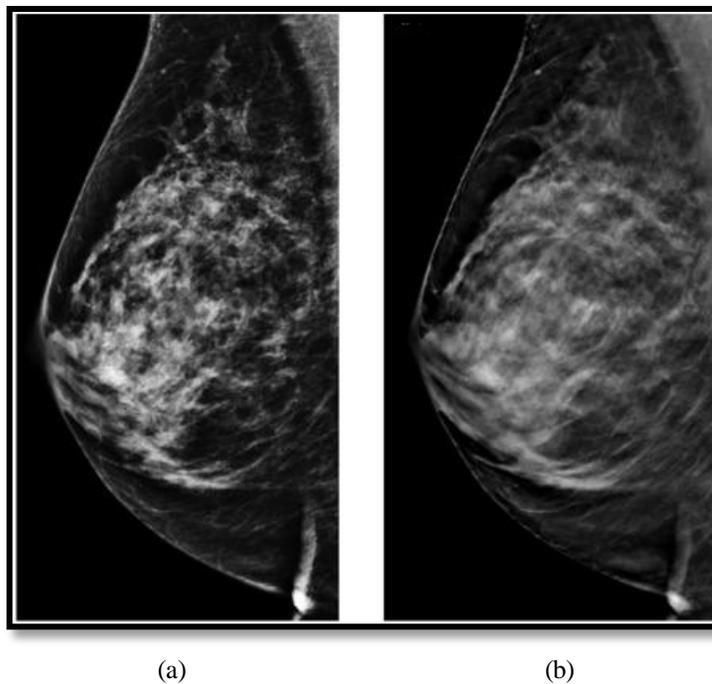


Figure 2.13 Tomosynthesis: Working Mechanism [128]

On the other hand, tomosynthesis is unable to offer the extremely narrow slice widths that conventional mammography (CT) offers because that the data acquired are incomplete. Breast imaging the first tomosynthesis interest and it is an expansion and development of mammography. it offers better detection results with slightly increasing in radiation. [38].

Moreover, there are probable clinical problems that includes tissue distortions, asymmetries between (left, right) breasts and confluent densities associated with benign fibrosis can be detected by mammography.

Figure below shows the difference between image taken by CT and that one taken by tomosynthesis, where (a) shows the calcifications, but no more abnormalities can be notated but , (b) tomosynthesis shows a distortion type of star upper section of image and it is clear to note that the calcifications are not shown very well on the tomosynthesis image as compared to the conventional CT image 2D FFDM.



(a) (b)
Figure 2.14 CT vs Tomosynthesis
(a) Right MLO (2D) ,(b) Right MLO Tomosynthesis

2.2.2 Computer-aided detection (CAD)

To support medical decision making , increase diagnostic efficiency advanced image processing methods and machine learning techniques can be used to alert radiologist to areas which needs further analysis, to determine and classify suspect lesions (mass or calcification).

These methods and techniques pose together to create automated systems such as (CADe) Computer Aided Detection system which can be used to aid radiologists to detect suspicious regions in the mammography image and (CADx) Computer Aided Diagnosis system to assist them to classify the lesion either benign tumor or malignant.

Essentially, CAD systems fundamentally based on highly complex pattern recognition, image processing and artificial intelligence algorithms to be able to deal with mammography images which are scanned for suspect structures in the breast, for optimizing those algorithms, ordinarily a high number of images are needed.

In general, to develop CADs it must copy a dataset of mammography digital images (with Digital Imaging and Communications in Medicine (DICOM) format) into a CAD server to pass through several steps in order to prepare, process and analyze them. (CADe) and (CADx) act as a radiologist assistant to give a second opinion but the last decision is made by the radiologist him/herself.

Any CAD system does the following steps sequentially (Chapter III discusses these steps in more details):

- 1- Digitizing Mammogram image
- 2- Image Preprocessing
- 3- Image Detection
- 4- Image Segmentation
- 5- Feature Extraction & Selection
- 6- Classification
- 7- Evaluation

CHAPTER III

REVIEW OF IMAGE PROCESSING METHODS

3.1 Image Enhancement

The nature of masses and calcifications in the mammography image has no clear color contrast to somewhat with surrounding normal breast tissues. This characteristic makes detecting and diagnoses lesions harder, and thus it should firstly enhance the image contrast before processing to improve the interpretability or perception of information in images for human viewers, or to provide better input for other automated image processing techniques.

Image enhancement develops a digital image quality for visual inspection or for machine analysis by various techniques including many and many image processing filters, algorithms and methods. Also it is a very important step to prepare a concerned image for next operations; therefore it must use appropriate methods to suppress images' problems such as (Contrast, noise and Lighting problems, etc.) for more accuracy.

The techniques can be divided into three categories: Spatial Domain Method, Frequency Domain Methods and Multiscale Approach.

3.1.1 Spatial domain methods

3.1.1.1 Simple intensity transformations

3.1.1.1.1 Contrast stitching

The idea of this technique is increasing the range of the gray-levels to enhance an image with low contrast by using a linear curve $T(r)$ which may be a thresholding function.

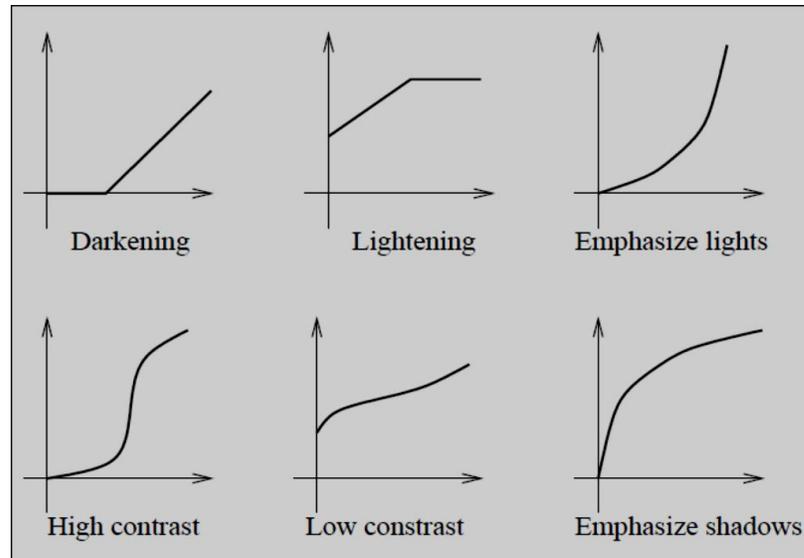


Figure 3.1 Gray-Scale Adjustments

3.1.1.1.2 Grayscale adjustments

Thresholding is the easiest case, where the intensity profile is replaced by using a step function, and select T as a threshold where the pixel value is replaced by 0 or 255 value, if its gray level value above the threshold value in the original image, it given a value of 255 in the resulted image, otherwise it is mapped to 255. By using this technique it can be achieved some interesting effects shows in Figure 3.1

3.1.1.1.3 Image negatives

This technique transforms high intensities to low and vice versa as shown in the following function:

$$O = N(I) \quad (1)$$

Where, $N(I) = (L - 1) - I$; L number of gray levels in I and O output image.

3.1.1.2 Histogram processing

Histogram equalization (HE) is one of the most common enhancement technique; it can be categorized into two methods:

3.1.1.2.1 Global Histogram Equalization (GHE)

Applying this technique is easy and effective with low contrast image, the main idea is standardize the distribution of the image values in its histogram; in other words standardize the intensity values of pixels on the scale by using a gray scale transformation functions that leads to get an enhanced image. By using HE technique as follows function will get an output image I_{new} :

$$I_{new} = G_{min} + (G_{max} - G_{min}) \times \sum_{i=0}^{I_{org}} H(i) / (m \times n) \quad (2)$$

Where $m \times n$ a size of image has a histogram $H(i)$ and G_{max}, G_{min} Intensities range. Figure 3.2 shows the result before and after applying GHE it easy to note the resulting histogram became flat.

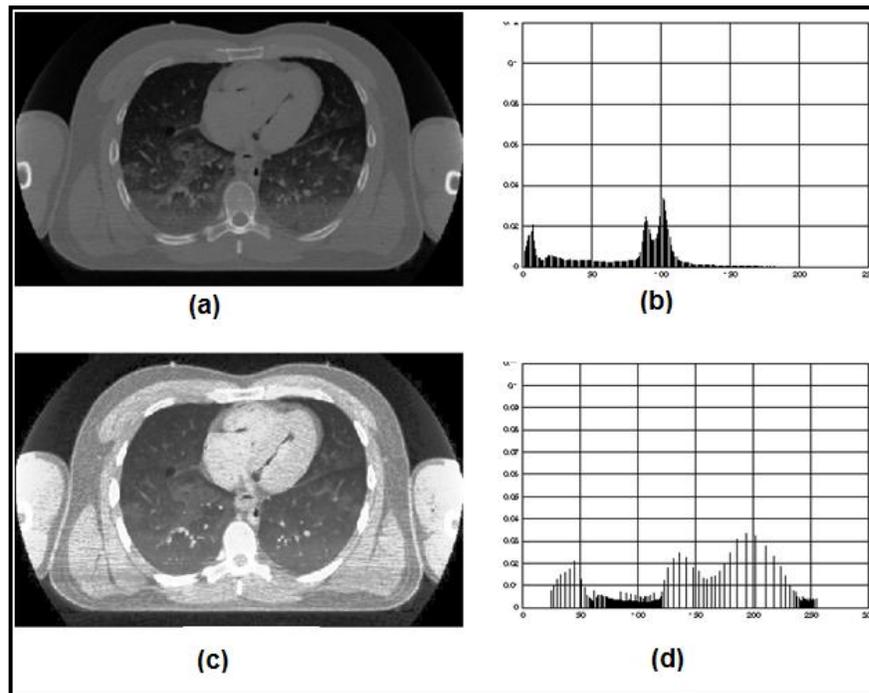


Figure 3.2 (a) Original chest CT (b) Histogram of original, (c) Result image processed by GHE and (d) Histogram of output image.

Histogram stretching; is included within global histogram modification techniques; it uses a linear transferring technique as the following function:

$$I_{new} = G_{min} + (I_{org} - I_{min}) \times (G_{max} - G_{min}) / I_{max} - I_{min} \quad (3)$$

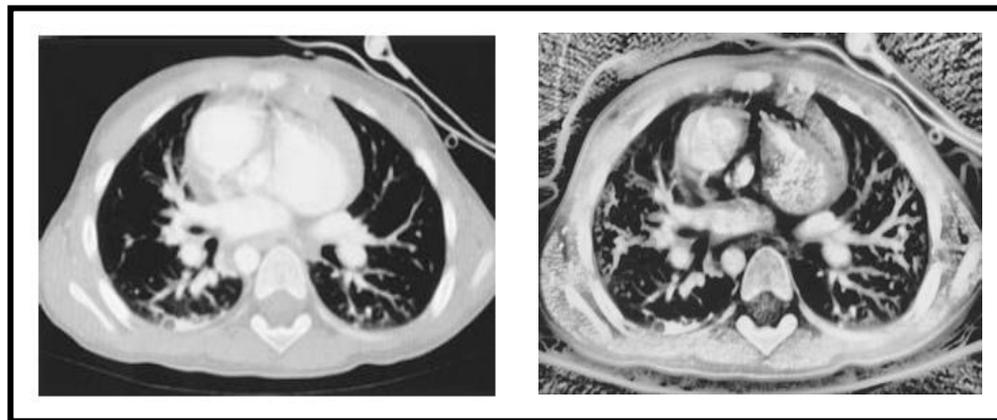
Where $[I_{org}, I_{min}]$ is an intensity range of the original image and $[G_{max}, G_{min}]$ is for the resulted image.

3.1.1.2.2 Adaptive histogram equalization (AHE, or LHE: Local histogram equalization).

When a new value of pixel intensity is calculated from its input value and any additional image information derived from the properties of a local image, this method called adaptive contrast-enhancement mapping.

$$I(x,y) = f(i(x,y), DN(x,y)) = fN(i(x,y)). \quad (4)$$

$N(x, y)$, the contextual region, is some spatial neighborhood of (x, y) in the image that includes the pixel of interest and $DN(x, y)$ is some collection of measurements over $N(x, y)$.



(a)

(b)

Figure 3.3 (a) Original Image of chest CT scans and (b) Result image after using (AHE)

Regrettably, besides that this technique offers an excellent result, but it also AHE enhances a noise if any.

In order to lessen the previous behavior, *Contrast-limited adaptive histogram equalization* (CLAHE) was designed in which the perceptual task leads to the desire to disregard the “salt-and-pepper” noise, In other words, In the comparatively homogeneous areas of the image, this technique tends to magnify the noises and to avoid this problem "contrast limited adaptive histogram equalization" (CLAHE). [40]

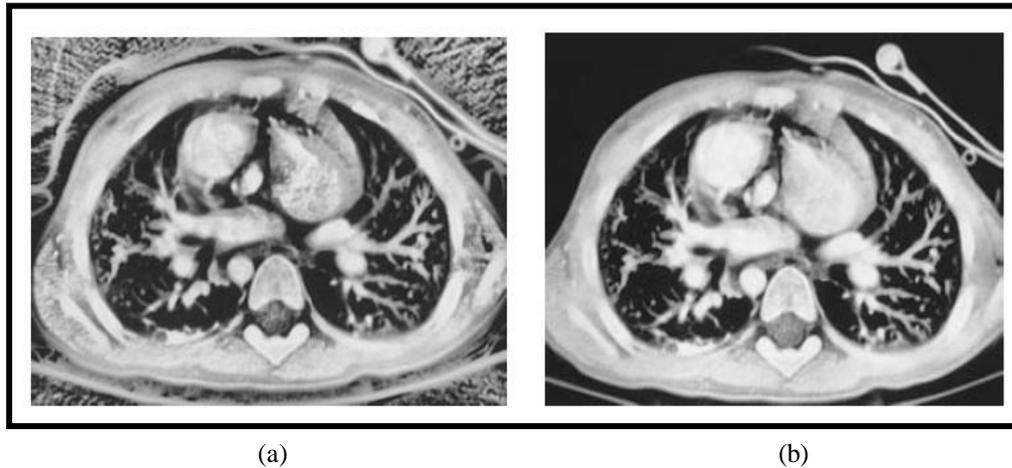


Figure 3.4 (a) Image after applying (AHE) (b) Image after applying (CLAHE)

3.1.2 Frequency domain methods

Image enhancement in the frequency domain is unambiguous way to improve the input image. It calculates the fourier transform of the image. Enhancing the image by multiplying the result using a filter (rather than convolve in the spatial domain), then to produce an enhanced image it takes the inverse transform.

Types of Filtering:

3.1.2.1 Low pass filter (LPF)

LPF runs on eliminate the high frequency components in the image and the resulting image be in blurring and smooth form (and thus a reduction in sharp transitions associated with noise).

Types of LPF:

- Ideal filter.
- Butterworth filter
- Gaussian filter

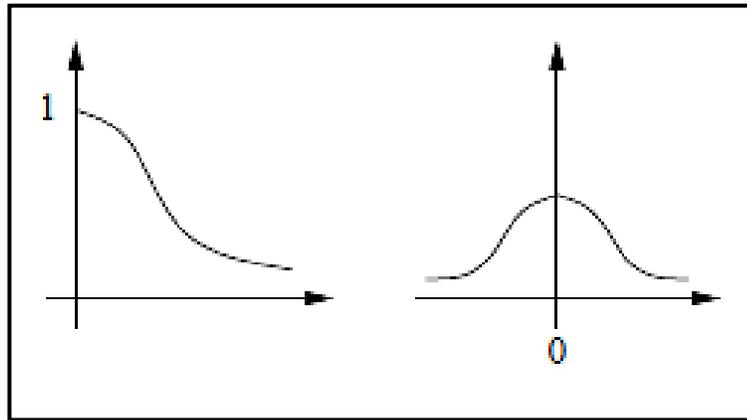


Figure 3.5 Low Pass Filter

3.1.2.1.1 Mean filter

The Average (mean) filter smooths image data, thus eliminating noise. This filter performs spatial filtering on each individual pixel in an image using the gray level values in a square or rectangular window surrounding each pixel.

3.1.2.1.2 Median filter

A median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical pre-processing step to improve the results of later processing (for example, edge detection on an image). It is very widely used in digital image processing, because under certain conditions, it preserves edges while removing noise.

3.1.2.2 High pass filter (HPF)

HPF works to minimize low frequency contents without changing high frequency components in the Fourier transform, Moreover edges and other sharp changes in gray scale linked with high frequency contents where it can get a sharp image by using a high pass filter.

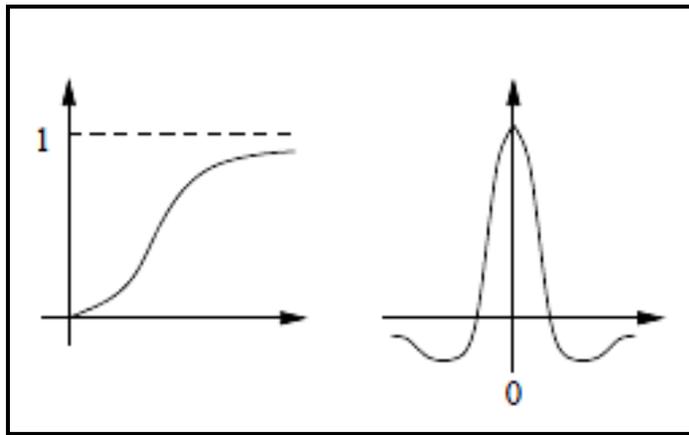


Figure 3.6 High Pass Filter

3.1.2.2.1 Gradient operators

One of the mammography enhancement techniques and includes for instance a sobel gradient operators and unsharp masking, used to reduce an image influence. After using this technique, the output image appears sharper because the high frequency information is amplified while the details in low frequency have been reduced; figure 3.7 shows the result of this technique application [41].

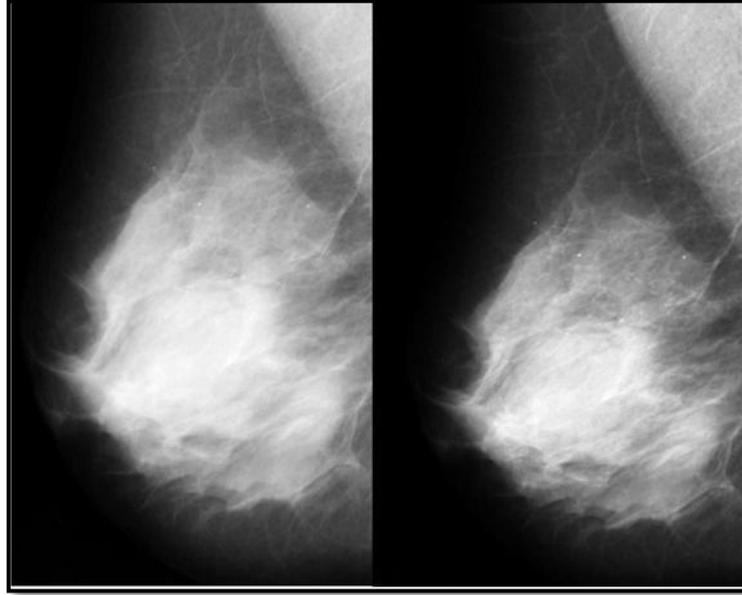


Figure 3.7 Example of enhancement of mammography with Unsharp Masking: Left: Original image; Right: image filtered using an adaptive un-sharp filter

3.2 Image Segmentation

This phase is one of the most complicated and difficult tasks in this field, and the accuracy of any CAD system associated with its success. It refers to extract exact object from the background (Parenchyma). Main goals determining the abnormality (mass, calcifications) locations in the image for helping radiologists in diagnosing and providing abstracted ROI for extracting its features in the next step. Segmentation algorithms are usually based on one of two properties of intensity value, discontinuity based on abrupt changes in the image as edges, and similarity based on the partition of the image into regions according to some similarity criteria [42].

Image segmentation processing has 4 types:

- i. Traditional Techniques (Thresholding Techniques as OTSU method)
- ii. Fuzzy Techniques (Based on the contrast between the normal and abnormal tissues)
- iii. Bilateral Image Subtraction (Based on the asymmetry between both breasts left and right)
- iv. Multiscale Techniques (Discrete Wavelet Transform DWT)

3.2.1 Classical Techniques

Some of these groups are based generally on the histogram, since the suspicious regions are brighter than the normal tissue of the background makes thresholding methods good for using in segmentation.

3.2.1.1 OTSU Method

Used to automatically perform histogram shape-based image thresholding [45].

3.2.1.1.1 Thresholding And Multilevel Thresholding by OTSU [46]

In the case of bi-level thresholding of an image, the pixels are weighted sum of variances of the two classes and distributed into these classes:

$$\sigma_{\omega}^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad (5)$$

Weights ω_i are the probabilities of the two classes separated by a threshold t and σ_i^2 variances of these classes.

$$\sigma_b^2(t) = \sigma^2 - \sigma_{\omega}^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2 \quad (6)$$

The class probability $\omega_1(t)$ is computed from the histogram as t :

$$\omega_1(t) = \sum_0^t P(i) \quad (7)$$

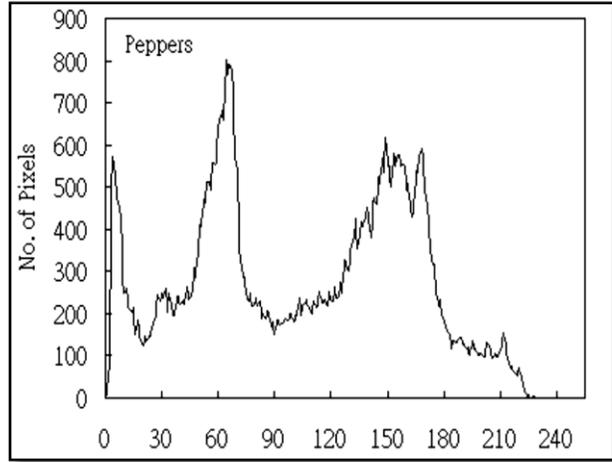
While the class mean $\mu_1(t)$ is:

$$\mu_1(t) = [\sum_0^t P(i)x(i)]/\omega_1 \quad (8)$$

Where $x(i)$ is the value at the center of the i^{th} histogram bin. Similarly, it can compute $\omega_2(t)$ and μ_i on the right-hand side of the histogram for bins greater than t , the class probabilities and class means can be computed iteratively, multi otsu method is the expansion of the original method to segment the image by applying multilevel thresholding (Multilevel thresholding). Figure 3.8 shows the result after applying OTSU method for multilevel thresholding with T=4 and classes = 5.



Original Image



The Histogram



Bi-level



Tri-level



Fourth-level



Fifth-level

Figure 3.8 Result of Multiple OTSU Method [46]

3.2.2 Multiscale approach (Discrete Wavelet Transform)

When digital images are to be viewed or processed at the multiple resolutions, discrete wavelet transform (DWT) is the mathematical tool of choice. In addition to being an efficient, highly intuitive framework for the representation and storage of multi-resolution images, the DWT provides powerful insight into an image's spatial and frequency characteristics. The Fourier transform, on the other hand, reveals only an image frequency attribute. The wavelet transform is based on small waves of variable frequency and limited duration, which are denominated wavelets [42].

A wavelet decomposition of an image divides the input image into several subbands containing features at different scales. The advantage is that small features like microcalcifications are dominant in one subband, whereas larger features like masses prevail in a different subband [43].

There are a lot of wavelet families in following some of them:

3.2.2.1 Daubechies

Ingrid Daubechies, one of the brightest stars in the world of wavelet research, invented what are called compactly supported orthonormal wavelets thus making discrete wavelet analysis practicable (See Figure 3.8).

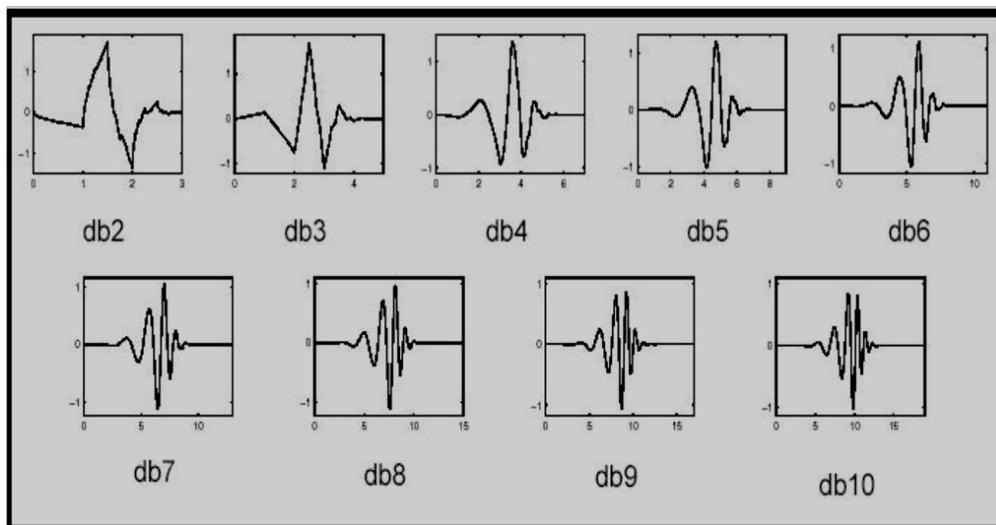


Figure 3.9 Daubechies Families [44]

3.2.2.2 Coiflets

Has built by Ingrid Daubechies at the request of R. Coifman. The wavelet function has $2N$ moments equal to 0 and the scaling function has $2N-1$ moments equal to 0 (See Figure 3.9). The two functions have a support of length $6N-1$

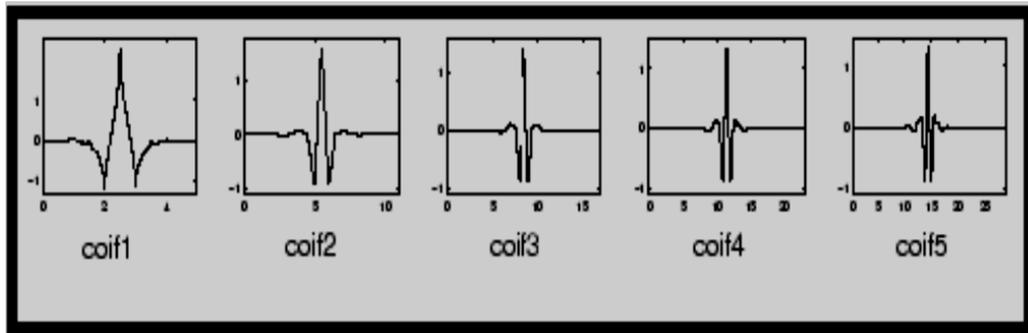


Figure 3.10 Coiflets Families [44]

3.2.2.3 Symlets

The symlets are nearly symmetrical wavelets proposed by Daubechies as modifications to the db family. The properties of the two wavelet families are similar (See Figure 3.10).

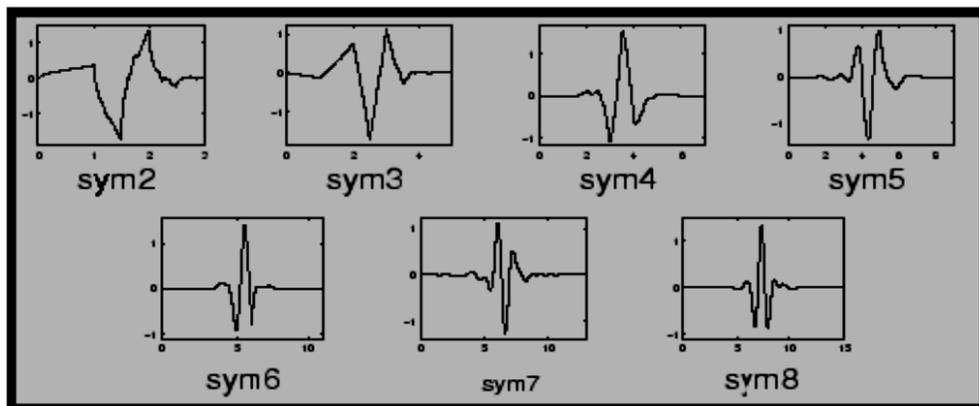


Figure 3.11 Symlet Families [44]

3.3 Features Extraction And Selection

The third phase and the key of optimal performance is extracting and selecting the features which can be obtained by applying mathematical and statistical operations for characteristics of the region of interest (ROI) such as contrast, edges, shape, size, etc.

Features selection, is the task of selecting the optimum features from the features available group after the image segmentation in particular problem[47].

The feature space is massive but might be divided into three main sub- spaces: intensity features, geometric features, and texture features [48].

3.3.1 Intensity features

Among the three subgroups, this group of features is easiest to get them and they have easy statistics. Some of features included in this group contrast measure of ROIs, standard variance, skewness and kurtosis, etc. Mostly, the statistical values of these features are associated with the difference between a rate of the gray scale of the ROIs and a rate of gray scale of the surrounding normal breast tissue.

3.3.2 Shape features

This group is concerned about those features of ROIs which are associated with the shape for that, this group also called geometric or morphological features group.

3.3.3 Textural features

Mainly, the features included within this group are based on ROIs texture, In turn; the features can be divided into three classes:

- 1- Spatial Gray Level Dependence (SGLD) or Gray Co-occurrence Matrices (GCM)
- 2- Gray Level Co-occurrence Matrices (GLDS)
- 3- Run Length Statistics (RLS)

SGLD is a 2-D histogram and an element of the SGLD matrix $P(I, j, d, \theta)$ is defined as the joint probability that the gray levels I and j occur separated by distance d and along direction θ of the image. In order to simplify the computational complexity of the algorithm, the θ is often given as 0° , 45° , 90° , and 135° , and the distance d is often defined as the Manhattan or city block distance [49].

3.4 Classification

3.4.1 Artificial neural network

This concept was inspired by systems of Biological Learning (BLS), where a huge number of neurons are connected interlaced forming a complex network, This net has high ability to process the information and data in part of second efficiently of BLS due to parallelism in processing where distribute the operations over more than one neuron, from here began the concept of ANN neural networks.

ANN simulates BLS in distributing and processing the operations in parallel, In brief, ANN (also called simulated neural network (SNN)) is a group of data structures and software contains a number of processors each has a small amount of knowledge with access to data in its own memory; initially this network must be trained by amount of rules and data.

3.4.2 K-nearest neighbors

Based on known similar samples the classifier (KNN) K-nearest neighbor differentiates unknown patterns by computing distances between every unknown one to every known sample, and then select the nearest sample K . Unknown patterns are assigned to that class which contains the most nearest patterns in the same known sample. For this type of deterministic classifier, it is necessary to have a training set which is not too small, and a good discriminating distance.

3.4.3 Support vector machines (SVM)

This method based on the precept of minimize structural risks, SVM performs better if applied to dataset outside the training dataset. After training, microcalcifications detection is formulated as one the classifications problem called a “supervised-learning” then SVM is applied to improve the detection algorithm. SVM is a promising method for abnormality detection in medical imaging applications, here it used for microcalcifications existence and detect the location in the mammogram image [124].

CHAPTER IV

LITERATURE SURVEY

So far, many experiments and researches have been done to detect breast cancer. This chapter summarizes the previous works which done by other researchers.

4.1 Enhancement

4.1.1 Spatial domain methods

Adin Ramirez Rivera et. al. proposed a "content-aware" algorithm that improves the dark images, detects details in textured areas, makes the edges sharp and protects the softness of flat areas. This method produces an 'ad hoc' transformation for each image, adjusting the mapping functions to each image features to generate the enhancement, then has been analyzed the contrast of the image's border and regions and gathered all the information with common properties. The functions were extracted from these groups model relation within the image. The results were then adaptively mixed to boost the details in the image. The disadvantage with this method that is incapable to improve the darkness into the image area while "image fusion" techniques overcome this limitation and enhances all regions of an image such as mammogram images [74] and that what has been proved by studying for Wanga Z et. al. proposed an image fusion scheme based on the multiple-resolution (MR) framework.

The three signals are first transformed into multiple bands presenting information at various frequencies and then a two-step processing follows:

- 1- Using a novel noise estimation method to enhance the local signal-to-noise ratio.
- 2- Context modeling, then restructured the fused image (which looks similar to traditional mammogram, but with significantly enhanced detail characteristics) by inverse transform [125].

R. Mekanle et. al. introduced a study which has been done in cooperation with radiologists at the Breast Imaging Center in Columbia-Presbyterian Medical Center and the Biomedical Imaging Laboratory of Columbia University where reported on the functional accomplishment of the first ROC study to evaluate the interests of contrast enhancement by means of over complete multiscale expansions of mammograms by using a "dynamic spline wavelet" as the base for multiscale extension and designed an "associated nonlinear sigmoidal" enhancement function [56].

Papadopoulos A. et. al. in their work assessed the impact of image enhancement processing stage and the parameter tuning of a computer-aided detection (CAD) system for microcalcifications detection in mammograms. Five image enhancement algorithms were used: introducing the contrast-limited adaptive histogram equalization (CLAHE), the local range modification (LRM) and the redundant discrete wavelet (RDW) linear stretching and shrinkage algorithms.

Computer Aided Detection (CAD) optimum was targeted to the percentage of the most contrasted pixels and the size of the minimum detectable object which could satisfactorily represent a microcalcification. LRM and wavelet-based linear stretching have achieved the highest rate of performance where LRM achieved ($A(Z)=0.932$) and the wavelet-based linear stretching has achieved ($A(Z)=0.926$)[78].

Rajib Kumar Jha et. al. suggested a nonlinear non-dynamic stochastic resonance-based technique, to enhance the low contrast and dark images. They treated a low contrast as a sub-threshold signal, then they used a noise-enhanced signal processing to improve the image contrast.

The proposed technique has applied uniquely with outer noise to equalize the effect of inner noise of a dark image due to insufficient lighting randomly and repeatedly, the noise was added to the image. Changing the noise intensity, noise induced resonance was obtained at a particular optimum noise intensity. The proposed technique can be applied for four distribution types of noise gaussian, uniform, poisson and gamma. Quantitative valuation of their performances had been done whence of contrast enhancement factor, color enhancement and perceptual quality measure [79].

R.Garg et. al. proposed a method based on the main idea of the histogram equalization enhancement techniques, they compared and analyzed the performance of the techniques contrast limited adaptive histogram equalization (CLAHE), equal area dualistic sub-image histogram equalization (DSIHE), dynamic histogram equalization (DHE) in terms of peak signal to noise ratio (PSNR), absolute mean brightness error (AMBE) and contrast. The result of this work proven that (DSIHE) technique in term of AMBE gave the best results among the rest techniques (Table 4.1) [39].

W.Zhiming et. al. implemented FAHE by adopting three techniques to enhance the quickness of AHE through firstly, acquiring the local histogram by a sliding window with a reiterate approach; Secondly, computing a half of the histogram cumulatively; last, keeping the block size (W^2 equal to the product of integral power of two and (GLN) gray level number) and all the operation of multiplication and division has been changed with the fast bitwise shift, moreover this study can be used with other enhancement algorithms (CLHE, MAHE) [59].

Stephen M. Pizer et. al. proved by their survey study that the advanced techniques which enhance digital images' contrast :(GCE) Global Contrast Enhancement , ACH and CLAHE have shown success to grant marks point to the significance of an exact visual model in the techniques' evolution [55].

| Parameter Technique | AMBE | Contrast | PSNR |
|---------------------|---------|----------|--------|
| CLAHE | 13.8521 | 23.5878 | 0.0366 |
| DSIHE | 4.9081 | 33.8767 | 0.0327 |
| DHE | 13.0886 | 12.1438 | 0.1107 |

Table 4.1 Comparison of Various Parameters [39]

As a result of the foregoing the conclusion can be drawn is that, although the concept of histogram equalization easy to use, fast and some of its techniques like AHE has proved its efficiency in the images displayed from a wide range of imaging methods such conventional technique, radiotherapy portal films and magnetic resonance imaging (MRI) [70][71], it has some of disadvantages may impede getting the desired result in some cases : if the histogram covers almost the full gray scale, if there are gray values that physically far apart from each other in the image and using LAHE techniques consumes time due to that techniques requires a high number of operations per pixel.

4.1.2 Frequency Domain Methods

Harish Kumar et. al. proposed a new enhancement technique to enhance digital mammogram images, the idea is based on biorthogonal wavelet transform and modified mathematical morphology. In this method the model of enhancement takes the mammogram image $M(x, y)$ as input image then separate it into low frequency and high frequency components by using Gaussian low pass filter for a high degree of control over all dynamic range. Modified mathematical morphology is applied to get a low pass filtered image $LI(x, y)$. A high pass filtered image $HI(x, y)$ contains the noise and edges information.

For enhancing the edge information in the high pass filtered image, one of the edge enhancement algorithms is used. To get the contrast enhanced Image $C(x, y)$ morphologically processed image $M(x, y)$ and Edge Enhanced image $EE(x, y)$ should be assembled. Wavelet transform technique is applied to suppress the noise.

Wavelet transform composed three operations: a) wavelet reconstruction, b) wavelet decomposition and 3) thresholding detail coefficients (applying dependent threshold). Finally, approximate and modify the detail coefficients $E(x, y)$ is done to reconstruct the decomposed image [75].

J. Scharcanski et. al. Proposed enhancement method and noise lessening which achieved very good results for improvement the contrast as well as denoising in mammogram images. This method firstly requires processing the mammogram image to enhance its local contrast and then transforming the pre-processed image into wavelet

domain to enhance the image edges and remove the noise. This method is somehow complex due to complexity working in the wavelet domain [76].

A new sharpness enhancement model has proposed by Seung-Won Jung et. al. for stereo images where they introduced a novel application of the "BJND model " to enhance the stereo images sharpness. They proposed an efficient solution to reduce the enhancement problems in the sharpness enhancement of stereo images. The proposed solution is optimizing the framework with additional constraint terms to remove the unnecessary excess values of luminance. In addition, the reliability of this model was taken into account by estimating the accuracy of stereo matching. "BJND model "not applicable to enhance the contrast in the colored images [80].

Xiaoying Fang et. al. proposed a method to develop the enhancement result while using "image fusion method" with evaluation on sharpness. They discussed various evaluation methods and fusion policies, then they compared the problem with the method, which evaluates the sharpness as a local statistical feature [81].

Leiner, B.J. et. al. introduced a method for microcalcifications detection in digital mammogram images by using 2D discrete wavelet transform which is an effective technique used for enhancing mammogram images and achieves satisfactory results with lesions detection. Besides using discrete wavelet transform, image enhancement methods were used to eliminate the noise as well as to improve a contrast. Firstly, segmentation algorithms were used to remove some unnecessary objects in the image which are not useful for interpretation the mammogram image. Then, using an unsharp masking filter to enhance the image contrast and make some detail like microcalcifications more clarity. Likewise, using histogram modification techniques to get better visualization of the microcalcifications and finally discrete wavelet transform was applied for abnormality detection [85] [126].

4.1.3 Mixed methods

Two different methods for image enhancement were employed by H. Stefanou et. al. then compared these methods in order to determine which method offers better results in each mammogram image. The first method is called adaptive enhancement algorithm measures the relation between wavelet parameters in analysis at sequential levels to decide if the parameter was derived from noise or not. Non-linear mapping used to enhance the image. The second one, which is called typical method for image enhancement, it works by using thresholding for noise reduction process, then non-linear mapping for image enhancement and finally filtering for blur reduction and make the mammogram sharper. This work observed that the typical method introduced better results when the analysis process is made at the first level. Due to the adaptive algorithm does not blur the image, its results are better when analysis process is made at higher levels [73].

Khairunnisa Hasikin proposed a fuzzy grayscale enhancement technique for low contrast image. This technique was proposed by maximizing fuzzy measures contained in the image. The membership function was then modified to enhance the image by using power-law transformation and saturation operator [82].

A new approach has proposed by Zailani Mohd Nordin et. al. this approach deals with contrast enhancement process for mammogram images. The proposed approach generates group of multiple different images Instead of generating just an one image where each one of these images group corresponds to various range of the level intensities, then these images are grouped together to form moving video frames of which would show smooth transmission from one level of intensity to another. As a result, the radiologists who are implementing visual examination of the image can sweep through the entire range of enhanced contrast image intensity level [84].

J. Jiang et. al. proposed a method combines a fuzzy logic and structure tensor in order to enhance microcalcifications contrast. A structure tensor operator firstly was produced and then applied to all the pixels of the mammogram image separately, this process resulted an "eigenimage". Eigenimage has used to merge with fuzzy image that resulted by a fuzzy transform from the input image to improve the image contrast. Although there is an advantage of this method that it can remove non-microcalcifications' regions while improving the microcalcifications regions, using it is complicated due to ambiguity and the need for experience to work with two different domains [123].

4.2 Segmentation

4.2.1 Thresholding techniques

Younesi F et. al. Introduced a CAD scheme with two parallel models: fuzzy entropy feature and adaptive thresholding method. The algorithm contains the following steps:

a) Preprocessing the digital mammogram image by identifying the region of interest (ROI) as a candidate for abnormality through extracted the breast region , b) Enhancing the image by using linear transformation and subtracting enhanced from the input image, c) Characterizing of the ROI by extracting the fuzzy entropy feature, d) Using a local adaptive thresholding for abnormality regions segmentation , e) To detect the abnormality using the combination of the last two parallel models. The introduced method achieved 90.73% as sensitivity and specificity of 89.17% has been proved that the behavior of local adaptive thresholding and fuzzy entropy technique could be a useful method for abnormality detection in digital mammogram images [90].

Dominguez, A.R. et. al. proposed a method with a following scheme:

(1) Preprocessing the original image in order to increase the signal-to-noise ratio to detect the abnormality, (2) Segment the prospective abnormal region, and (3) Lessening the false-positive rate.

Region segmentation is applied after using morphological operations, local scaling and wavelet decomposition and reconstruction via transforming the image multiple threshold levels to binary image. A set of features should compute from each of the segmented areas.

In this work 57 mammogram images were tested contain breast masses from MIAS data set, including speculated, circumscribed, and ill-defined masses. The result of this method has reached 80% sensitivity rate at 2.3 as a false-positives per image [91].

4.2.2 Region based techniques

Morphological operators and fuzzy c- means clustering (FCM) algorithm have been used for intensity – based segmentation and it was done by BASHA S. This algorithm used to separate cancerous region from the fatty breast regions. Using the proposed method is helpful in clustering medical diagnosis at early stage and has showed good results. As the number of clusters increases, more and more information is obtained about the tissue which cannot be identified by pathologists [89].

N. A. ISA et. al. presented a mass detection and segmentation system in mammogram images; the system in the beginning separates the breast from the background then improves the contrast of tissues. The basis of the work is applying constraint region growing based on local statistical texture analysis in order to detect the mass and segment it out. The results were showed that the proposed technique has 94.59% as sensitivity and the number of false positive is 3.90 per image [62].

Jue Wu et. al. proposed a new method using compound "MRFs" based on a general boundary model. The main goal of this method is using concept of the relationship between labeling and boundaries to enhance the performance of segmentation. The experiments and comparisons of proposed model with other existing MRF methods have proved that the proposed model showed good and satisfactory segmentation results in high noise regions [88].

4.2.3 Edge detection techniques

Indra Kanta Maitra et. al. They proposed using " Binary Homogeneity Enhancement Algorithm (BHEA), followed by Edge Detection Algorithm (EDA), Breast Boundary Detection Algorithm (BBDA), Pectoral Muscle Detection Algorithm (PMDA), and Anatomical Segmentation of Breast ROI (ASB) ", and then using Seeded Region Growing Algorithm (SRGA) which is a modified algorithm of the seeded region growing method for each region segmentation (coloring objects boundaries) in the mammogram. In other words, each region is bounded by a single pixel boundary as obtained during the edge detection process followed by the process of anatomical segmentation. This method has registered accuracy rate around 99.87% [77].

H. B. Kekre et. al. Proposed extended Sobel, Prewitt and Kirsch edge operators to locate and segment tumor from the mammography images. this method can detect the location and edges of the tumor clearly better than using Sobel, Prewitt and Kirsch edge operators, as well as , using proposed extended 5x5 Sobel operator with mammography images gives better segmentation than extended Kirsch and Prewitt operator [86].

A gradient vector flow field (GVF) is a new method has been introduced by Zou et. al. is a parametric deformable contour model. This method uses an adaptive histogram equalization for mammogram image enhancement and followed by gradient vector flow field (GVF) component with a larger entropy to segment the region of interest (ROI). In the Figure 4.1 shows an example of results after using GVF with applying and without applying enhancement technique [87].

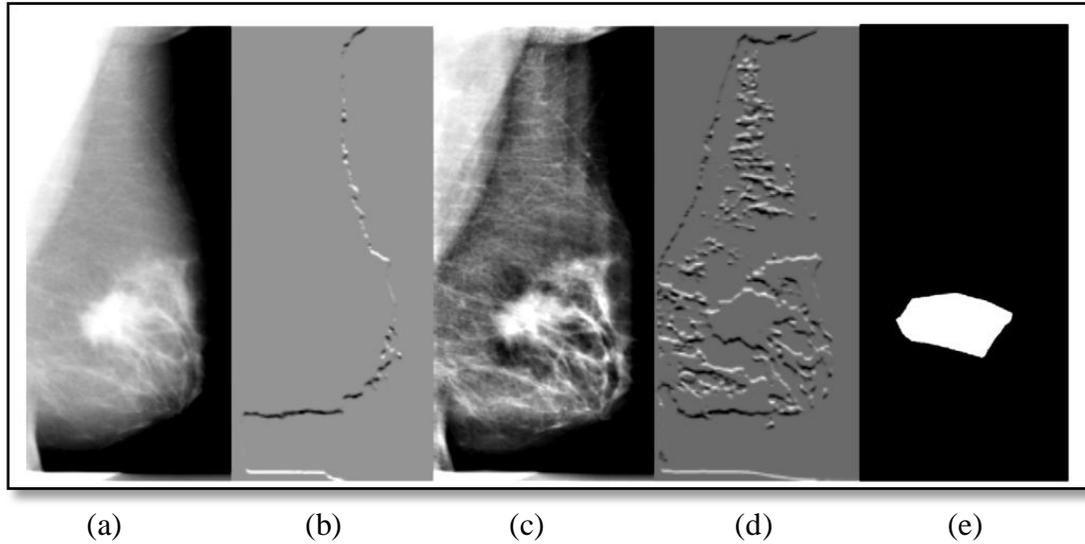


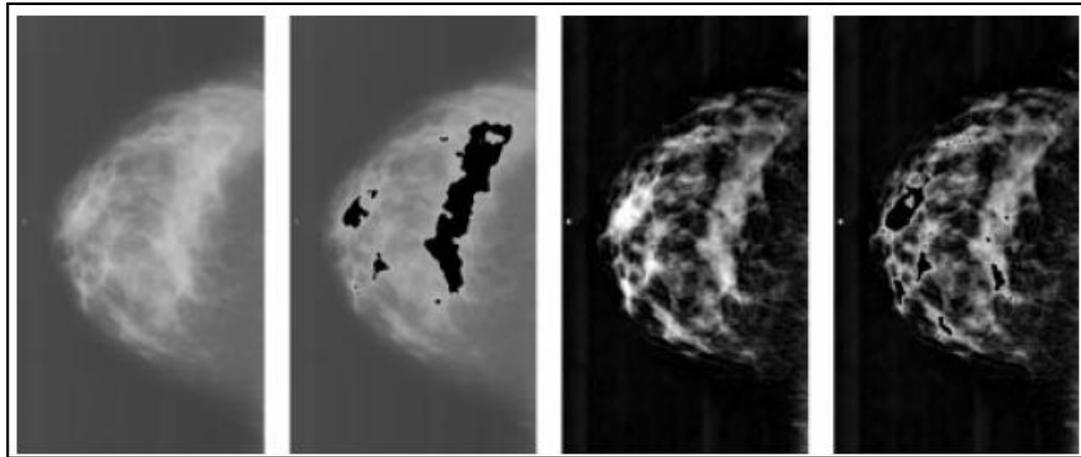
Figure 4.1 (a) Original Image, (b) The Horizontal GVF , (c) Resulted image after applying AHE, (d) Horizontal GVF, (e) Generated mass mask.

4.2.4 Hybrid techniques

Li et. al. employed a statistical method using with enhanced segmentation and extraction of abnormalities regions this method is “a finite generalized Gaussian mixture” (FGGM) distribution. FGGM distribution was used to model mammogram images and basically, it relies on two information theoretic criteria with each other to define the best number of image regions. Finally, a "contextual Bayesian relaxation labeling" (CBRL) technique was used to perform the selection of the suspected tumors. Figure 4.2 shows the result of this work [111].

Ball and Bruce suspected masses were segmented in polar field where "adaptive level set segmentation method" (ALSSM) was utilized to adjust the edge threshold at each angle in order to supply an accurate results of the segmentation [112]. Then, this work was extended in [124] where "speculation segmentation with level sets" (SSLS) has been used for the speculated masses Identification, detection and segmentation. In the same time "Dixon and Taylor line operator" (DTLO) and "a generalized version of DTLO" (GDTLO) were used with level set segmentation method.

The great advantage of hybrid segmentation is that any segments that are discovered are automatically “identified”, since they come with a “built-in” demographic profile.



(a)

(b)

(c)

(d)

Figure 4.2 (a) Original Image, (b) Segmented Image , (c) Enhanced Image, (d) Segmented Image Based on Enhanced Image

4.3 Feature Extraction and Selection

M.Sameti et. al. applied a study represented in collecting mammography image to 58 biopsies indicate to breast cancer, in each mammogram were taken 10 -18 months earlier than breast cancer detection to evaluate them. Marking two regions first one corresponded to a malignant mass area were developed later, the second indicates the same region in the first region but on the normal breast, then 62 features of texture and photometric have been calculated. This method can detect breast cancer by nearly 72 % [68].

Varela et. al. in this work the features were extracted belong to the iris filter output, in conjunction with texture, edge-related, morphological features and the gray level. The best performance was achieved with the integration of seven features. These features precisely are: The maximum average of iris filter result, the medium value of the gray level for segmented area, the mean enhanced filter output value, size, location, isodense, compactness and eccentricity [92].

In Yuan et. al. study three sets of features were used. The first set containing features that distinguish speculated masses, edge, appearance and contrast of tumor. The second set containing texture features and the last set containing a distance feature which was calculated as a Euclidean distance from the breast nipple to the center of the mass. To select a subset features for the classification phase, a linear stepwise feature selection method with a Wilks lambda criterion was utilized [93].

A temporal feature group was proposed by Timp and Karssemeijer, this group contained a complete group of single vision features together with temporal features. They extracted 38 individual view features and they have been compiled into 12 main classes depending on characteristic type they represent. The features from the current mammogram image are put together with the features calculated from the identical area in mammogram images taken in previous tests to offer temporal information. Temporal features are acquired by subtracting previously from current feature values which produced in 29 the temporal features. Sequential forward floating selection (SFFS) is used to choose the best subset from the feature group [94].

Timp et. al. [95] they proposed two types of the temporal features similarity features and difference features.

Kim and Yoon evaluated recursive feature removal-based support vector machines (SVM-RFE) in order to improve classification accuracy. SVM-RFE combines a recursive elimination method in feature selection to get the features ranking that are mostly significant to (SVMs) and the topmost rank features are selected to be used in classification. SVM-RFE has detected that using only gray level features from a subset of the 22 BI-RADS and simplified increased computer aided detection (CAD) accuracy compared to using all the 22 features [96].

Nandi et al. this study proposed using five techniques to select stand-alone features: Kullback- Leibler divergence (KLD), Kolmogorov-Smirnov test (K-S test), Students t test (t test), sequential forward selection (SFS) and sequential backward selection (SBS) to make the space of features more cramped and prepare it for classification phase. They found that the most significant feature for the designed classifier was the measure of shape of fractional concavity [97].

Hupse and Karssemeijer, their work is based on comparison between two techniques in feature selection standards, these standards "Wilks lambda" and "the mean sensitivity of the FROC (free response operating characteristic) curve", both standards were used with and without normalize the features. This study proved that the used selection technique that performed well was the technique in which Wilks lambda was used as selection standard in integration with features normalization [98].

4.4 Classification

4.4.1 Neural networks approache

Varela et. al. integrated the feature group into a backpropagation neural network (BNN) classifier to minimize the number of false positives (FP). The results registered two results. When the evaluation was based on lesion the rate sensitivity of false positive was 88% and was 94% at false positive equals 1.02 results per image when the evaluation based on the case [92].

Li et. al. they used a "Bayesian artificial neural network (BANN) classifier" to integrate the selected features, then use the classifier to get an estimation of the eventuality of malignancy. Important results of integrated features performance have been shown compared with the individual features when discrimination between malignant and benign tumors. The findings of this method were 0.83 as an A value under the ROC curve and 0.02 as a standard error [99].

Fauci et. al. designed a classifier by using an artificial neural network (ANN). The classifier was designed by 12 input neurons, changing number of the hidden neurons to obtain the best classification performance and one output neuron. The output neuron provides the probability incidence of breast cancer. The neural network was adopted on a feed-forward back-propagation supervised network trained with gradient descent learning rule with momentum. Momentum represents a sort of inertia, which is added to quickly move along the direction of decreasing gradient, thus avoiding oscillations around secondary minima. The results of this work were: an A=0.85 with standard error of 0.08 [100].

4.4.2 Linear and non-linear discriminant analysis

Ball and Bruce generalized discriminant analysis (GDA) was used to analyze the features vector in order to provide a non linear classification and to classify the tumor as a spiculated mass or not. Classification the selected features were classified as benign or malignant using two classifiers k-nearest neighbor (k-NN) and maximum likelihood (ML). This study showed that the k-NN classifier outperformed the ML classifier slightly in fewer numbers of false negatives and the higher overall accuracy. Using classifiers 1-NN or 2-NN has been achieved 93% overall accuracy with 3 FP and 1FN, while the use of ML classifier achieved 92% overall accuracy with three FP and two FN [101].

Nandi et. al. presented an adaptive genetic programming for masses classification. The performance of the genetic programming classifier was good in the discriminating between malignant and benign masses where registered 99.5% an accuracy in training and 98% in testing [97].

A 2-plane learning method for binary classification was proposed by Mu et. al. it is also called a strict 2-surface proximal (S2SP) classifier. The proposed classifier S2SP was proposed to classify both linear and nonlinear patterns. The S2SP classifier ameliorates the performance accuracy for distinguishing between malignant and benign masses. This classifier based on the features using conventional pattern recognition techniques and provides weak performance. The performance of the linear classification was yielded $A=0.97$ and nonlinear classification $A=1.0$ [102].

4.4.3 Support Vector Machine

Leonardo et. al. presented a methodology registered 85% of accuracy of mass detection in mammography images by using K-means algorithm to segment the interesting region and co-occurrence matrix to extract its features, to classify this structure SVM was used to separate it as one of two classes depending on shape and texture [67]. Yoshida et. al. applied a supervised learning techniques with wavelet transform to detect clustered MCs in mammography images [60][61].

Yongyi et. al. achieved 94% as sensitivity by a method based on using “Support Vector Machine” (SVM) for microcalcifications detection in mammography images [65]. Netsch described MCs as shiny dots of certain contrast and volume by using mathematical models with a multiscale analysis based on the (LGF) “Laplacian-of-Gaussian Filter” this proposed detection scheme has the ability to de-clustered microcalcifications effectively[66].

A new methodology combines the use of Independent Component Analysis (ICA) along with Support Vector Machine (SVM) and Linear Discriminate Analysis (LDA) was proposed by D. D. Costa et. al. to differentiate between benign or malignant mass or non-mass breast tissues in mammogram images. As a result, LDA was reached 89.5 % of accuracy to discriminate mass or non-mass and 95.2% to differentiate between benign or malignant in DDSM data set.

Applying the proposed methodology on MIAS database obtained 85 % to distinguish between mass or non-mass and 88% to differentiate malignant or benign tissues; in SVM the rate of accuracy reached 99.6% in discrimination between mass or non-mass and to discriminate malignant and benign tissues was registered 99.5% in DDSM data set and 97% was registered while using MIAS data set to differentiate mass or non-mass, and the rate of the accuracy was obtained 100% in discrimination between benign and malignant tissues [121].

4.4.4 Other methods

A fuzzy binary decision tree (FBDT) technique was used by Li et. al., this technique is based on density-related features in a group of radiographies. In classification the region of interest (ROIs) was classified as normal or abnormal, the results have proved that this technique might be partly effective with high accuracy for tumors with size less than (10 mm) which are not easily distinguishable in mammogram images. This method has achieved 90% sensitivity with 2 false positive [103].

A multiple-instance learning (MIL) was proposed by Krishnapuram et. al. which is an automatically selection algorithm. The proposed algorithm selects small sets of useful features to breast cancer diagnosis. This work proved that the proposed algorithm is more accurate in use compared to using the support vector machine (SVM) classifier [104].

The effects of classifier diversity on the popularization accuracy of the ensemble were investigated by West et. al., the results showed that the best performance improvement has occurred with ensembles produced by forming 3-5 different classifiers. The most effective ensembles formed in this research resulted from a selected small subset of available classifiers, with prospect candidates' specific using both of the following together:

- 1- Considering the properties of error in classifier generalization,
- 2- The sensitivity of classifier and
- 3- Classifier independence in making decisions proportional with other ensemble members [105].

4.5 Proposed Detection Methods

Sankar and Thomas proposed a method for microcalcifications detection using a fractal modeling of mammograms based on both mean and variance. In this work, 28 mammograms were tested from the MIAS dataset and resulted the following: true positive rate (TPR) =82% with an average of false positive = 0.214 per image [106].

A semiautomatic segmentation technique presents some interaction between radiologists. In microcalcifications detection this method uses "Daubechies 6 wavelets" as a main element of the system, where the focus is on components detection that have higher or lower spatial frequency than the calcifications which resulting a pass-band filtered of the original image. Normally, calcifications segmentation is done by applying a threshold on that filtered image, but instead of this traditional approach, the proposed technique allows an interactive segmentation stage can be done by the radiologist him/her self. Results of the fully automatic and the semiautomatic segmentation when using images from DDSM that evaluated by reading the area under the ROC curve [108] gave outcome of $A=0.80$ and $A=0.84$ for semiautomatic, which noticeably higher [107].

Another microcalcifications detection method has presented and tested by using DDSM data set [109], this method is a new and different. The first step, extracting the areas that potentially might be microcalcifications by analyzing the distribution of high contrast and bright regions over the mammogram image; then identify the clusters as regions of interest (ROIs). The last step is restoring the information that might have been lost in the previous steps.

Thangavel and Karnan proposed an algorithm called "Meta-heuristic" uses the meta-heuristic methods such as "Ant Colony Optimization" (ACO) and Genetic Algorithm (GA) to identify the suspicious regions in mammogram images. This method relies on the bilateral asymmetry. If there is not a structural similarity between of both breasts left and right, the possibility for microcalcifications existence is high. Bilateral subtraction is used for determining structural asymmetry. In the first stage, using median filter to enhance the mammogram images, cutting the pectoral muscle part and detecting breast border. Genetic algorithm should apply for enhancing the detected breast border.

The next process is to modify the image alignment by using the border points and nipple position. After that extracting the suspicious region by subtracting the images. These step and algorithms were tested on the whole MIAS dataset (322 mammogram images for both sides breast which means 161 patients). The proposed algorithm results are shown by using FROC (Free-Response Receiver Operating Characteristics) curve where this method achieved $A=0.94$ area under the curve and this considered satisfactory results for a complete automatic detection system [110].

Leonardo de Oliveira Martins et. al. presented a methodology for masses detection on digitized mammograms based on three steps: Images Database formation; structures segmentation using K-means and texture descriptors.

The structures of classification are accomplished by support vector machines (SVM), which divided into two groups through describe the appearance of suspicious areas and texture as a mass or not. This methodology reaches to 85% of performance accuracy [113].

Eltonsy et. al. showed a technique for the automated detection of malignant masses in screening mammography. The technique is based on the presence of concentric layers surrounding a focal area with suspicious morphological characteristics and low relative incidence in the breast region. Malignant masses were detected with 92%, 88%, and 81% sensitivity of 5.4, 2.4, and 0.6 false positive per image [114].

Rojas et. al. explores the use of characterization features based on breast-mass contours that are obtained by automated segmentation. The method presented low sensitivity and high specificity values (e.g. 0.6 and 0.8, respectively) since the positive class in the classification experiments is the set of malignant masses [115].

Belloti et. al. presented an automatic computational tool for mass detection. The authors used an edge based segmentation algorithm in order to separate suspicious area, then order measures obtained from co-occurrence matrices were used to describe the texture of each region. Finally, artificial neural networks technique was used in classification process. The results showed a sensitivity of 80% and 4.23 false positives per image [116].

Chan et. al. proposes a CAD system based in 3D mammograms obtained from a tomosynthesis process. The area under the curve (ROC curve) was 0.91, with standard deviation of 0.03. The CAD tool obtained a sensitivity of 85%, with 1.32 false positives per image [117].

Martins et. al. presented a mass detection method that use Growing Neural Gas algorithm to perform the segmentation step. For each segmented region, shape measures are computed in order to discard bad mass candidates and texture measures are obtained from Ripley's K function. The classification step is performed using SVM classifier. The method provides an accuracy rate of 89.30% [118].

Campanini et. al. presented an approach for detection of masses in digital mammograms that reach sensitivity around 80%, with a false-positive rate of 1.1 marks per image. The work used images coming from the DDSM database [119].

Martins et. al. the rate of their work reached about 86.85% in regions classification of mammograms into benign, malignant and normal using a Bayesian Neural Network [120].

L. F. A. Campos et. al. proposed a method for discrimination and classification of mammograms with benign, malignant and normal tissues using independent component analysis and multilayer neural networks. The best performance was obtained with probabilistic neural networks, resulting in a success rate of 97.3%, 100% of specificity and 96% of sensitivity [122].

Two automatic detection system introduced by Jun Wei et. al., the systems have the same architectures that contain five stages. For FFDMs, the first stage is using a multiresolution preprocessing method for transforming and enhancing the input raw image. For digitizing SFMs, smoothing the input image and sub divided to size of $(100 \times 100) \mu\text{m}$ pixel. For both systems, the pre-processed mammogram image undergoes two operations gradient field analysis then clustering-based region growing in order to identify the abnormal tissues. Both structures were repeated in for image segmentation.

Then for each detected region, texture and morphological features were extracted. Rule-based and linear discriminate analyses (LDA) are two classifiers that have been designed to differentiate masses tissue from normal breast tissues, then two data sets were collected, the first data set contained 131 (104 benign masses and 27 for malignant masses) abnormal cases with 131 biopsy proven breast masses images while the other data set contained 98 cases for normal breast images. Experienced MQSA radiologist identified the exact mass location in the mammogram image. From 0 to 118 days were the interval time between the identical and the SFMFFDM [127].

CHAPTER V

APPLIED METHODS AND PROPOSED SYSTEM

5.1 Introduction

This work aims to build an accurate automatic system to determine the interested region and classify it as normal or abnormal in mammograms. As mentioned previously, any CAD system passes through several stages and many techniques can be used in each stage, these techniques play effective role in achieving the accuracy of the system, so the proposed system basically concentrate on enhancement stage which in turn affects on the result of the segmentation and on extracting the abnormalities' features.

5.2 Proposed System

This thesis mainly introduces a system to detect abnormality in digital mammographic images using image enhancement techniques for eliminating noise as well as to obtain a better image contrast in order to segment the suspicious region and figure out its features to train and test a classifier for discrimination between normal and abnormal regions in the breast.

This section explains in detail the stages of this system techniques and used methods of detection and diagnosis ambiguous regions in the breast's digital mammography image. As displays in the figure below (Figure 5.1) the mammographic image is subject to numerous processing actions in order to classify the lesion as a normal or an abnormal.

In the first stage is applying morphological techniques for removing extra objects in the image such as (imaging problem like light problem, labels and chest muscle) and then unsharp filters, histogram modification methods and other algorithms have been used to enhance the edges and the contents of the breast to be ready for using in the segmentation process. Image segmentation plays a very effective role to maximize the rate of the system accuracy,

In this step multi threading and region growing were used to detect abnormality. After that, extracting the ROI's features should be done to be used in the classification process.

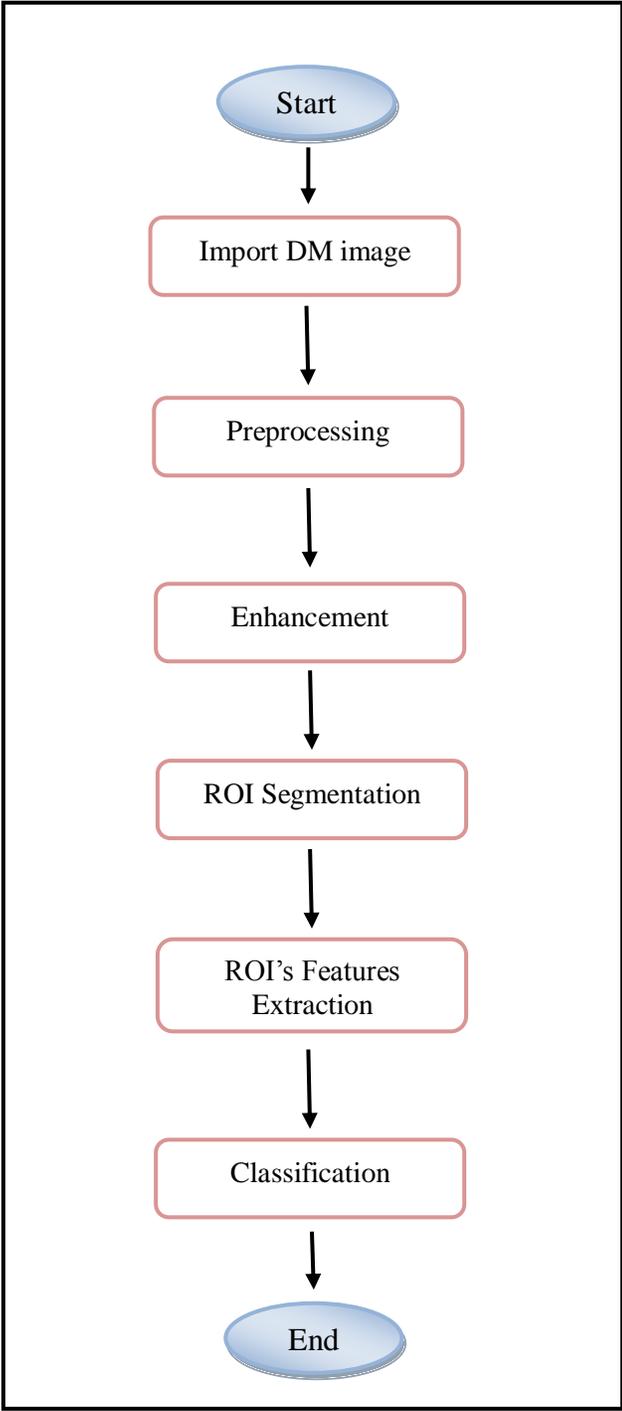


Figure 5.1 Flowchart Digram of The System

5.2.1 Preprocessing

The goal of this step is decreasing image objects as well as enhancing the breast's edges before extracting (ROI) from the original mammographic image. Figure 5.2 shows the sub steps on preprocessing stage.

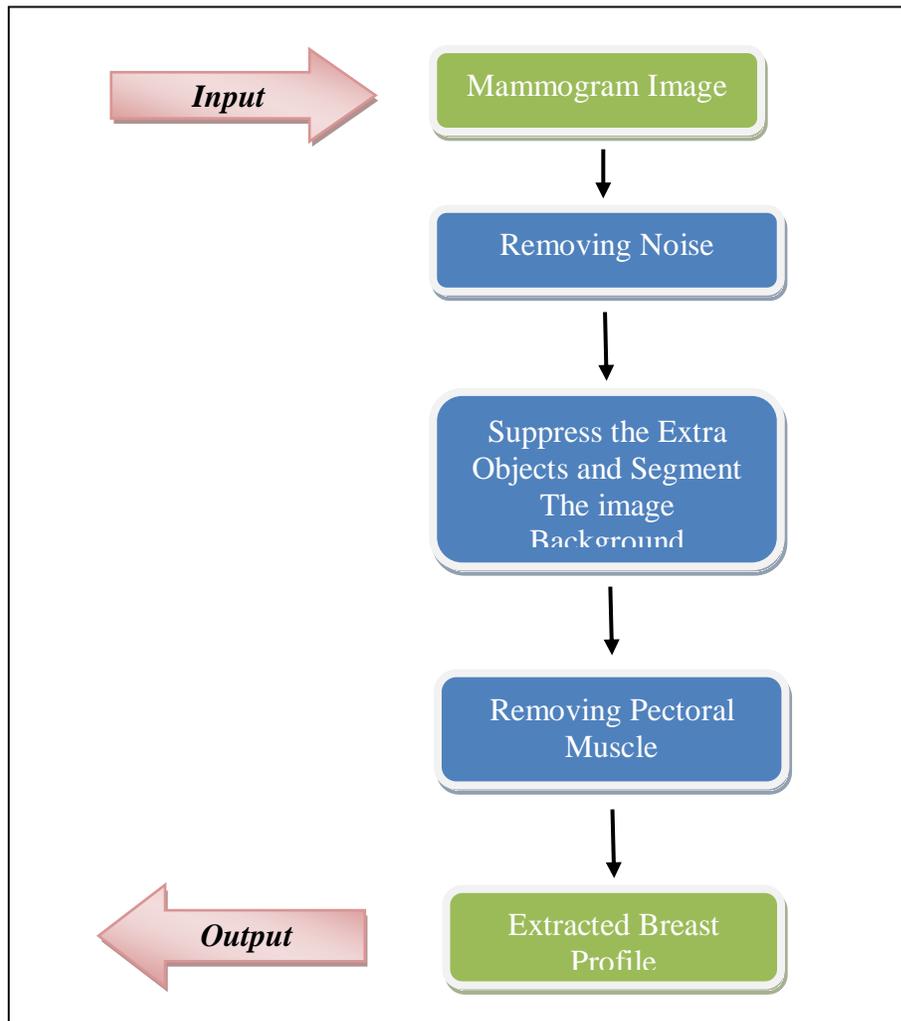


Figure 5.2 Sub steps of Preprocessing Process

The method, suppressing the noises such as straight lines caused by light disorders while taking the image or patient's movement can be removed by applying the 2D median filtering approach 3×3 neighborhood around the corresponding pixel in the original image (Figure 5.3) shows an original screen image with light problems shown as lines.

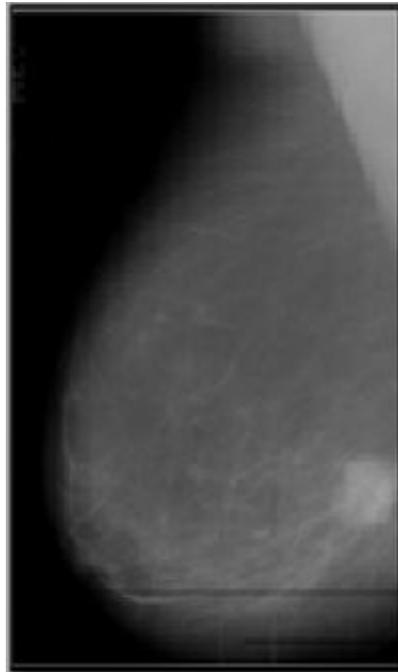


Figure 5.3 Mammographic Image with Light Problem

Then in order to remove the extra objects from the mammography images like wedges and labels, thresholding and morphological techniques have been used, by using global threshold to transform the image into binary $[0, 1]$ format then morphological operations (opening, closing, erosion and dilation) were used on the threshold images to enhance it; to suppress labels and wedges the actual number of each object in the binary image is calculated where the largest number represents the breast profile, after removing all the objects except the largest one.

After getting the final result of step b, the last substeps in the preprocessing stage is removing the pectoral muscle by using geometrical operations; firstly, the direction of the breast in the image was determined then enhances the image's contrast by using `imadjust` and `stretchlim`.

Finally, to transform the enhanced image into [0 1] image then by scanning the image horizontally and stop at the first change in the pixel's value from 1 to 0 and keeping the coordinate of this pixel; the same procedure have done vertically then link these points together by straight lines to form a triangle, then extract the corresponding part of the chest muscle in the original image. The achieved success of this phase was almost 100% with all mammograms.

5.2.2 Enhancement

In this work more than one image enhancement technique were used, the first part of this section explains the main enhancement techniques which have been used with all the images to get an improved image (especially, suspected part that may contain the lesion)in order to facilitate the implementation of segmentation algorithms.

The rest of enhancement techniques which can be used manually according to the image's problem which can be observed by human vision system were explained in the second part.

5.2.2.1 The main enhancement techniques

5.2.2.1.1 Unsharp masking

Laplacian linear high pass filter is employed here. This method is simple and effective in using to enhance the quality of digital mammography image also it is widely-used for conservative high frequency components and enhances the edges; however it is sensitive to the noise.

$$g1(x,y) = g(x,y) + \alpha f1(x,y) \quad (9)$$

Where $g_1(x, y)$ is output image, $g(x, y)$ is the input image (output of preprocessing the previous step), α is the factor of improvement that controls the level of improvement of contrast and $f_1(x, y)$ is the filtered image by linear high pass filter.

The following equation described this process:

$$f_1(x, y) = 4g(x, y) - g(x - 1, y) - g(x + 1, y) - g(x, y - 1) - g(x, y + 1) \quad (10)$$

Equation (11) can be implemented by the involution masks in the following:

| | | |
|----------|----------|----------|
| 0 | 1 | 0 |
| 1 | 4 | 1 |
| 0 | 1 | 0 |

The alternative implementation of the Laplacian takes into account all the 8 neighbors of the center pixel in the input image, and it may be applied by the masks below:

| | | |
|-----------|-----------|-----------|
| 1 | -2 | 1 |
| -2 | 4 | -2 |
| 1 | -2 | 1 |

Another expression for Laplacian mask can be used as shown below:

| | | |
|----|----|----|
| -1 | -1 | -1 |
| -1 | 8 | -1 |
| -1 | -1 | -1 |

In this system the previous three types of Laplacian masks have been used.

5.2.2.1.2 Histogram modification:

Image histogram displays the relative frequency of occurrence of different gray levels in the image. Histogram modification methods are favored due to their speed and easiness. A transformation function is nonlinear and first derived from a desired histogram and the histogram of an input image. (For more details return to Chapter IV)

In this stage were used two approaches for histogram modification:

5.2.2.1.2.1 Exponential modification approach

The following formula describes this concept:

$$g_2(x, y) = \begin{cases} e^{\frac{g_1(x,y)}{m-c}} & \text{if } g_1(x,y) \geq T \\ e^{\frac{g_1(x,y)}{m+c}} & \text{if } g_1(x,y) < T \end{cases} \quad (11)$$

Where g_2 is output image, g_1 is the resulted image after using unsharp mask stage, T is a threshold, c is a positive integer and m represents the mean values in the mammography image.

5.2.2.1.2.1 Linear Modification Approach

Which described by:

$$g_2(x,y) = \begin{cases} g_1(x,y) & \text{if } g_1(x,y) \geq T \\ g_1(x,y) - c & \text{if } g_1(x,y) < T \end{cases} \quad (12)$$

Where g_2 is an output image, g_1 is an image from the unsharp mask stage, T is a threshold and c is a positive integer.

5.2.2.2 Enhancement techniques manually used

These options can be used in case the digital image has observation problems such as darkness, noise or other problem, the options include some techniques to modify contrast and eliminate noise. These techniques in the following:

- a. Contrast stretching.
- b. Local enhancement.
- c. Negative.
- d. Custom equalization: this option allows the user to insert the appropriate parameter value for image equalization process.

5.2.3 Segmentation

This stage involves using two techniques to segment (ROI) that includes a lesion and extract it from the enhanced image which obtained from the previous stage.

5.2.3.1 Seed-based region growing (SBRG)

CRG is a region based technique to segment the image and it is iterated on, where the main idea in this approach is selecting an initial point called “seed” which is a vital step for the segmentation result; then examines adjoining pixels of this seed, whether they have the same characteristics or not, in case the neighbors are matching with the seed then they be added to the region and so on till reaching to the border of the region where the pixels have a different frequency.

Since lesions are mostly denser than normal surrounding tissue, In this work, the image should be prepared to make the seed selection process easier before applying SBRG to segment the suspicious region by using one of the previously mentioned unsharp masks then using one of the histogram modification approaches to magnify the contrast between the ROI and the natural tissue to.

This technique achieved satisfactory results with all mammograms especially, those contain masses. When SBRG was applied 25 images were used this technique extracted masses from 23 out of them which means 92 % success rate.

5.2.3.2 Multi thresholding OTSU’s method

This method is the mechanism used for the extraction of ROI through the adoption of more than one threshold, and thus dividing the image into several classes, where often ROI belongs to the class value-255. Thus, the work of this technique depends on the nature of the lesion contrast. So, applying this method to segment the abnormality requires careful in choosing the number of the classes.

Using multi thresholding in fact doesn't register desired results with mammograms where the rate of success was approximately 79.7 %. This can be deduced through the results of two systems using these techniques when using the first technique to extract the mass in the system proposed by N. A. ISA et. al. achieved 94.59% a sensitivity and the number of false positive per image is 3.90[62] whilst the system which uses thresholding segmentation method proposed by Dominguez, A.R. et. al. achieved a sensitivity of 80% at 2.3 false-positives [91].

5.2.4 Feature extraction

Features are often used to point the numerical information group that is pertinent to solve the computational tasks related to certain applications, specifically, features can refer to:

- A result of neighborhood operations applied to the images to extract or detect the feature.
- Defined simple structures of the image such as points or edge and more complex structures such as the objects.

A lot of the features have been extracted in order to identify and classify the abnormalities of mammogram images, where the extraction method of texture, photometric, shape and etc.

Features play a very effective role to detect the mammographic image abnormalities due to its nature. In this work 15 features (photometric, texture) have been extracted and then 7 features of them were selected and were utilized to build ANN according to their ability to segregate normal and abnormal lesions with masses and microcalcifications, and this has proved with good results. Figure 5.4 displays the sub steps of this stage.

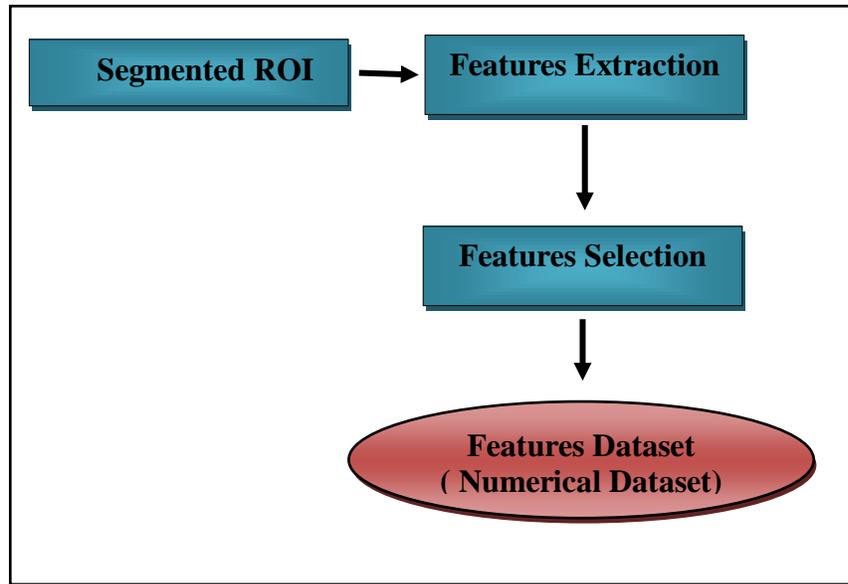


Figure 5.4 Block Diagram of Features Extraction and Selection phase.

For each of the extracted region, 15 features were calculated. To calculate these features, the lesion's area and boundaries should be calculated firstly. These features can be grouped into two groups as the following:

5.2.4.1 Photometric features

This group gives estimations of optical density levels of (ROI) area and the absolute intensity according to the way that has been used to extract photometric features was done by M.Sameti et. al. where their proposed method can detect breast cancer by nearly 72 % [68].

The value of optical density for each pixel at point (i,j) of (ROI) pixel (i,j) can be defined by the following::

$$OD_{ij} = \log(I_{ij}/I_0) \quad (13)$$

I_{ij} is the value of the intensity of pixel at point (i, j), and the intensity of the background is I_0 which can be calculated by way of three:

The first way, calculate the mode value of all pixels intensity in the image excluding those pixels belonging to the abnormal area, could also calculated the mean value of the

whole intensity or the intensity value of the first image pixel of the background. In this work, I_0 is the mean of overall intensity.

The photometric features that have been calculated were three different features rely on the optical and image intensity value. These features are:

1. **OD_Sum:** The measurement of the integrated abnormality optical density (un-normalized value).
2. **OD_max:** The highest value of the abnormality optical density.
3. **OD_variance:** The variance function value of the abnormality optical density (normalized value).

5.2.4.2 Haralick Texture Features:

12 different features were calculated as the texture features.

4. **Mean:** estimates the value in the image in which central clustering occurs and can be obtained from the following using the following formula:

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N p(i, j) \quad (14)$$

Where in the image of size $M \times N$, the gray value of pixel at point (i, j) is $p(i, j)$.

5. **Standard Deviation:** deviation describes the dispersion within a local region. It is determined using the following:

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (p(i, j) - \mu)^2} \quad (15)$$

Where σ estimates the square deviation of the grey pixel value $p(i, j)$ mean.

6. Smoothness:

$$R = 1 - \frac{1}{1+\sigma^2} \quad (16)$$

Where R measures the gray level contrast that can be utilized to create smoothness descriptors.

7. Skewness: A number only describes the form of the distribution while S describes the level of distribution of a pixel asymmetry in the particular window around the pixel mean.

$$S = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(\frac{p(i,j)-\mu}{\sigma} \right)^3 \quad (17)$$

Where p is the pixel gray value at point (i,j), σ is the standard deviation and μ is the mean.

8. Kurtosis: K calculates the flatness and peakness of a distribution proportional to the distribution in normal case.

$$K = \left\{ \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(\frac{p(i,j)-\mu}{\sigma} \right)^4 \right\} - 3 \quad (18)$$

Where, $p(i,j)$ is the pixel value at point (i,j), μ and σ are the Mean and Standard Deviation respectively. The -3 term makes the value zero for a normal distribution.

9. Homogeneity : Homogeneity returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. The range of Homogeneity is [0 1]. Homogeneity is 1 for a diagonal GLCM.

$$H = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (19)$$

10. Entropy: A statistical measure of randomness that can be used to characterize the texture of the input image. Entropy, h can also be used to describe the distribution variation in a region:

$$h = - \sum_{k=0}^{L-1} Pr_k (\log_2 Pr_k) \quad (20)$$

Where, Pr is the probability of the k -th gray level, which can be calculated as $Zk / m*n$, Zk is the total number of pixels with the k -th grey level and L is the total number of grey levels.

11. Energy: returns the sum of squared elements in the Grey Level Co-Occurrence Matrix (GLCM). Energy is also known as uniformity. The range of energy is [0 1]. Energy is 1 for a constant image.

$$E = \sum_{i,j} P(i,j)^2 \quad (21)$$

12. Contrast: returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. The range of Contrast is [0 (size GLCM, 1)-1] 2]. Contrast is 0 for a constant image.

$$C = \sum_{i,j} |i - j|^2 P(i,j) \quad (22)$$

13. Correlation: returns a measure of how correlated a pixel is to its neighbor over the whole image. The range of correlation is [-1 1]. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN (Not a Number) for a constant image.

$$Corr = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j} \quad (23)$$

14. Information Measure of Correlation:

This feature can be determined using the formula:

$$IMC = (1 - \exp [-2(HXY2 - HXY)])^{\frac{1}{2}} \quad (24)$$

Where $HXY = -\sum_i \sum_j p(i, j) \log(p(i, j))$, HX and HY are the entropies of Px and Py .

15. Sum Variance: This feature puts relatively high weights on the elements that differ from the average value of $P(i, j)$.

$$VARIANCE = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 P(i, j) \quad (25)$$

5.2.5 Feature Selection

Then the probability density function (PDF) has been used to select the best features among the 15 features mentioned previously, which was a careful selection of those features that there is no overlap between their distributions in normal and abnormal tissue and the result was 7 features were selected by the end of this process.

The following figures, Figure 5.5 and Figure 5.6 show the PDF of (smoothness) in both normal and abnormal tissues respectively and there is not overlap between them, therefore, this feature has been selected, and vice versa, shown in Figure 5.7 and Figure 5.8 with correlation.

The selected features are:

1. Optical Density Sum
2. Optical density variance
3. Standard Deviation
4. Smoothness
5. Skewness
6. Kurtosis
7. Sum Variance

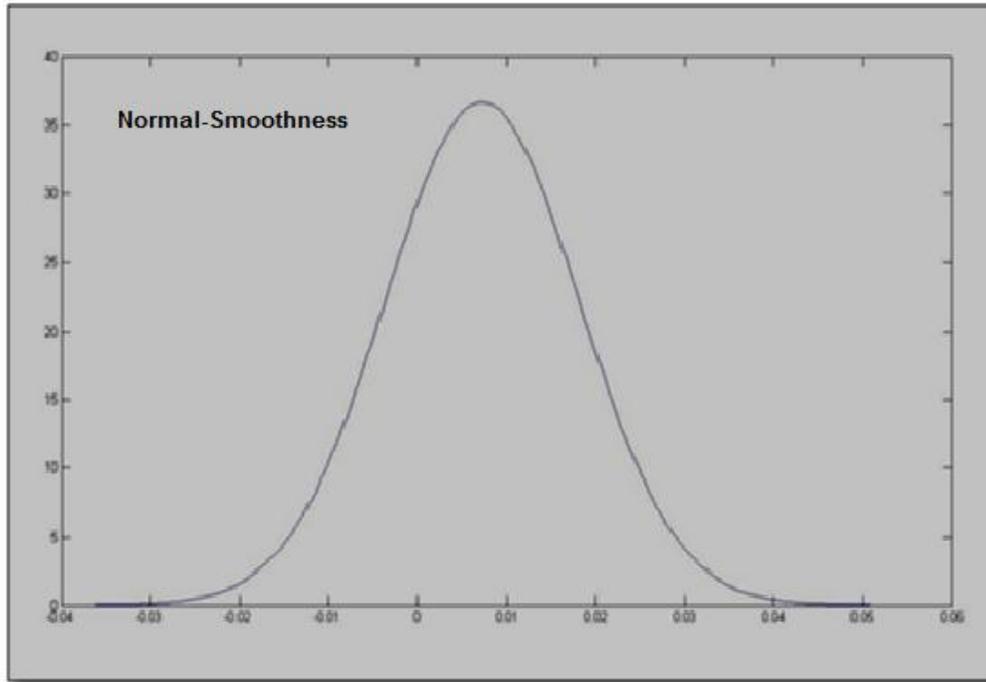


Figure 5.5 Probability Density Function of a Normal Distribution for Smoothness of Normal Tissue

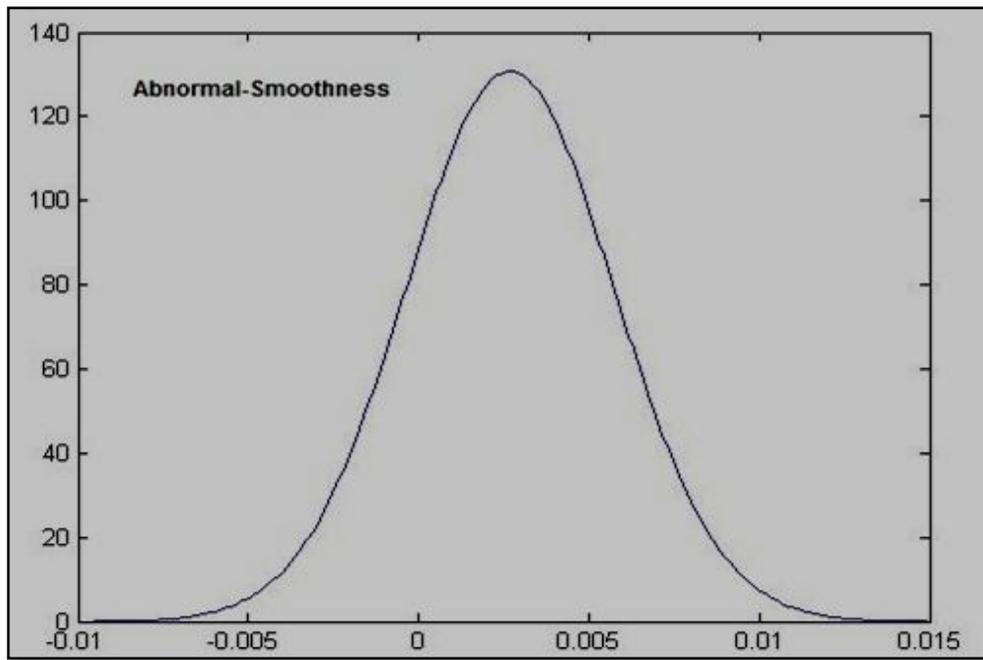


Figure 5.6 Probability Density Function of a Normal Distribution for Smoothness of Abnormal Tissue

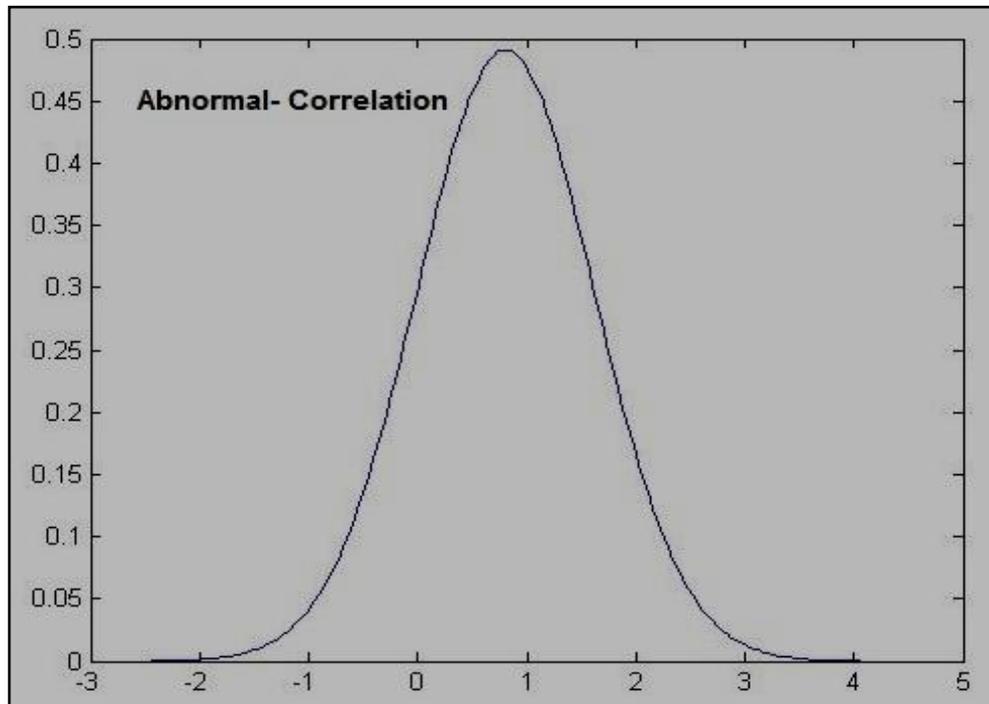


Figure 5.7 Probability Density Function of a Normal Distribution for Correlation of Abnormal Tissue

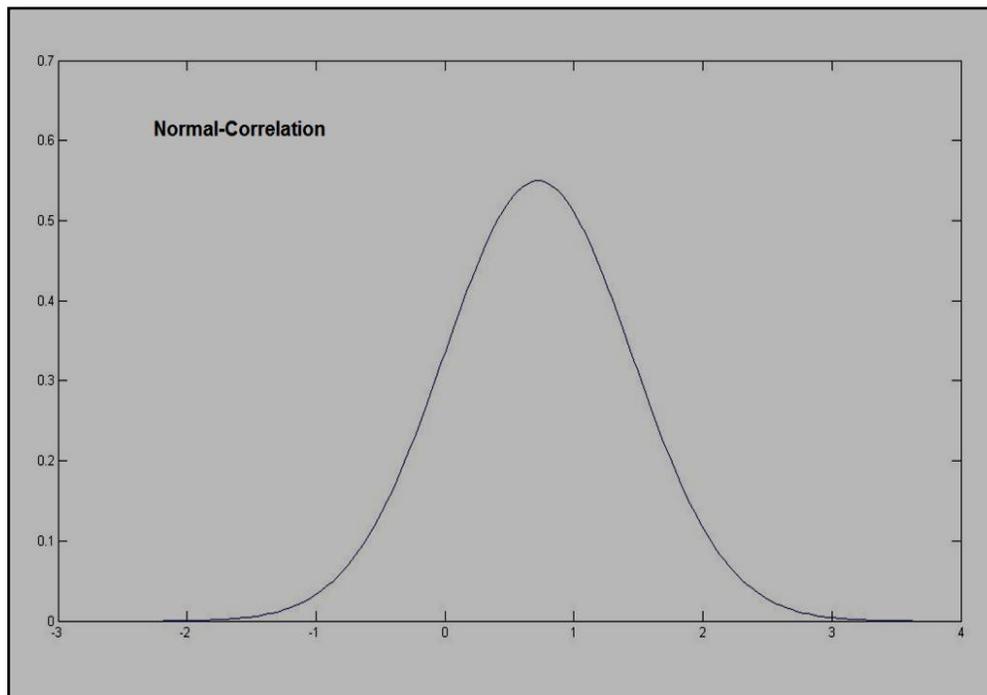


Figure 5.8 Probability Density Function of a Normal Distribution for Correlation of Normal Tissue

5.3.6 Classification

A classifier used in this work is a neural network was designed in Matlab (Figure 5.9 explains briefly how ANN works), candidate regions which has already been decided as abnormal or normal were utilized (feature dataset as an input and class attribute as a target) to train the classifier. The design of this network contains 21 input neurons, one output neuron and Hidden neurons number often is not fixed it is need to tune to reach the best performance.

The figures below summarize this phase, where figure 5.10 shows that the success rate of training phase is 100% and error rate is 0% while figure 5.11 shows the confusion matrices of training, validation and testing actions for the same phase, each matrix explains the number of used mammograms in each training action, validation and testing out of the total number of used mammograms (total used mammograms in this phase is 21), also the error rate was zero in each.

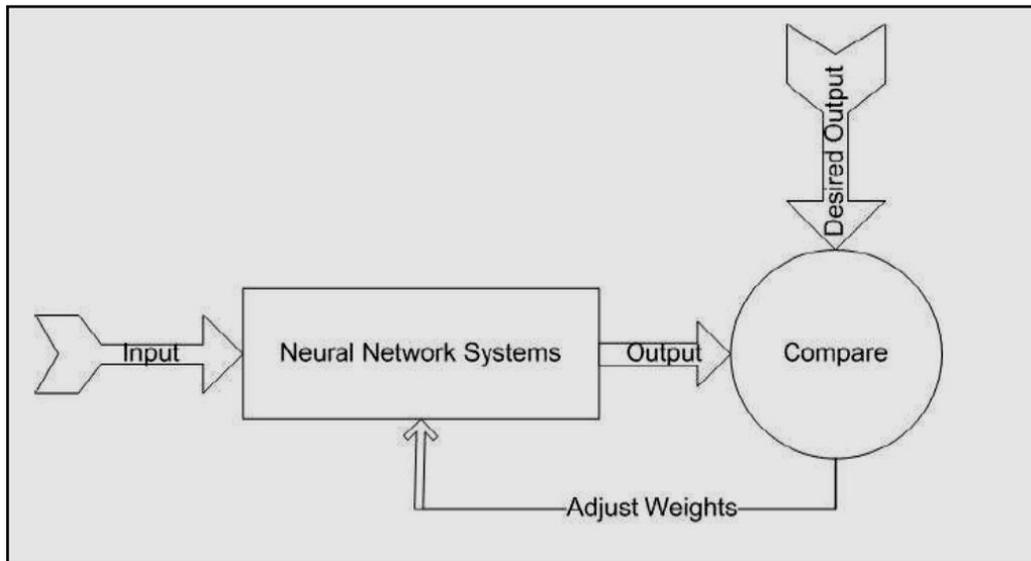


Figure 5.9 ANNs area adjusted or trained so that a particular input leads to a specific desired or target output

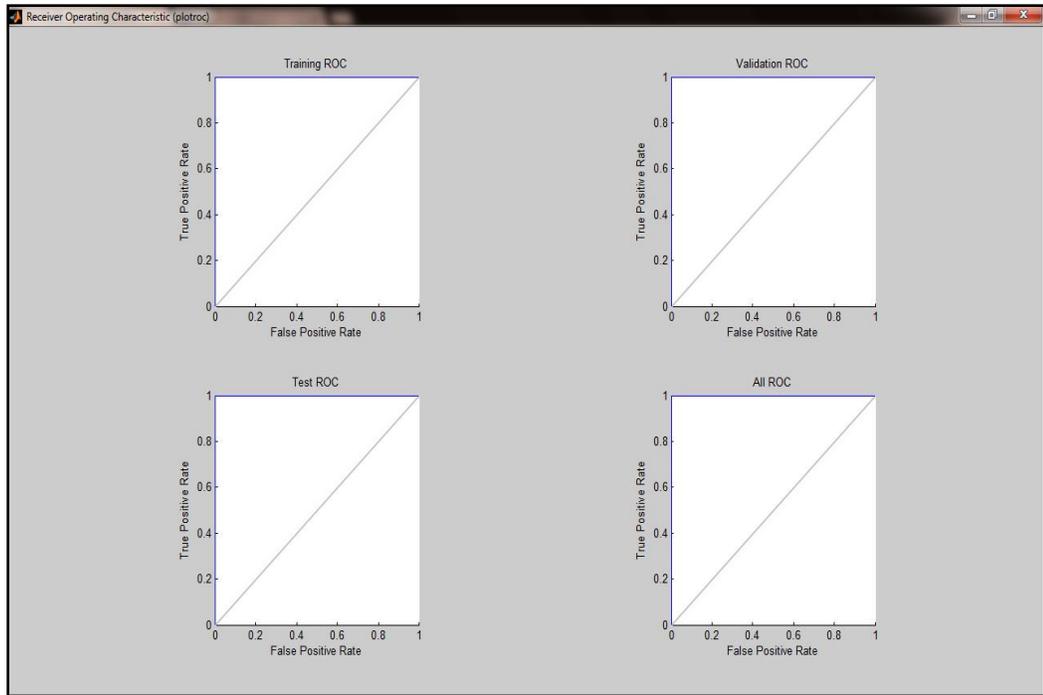


Figure 5.10 Snapshot of ROC for Training Phase



Figure 5.11 Snapshot of Confusion for Training Phase

CHAPTER VI

EXPERMENTS AND RESULTS

This section presents the detection methods and the results of applying these methods to the digital mammogram database that explained below.

6.1 Mammographic Dataset and Used Tools

The dataset used in this work is a known digital mammography database (MIAS) Mammography Image Analysis Society (British research groups organization <http://peipa.essex.ac.uk/>); includes breast images for both sides (left and right) belong to 161 patients which means 322 images, each one belongs to one of three types Normal, Benign or Malignant and distributed as follows :

- 208 images as normal,
- 63 images as benign and
- 51 images as malignant (abnormal) images.

Moreover, this dataset has an adequate information about each mammography image whence a class of the abnormality, position and the character of the background tissue. The images have been digitized with a” Joyce-Lobel” scanning microdensitometer with a resolution of ($50 \mu\text{m} \times 50 \mu\text{m}$), 8 bits symbolize each pixel with 1024×1024 pixel size and at 256 gray levels.

By using MATLAB® R2010b/2013a this automatic system was built in computer with following components:

- ✓ CPU Intel® Core™ 2Duo Processor T8100 (2.10 GHz)
- ✓ 2GB RAM.
- ✓ Hard Disk Drive 250 GB.
- ✓ Microsoft Windows 7 Ultimate 32-bit Copyright® 2009 service pack 1.

6.2 Experiments

Experiment: In this experiment we test our method on mammographic image contains a malignant mass.

- **Step I : Preprocessing**

1. *Noise Removal*

Here median filter is applied on the mammographic image (mdb271) for removing the light problem which is shown as dark lines (If necessary) Figure 6.1 shows the mammograms before and after applying this step.

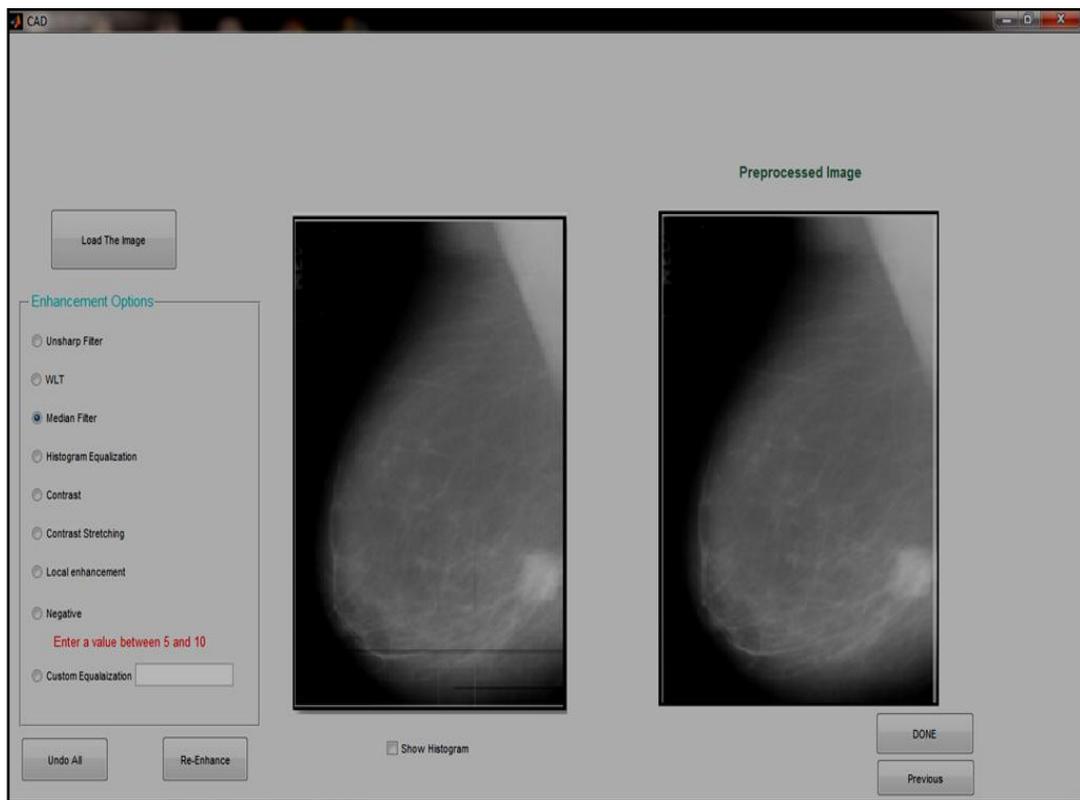


Figure 6.1 Snapshot Shows The Original Image and Filtered one by using the 2D Media Filter.

2. Suppressing Extra Objects

Here in this step the extra image's objects are removed by using morphological, thresholding and contrast enhancement techniques. Figure 6.2 shows the sub steps in this process and Figure 6.3 displays the system screen of this step.

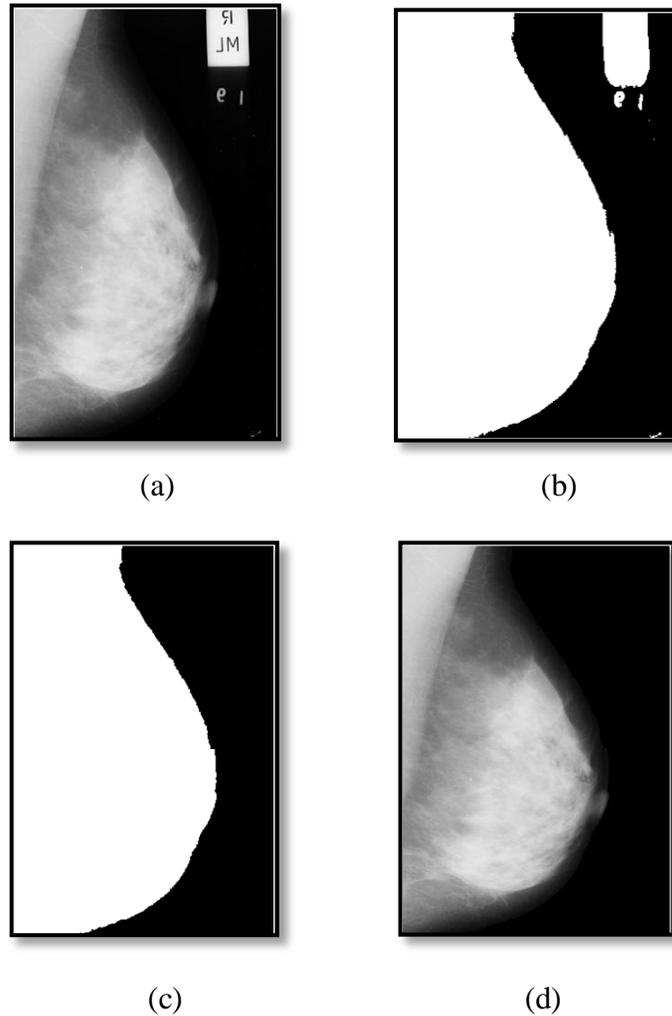


Figure 6.2 (a) Original Image, (b) Binary Image After applying Global Threshold, (c) Result After keeping the Object with the Largest number of pixels and (d) Final Result

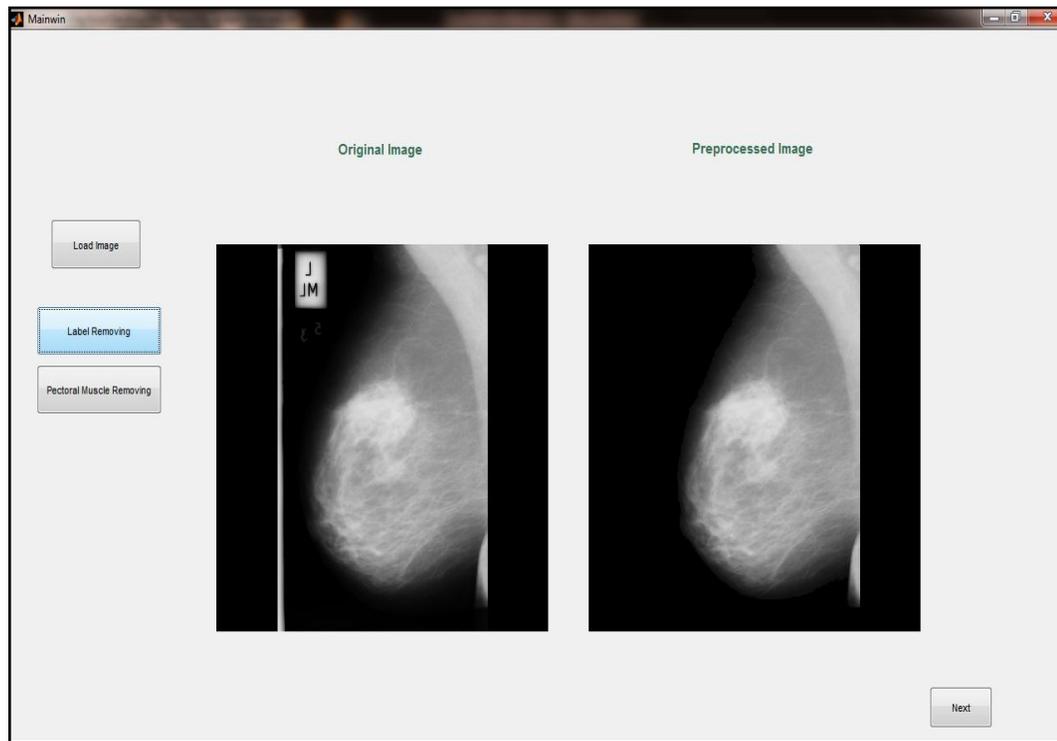
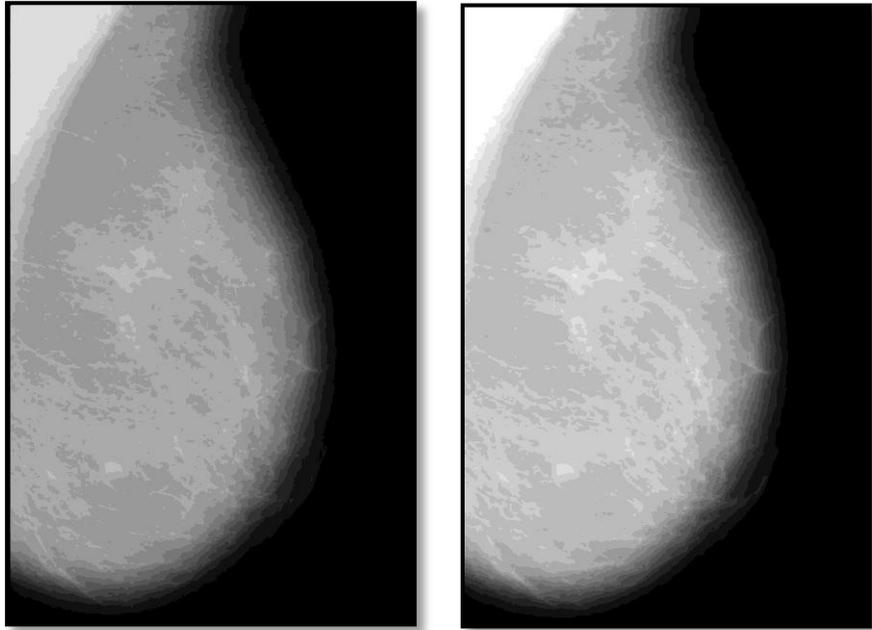


Figure 6.3 Snapshot Shows The Image Before And After Removing labels

3. *Pectoral Muscle Removal*

In this step pectoral muscle is removed by using contrast enhancement and geometrical techniques.

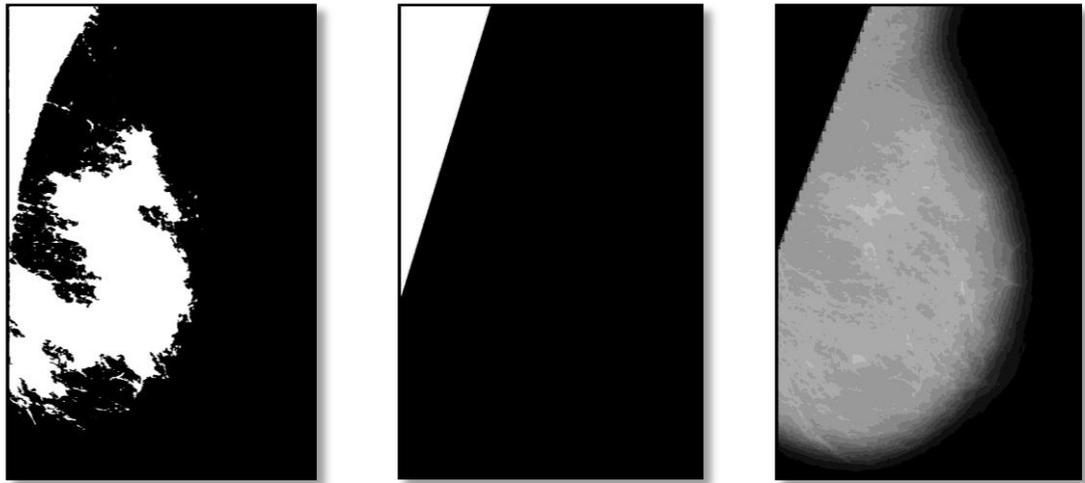
Figure 6.4 (a) and (b) show the result before and after applying contrast enhancement technique on the breast profile image and Figure 6.5 shows the steps for removing the muscle.



(a)

(b)

Figure 6.4 (a) Original Image, (b) Enhanced Image



(a)

(b)

(c)

Figure 6.5 (a) Binary Image of (Figure 6.4 (b)), (b) Segmented part of the muscle and (c) Output image.

Figure below 6.6 is a snapshot of this step on the proposed system screen.

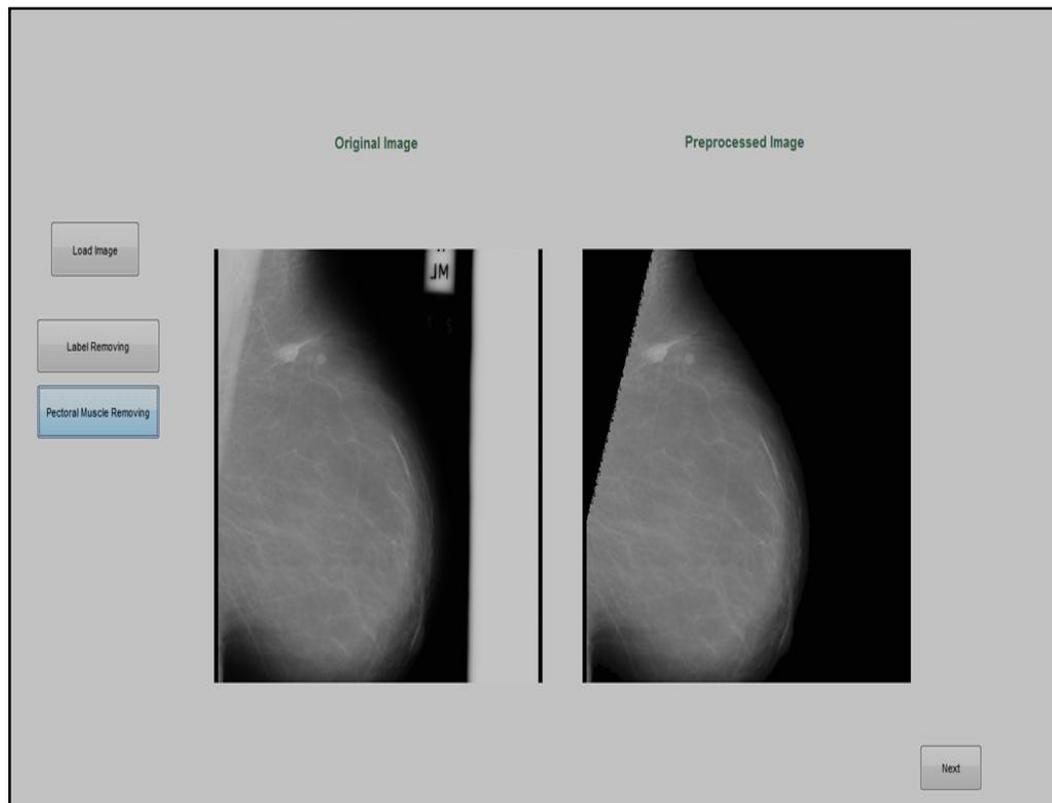


Figure 6.6 Snapshot Shows The Image Before And After Removing the Pecctoral Muscle

- **Step II : Enhancement**

This step prepares the image for ROI segmentation by applying image enhancement techniques. Figure 6.7 shows the resulted image from the previous step before and after applying one of the sharp masks and histogram modification techniques and the mass (suspicious region) appears lighter. Figure 6.8, Figure 6.9 and Figure 6.10 show the result after applying the rest of enhancement techniques that can be used if necessary.

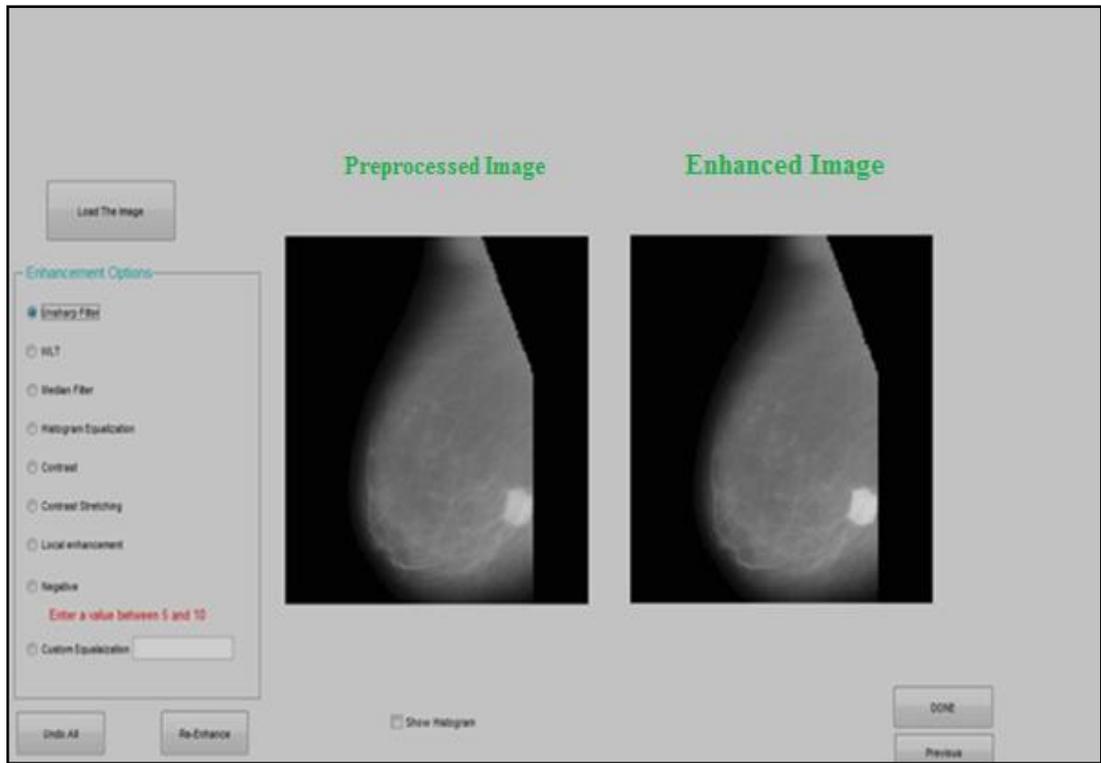


Figure 6.7 Snapshot Shows The Image Before And After Applying Enhancement Techniques

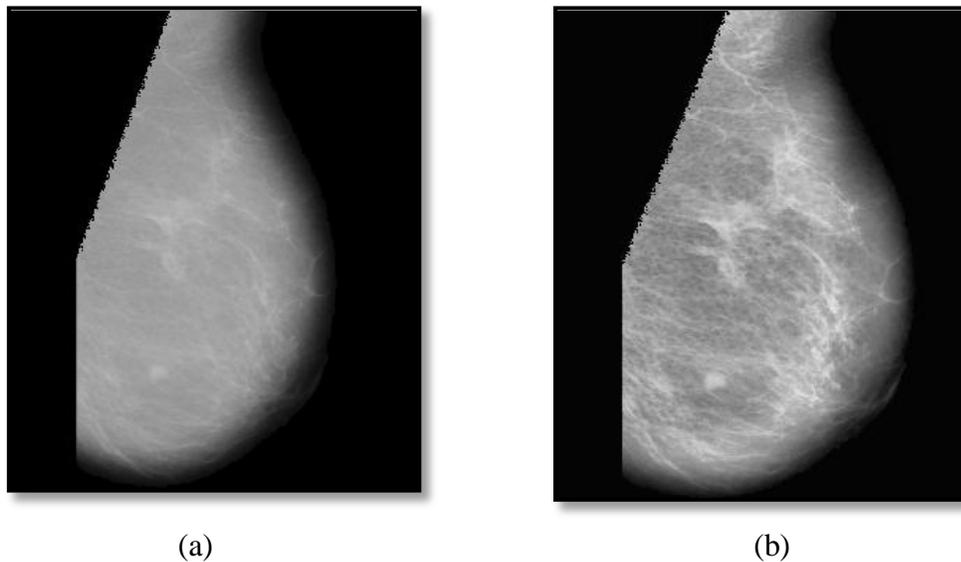
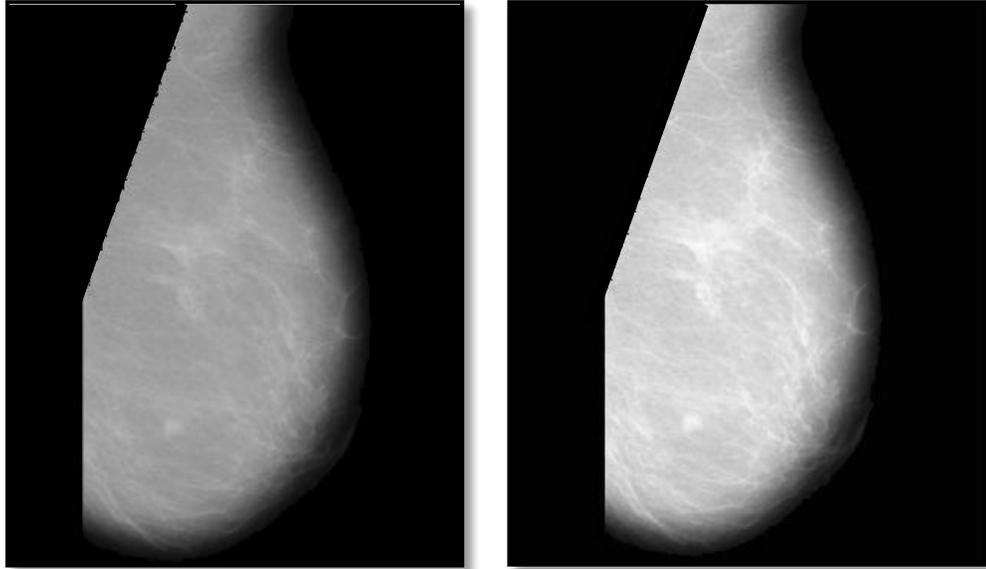


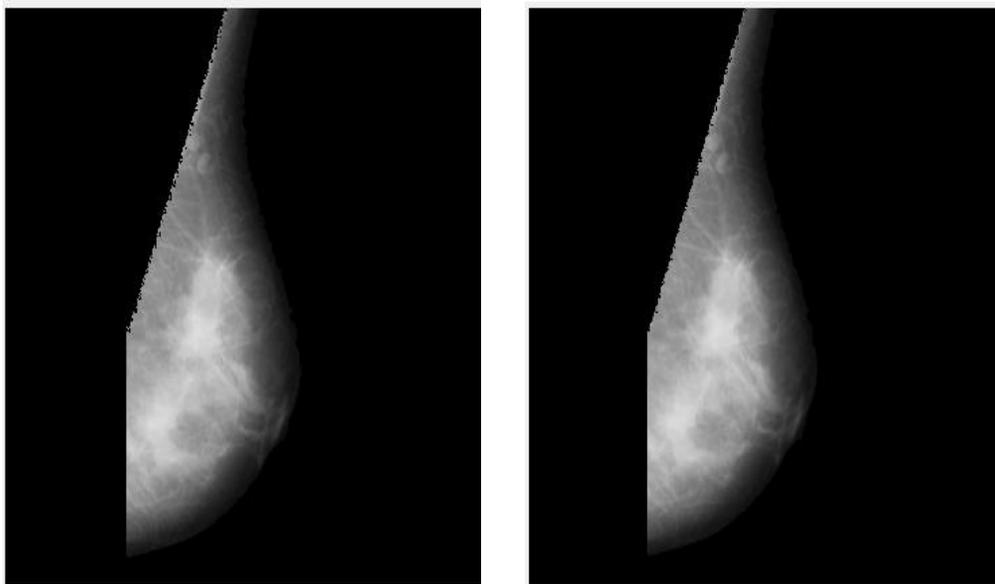
Figure 6.8 (a) Original Image, (b) Image After Modifying The Contrast.



(a)

(b)

Figure 6.9 (a) Original Image, (b) Enhanced Image by Equalizing the histogram.



(a)

(b)

Figure 6.10 (a) Original Image, (b) Enhanced the edge of the ROI by using the Median Filter.

- **Step III : Segmentation**

In this stage of system's work ROI (abnormal region) is extracted by using multi-thresholding OTSU or seed-based region growing (SBRG) image segmentation techniques. Figure 6.11 shows the result after applying SBRG, whereas Figure 6.12 (a) and (b) show the result of multi-thresholding OTSU.

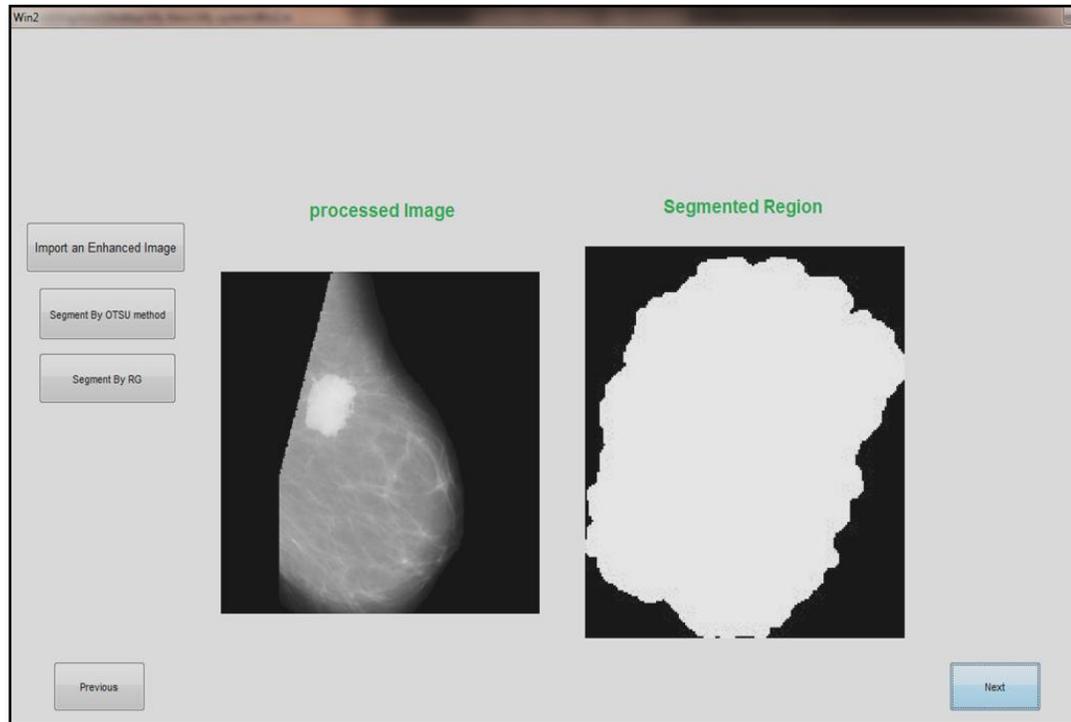


Figure 6.11 Snapshot shows the segmented region after applying SBRG

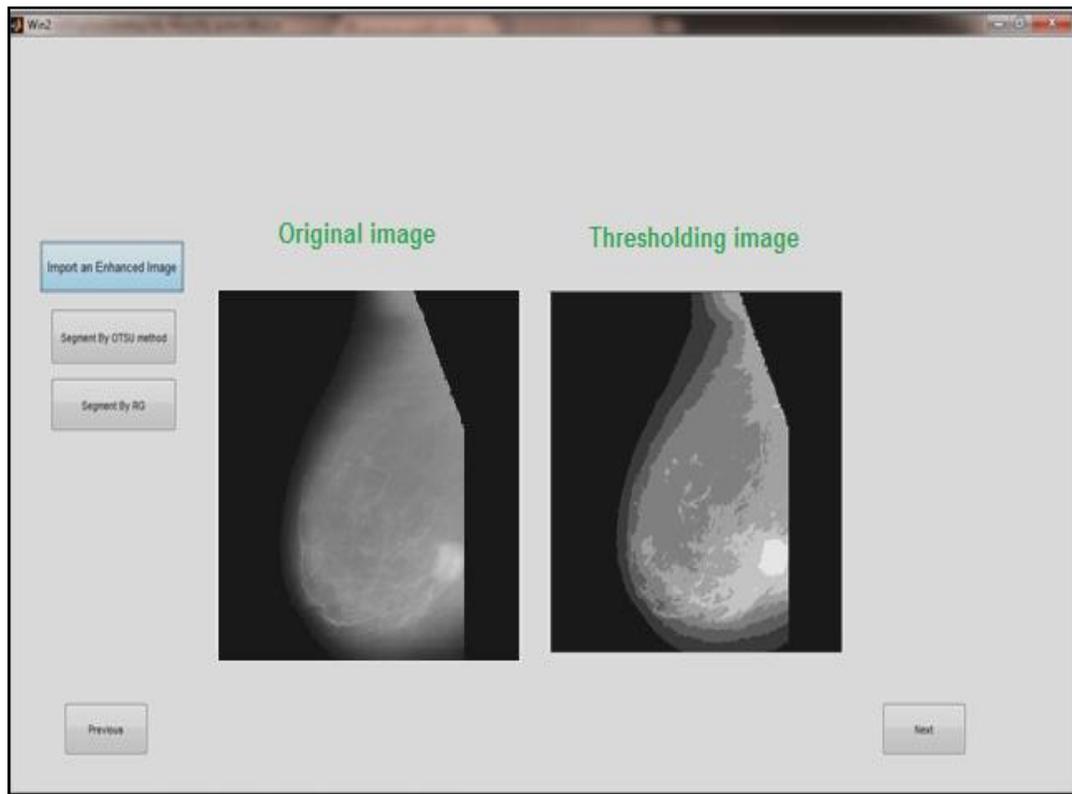


Figure 6.12(a) Snapshot shows the segmented region after applying multi thresholding OTSU's method

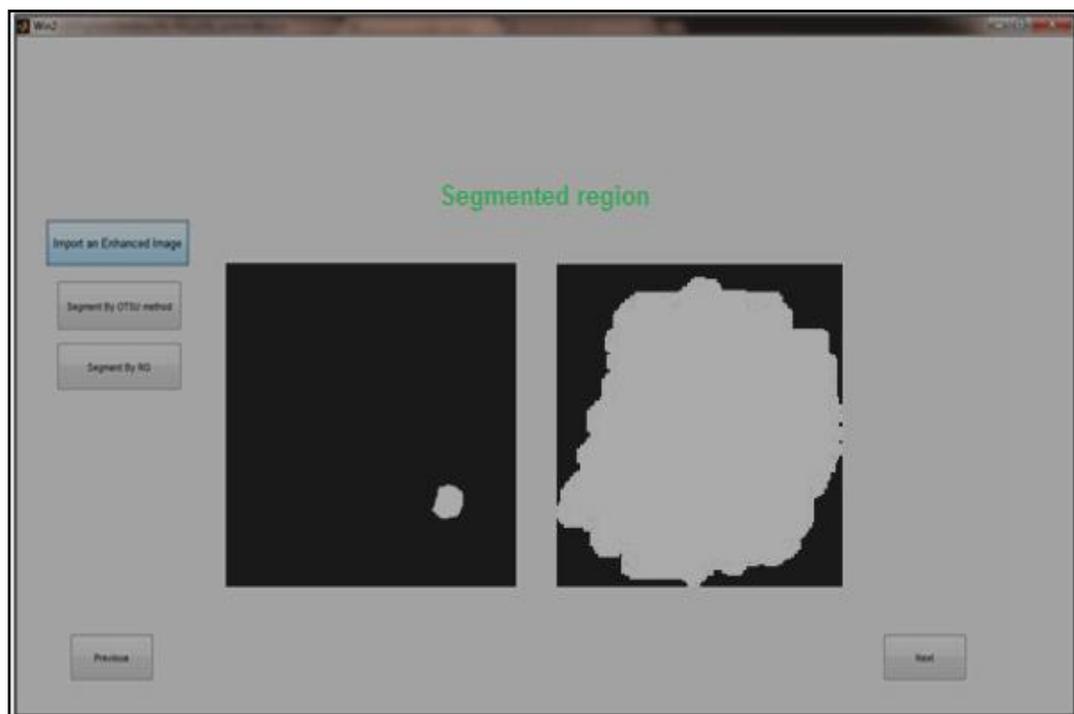


Figure 6.12(b) Snapshot shows the segmented region after applying multi thresholding OTSU's method

- Step IV : Feature Extraction

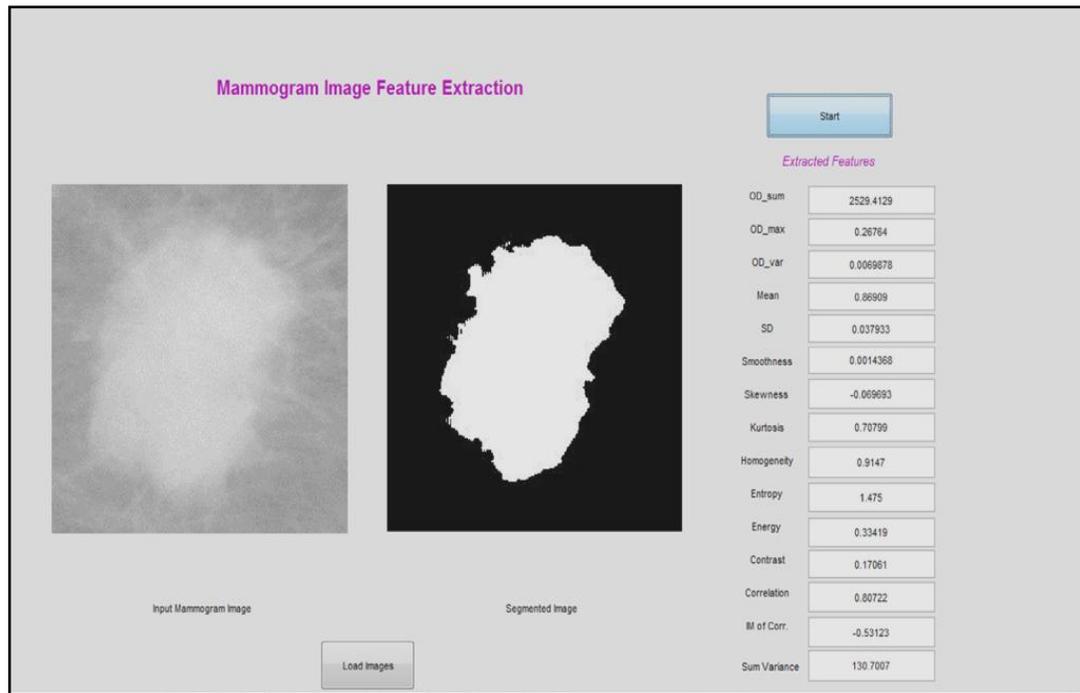


Figure 6.13 Snapshot of Output Result of the Features Extraction

After getting the feature dataset for all the mammograms images as shown in Figure 5.14, new attribute has been added to this data set represents the class of each image as normal or abnormal. This attribute is one of two values 0 for normal image or 1 for abnormal image as shown in Figure 5.16.

6.3 Classification

This process is divided into 2 stages:

- 1- Training process and
- 2- Testing process.

Training process has explained in the previous chapter and in testing stage, one test image is divided into sub images (16×16 parts) and then classification is performed by using trained classifier from the prior phase for each part. The number of test images which have been tested is 39 images: 24 abnormal and 15 normal.

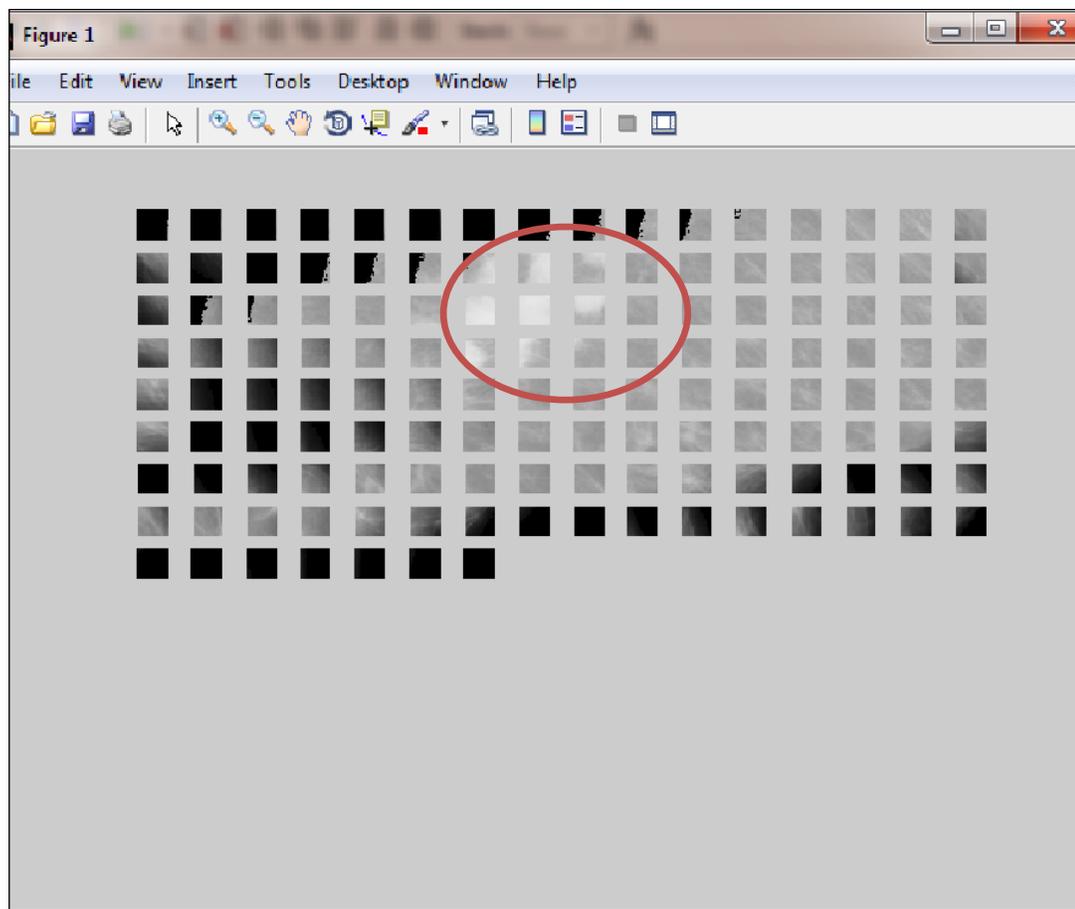


Figure 6.16 Snapshot of Sub-images (mdb184)

Figure 6.16 shows cancerous regions surrounding by the red circles in a divided image (mdb184) positions of these regions are: 23,24,25,38,39,40,41,55,56,57 and 58.

Table 6.1 below shows part of the result after classifying all sub images to appropriate class either around 0 or 1. Results in green are the correct result of the abnormality part whilst results in red shows false positive (FP) results.

Table 6.1 Part of The Results After Classifying a Divided Image

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|----------|--------|---------|----------|------------|---------|----------|----------|----------|-------|
| 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 2 | NaN | NaN | 0.00085 | 0.01312 | 0.0378 | 0.976 | 1 | NaN | NaN | NaN |
| 3 | NaN | NaN | 0.999 | 0.999073 | 0.944 | 0.002 | 6.11E-05 | 0.001 | 7.29E-05 | 0.034 |
| 4 | 0.0430 | 0.435 | 0.370 | NaN | NaN | NaN | 5.92E-05 | 0.845 | 0.9814 | 0.999 |
| 5 | 0.999 | 0.001 | 0.0001 | 7.36E-05 | 0.0023 | 0.001 | 0.003266 | 0.0747 | 0.99 | 0.415 |
| 6 | 0.759 | 0.442 | 0.231 | 0.120 | 0.99998006 | 0.9998 | 0.157377 | 4.32E-05 | 0.0005 | 0.003 |
| 7 | 7.75E-05 | 0.0001 | 0.0048 | 0.013 | 0.3287565 | NaN | NaN | 0.213 | 0.483 | 0.999 |
| 8 | 0.881 | 0.063 | 0.012 | 0.031 | 0.00029459 | 0.08462 | 0.002415 | 0.000 | 0.001 | 0.001 |

6.4 Evaluation

This section represents the results that were achieved in abnormality detection and diagnosis techniques and evaluates the system performance. These results were obtained of classification among normal, abnormal classes by using neural network classifier.

Performance evaluation depends on the calculation of the following elements (Figure 6.17):

- True Positive (TP).
- True Negative (TN).
- False Positive (FP).
- False Negative (FN).
- Completeness (CM) = Sensitivity = True Positive Rate (TPR) = $TP/(TP+FN)$.
- Specificity (SPC) = True Negative Rate (TNR) = $1 - \text{False Positive Rate (FPR)} = 1 - FP/(FP+TN)$.
- Correctness (CR) = Positive Predictive Value (PPV) = $TP/(TP+FP)$.

| | | Identified | |
|--------|-----|------------|----|
| | | Yes | No |
| Actual | Yes | TP | FN |
| | No | FP | TN |

Figure 6.17 Performance Evaluation

Table 6.2 (a) Statistical Values Of The System Performance

| NO. | Total of sub images | Real abnormality | TP | FP | FN | TN | P | N |
|---------|---------------------|------------------|----|----|----|-----|----|-----|
| mdb271 | 149 | 9 | 9 | 12 | 0 | 128 | 9 | 140 |
| mdb 184 | 135 | 10 | 8 | 12 | 2 | 115 | 10 | 127 |
| mdb 165 | 66 | 14 | 14 | 5 | 0 | 47 | 14 | 52 |
| mdb 150 | 152 | 17 | 17 | 3 | 0 | 132 | 17 | 135 |
| mdb 134 | 201 | 5 | 5 | 6 | 0 | 190 | 5 | 196 |
| mdb 132 | 159 | 17 | 15 | 2 | 2 | 142 | 17 | 144 |
| mdb 111 | 106 | 22 | 22 | 2 | 0 | 82 | 22 | 84 |
| mdb 83 | 116 | 23 | 23 | 3 | 0 | 90 | 23 | 93 |
| mdb 81 | 119 | 29 | 28 | 3 | 1 | 88 | 29 | 91 |
| mdb 72 | 66 | 10 | 10 | 4 | 0 | 52 | 10 | 56 |
| mdb 28 | 122 | 8 | 8 | 5 | 0 | 109 | 8 | 114 |
| mdb 25 | 146 | 16 | 16 | 7 | 0 | 123 | 16 | 130 |
| mdb 5 | 89 | 15 | 15 | 1 | 0 | 73 | 15 | 74 |

Table 6.2 (b) Statistical Values Of The System Performance

| NO. | CM=TPR | FPR | CR=PPV | F1 | ACC |
|---------|-----------------|-----------------|-----------------|----------|-----------------|
| mdb271 | 1 | 0.085714 | 0.428571 | 0.6 | 0.919463 |
| mdb 184 | 0.8 | 0.094488 | 0.4 | 0.533333 | 0.89781 |
| mdb 165 | 1 | 0.096154 | 0.736842 | 0.848485 | 0.924242 |
| mdb 150 | 1 | 0.022222 | 0.85 | 0.918919 | 0.980263 |
| mdb 134 | 1 | 0.030612 | 0.454545 | 0.625 | 0.970149 |
| mdb 132 | 0.882353 | 0.013889 | 0.882353 | 0.882353 | 0.975155 |
| mdb 111 | 1 | 0.02381 | 0.916667 | 0.956522 | 0.981132 |
| mdb 83 | 1 | 0.032258 | 0.884615 | 0.938776 | 0.974138 |
| mdb 81 | 0.965517 | 0.032967 | 0.903226 | 0.933333 | 0.966667 |
| mdb 72 | 1 | 0.071429 | 0.714286 | 0.833333 | 0.939394 |
| mdb 28 | 1 | 0.04386 | 0.615385 | 0.761905 | 0.959016 |
| mdb 25 | 1 | 0.053846 | 0.695652 | 0.820513 | 0.952055 |
| mdb 5 | 1 | 0.013514 | 0.9375 | 0.967742 | 0.988764 |

After getting the statistical values shown in Table 6.2 (a) and (b), we could evaluate the proposed system by getting the following values:

- **Average of CM = Sensitivity= 98.3134 %**
- **Average of CR= 74.997 %**
- **Average of Specificity =92.28 %**
- **Average of Accuracy = 93.8059 %**

And the following table (Table 6.3) shows the transactional relationship between sensitivity and specificity where the increasing in sensitivity causes decreasing in specificity, and the chart of Receiver Operating Characteristics (ROC curve) , Figure 6.18 shows that the best result was at threshold value more than 0.90 where it is close to 1, obtaining the following results have been got by choosing more than one threshold, specificity and sensitivity values have been calculated in each change of threshold value.

Table 6.3 Statistical Values Of The Sensitivity and FP at Threshold Point

| T | Sensitivity | Specificity | False Positive |
|-------------------------------------|--------------------|--------------------|-----------------------|
| T1 (>=0.50) | 0.33 | 0.99 | 0.01 |
| T2(>=0.60) | 0.53 | 0.9 | 0.1 |
| T3 (>=0.70) | 0.71 | 0.8 | 0.2 |
| T4 (>=0.81 and < 0.85) | 0.89 | 0.5 | 0.5 |
| T5 (>= 0.85) | 0.951 | 0.262 | 0.737979 |
| T6 (>= 0.90) | 0.983 | 0.188 | 0.922775 |

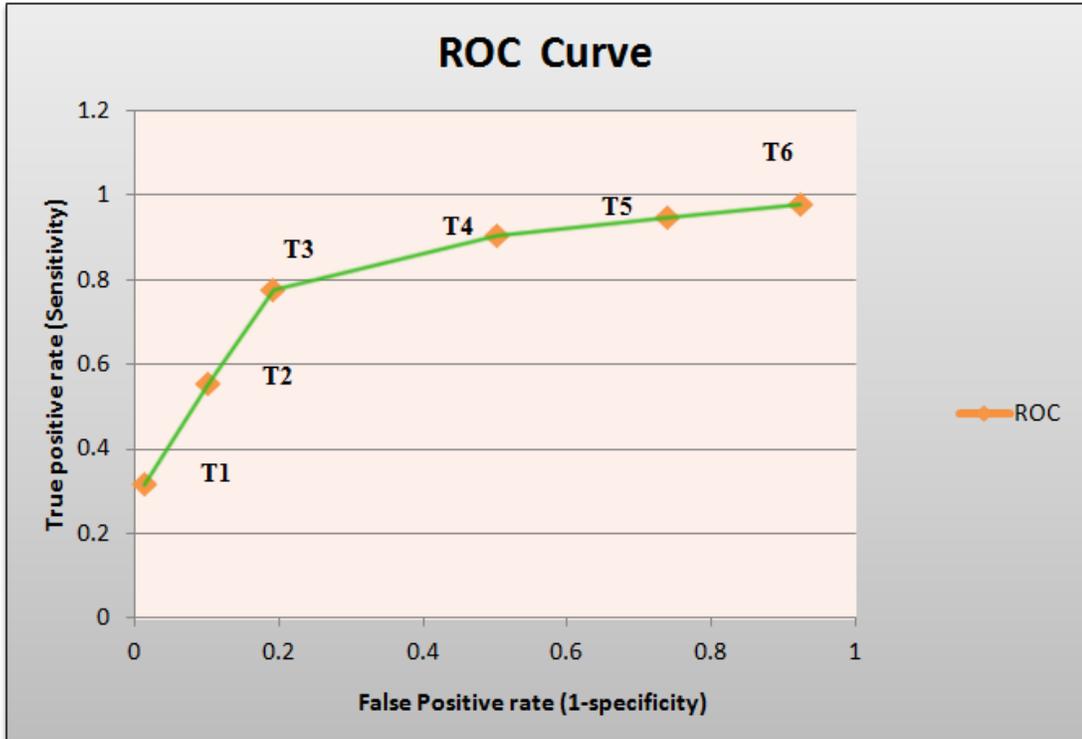


Figure 6.18 ROC Curve

Table 6.4 Comprehensive Summary of the Breast Cancer Detection Systems Using Different Techniques

| Author | Accuracy | Sensitivity | Specificity | FP |
|--------------------------------------|---------------------------|----------------|---------------|------------------------------|
| Proposed System | <i>93.8059%</i> | <i>98.31 %</i> | <i>92.28%</i> | <i>0.401</i> |
| Sankar and Thomas[106] | | <i>82%</i> | | |
| Elter, M. et. al. [107] | <i>84 %</i> | | | |
| Heath, M.[108] | <i>80%</i> | | | |
| Thangavel and Karnan[110] | <i>94%</i> | | | |
| Eltonsy et al[114] | <i>81%,88%and 92%</i> | | | <i>5.4, 2.4, and 0.6</i> |
| Rojas et. al. [115] | | <i>60%</i> | <i>80%</i> | |
| Belloti et. al. [116] | | <i>80%</i> | | <i>4.23</i> |
| Chan et.al [117] | | <i>85%</i> | | <i>1.32</i> |
| Martins et. al. [118] | <i>89.30%</i> | | | |
| Campanini et. al. [119] | | <i>80%</i> | | <i>1.1</i> |
| Martins et. al. [120] | <i>86.85%</i> | | | |
| L. F. A. Campos et. al. [122] | <i>97.3%</i> | <i>96%</i> | <i>100%</i> | |

CHAPTER VII

CONCLUSIONS AND FUTURE WORKS

7.1 Conclusions

Brest cancer is one of the most dangerous and that second leading reason of death among females. There are many screening programs play main role in early detection diagnose the disease which increase chances of recovery. One of the most effective program in this area is digital mammography screening.

Mammography imaging produces a wide set of digital mammography images which then be cleared by radiologist doctors. Due to the enormous number of the breast anomalies' properties and the nature of the human visual system, it is normal that some abnormalities be missed or miss clarified. Thus a great number of unnecessary biopsies as a result of the vast number of false positive (FP) results occur.

In the detection and diagnosis systems, there are enormous number of algorithms which have been improved for detection of breast abnormality (masses , calcifications) and to form an integrated system to increase an accurate diagnosis of the radiologists which lead to minimize the number of false positive. There are many challenges and problems faced by such systems, particularly systems for detection of breast cancer due to the nature of mammographic images that are handled, so had to be accurate and careful in the selection of appropriate algorithms to build the required system. Choosing carefully, conducting experiments and observe the results of the used algorithms and techniques play an important role in raising the efficiency of the system, so we find a clear disparity in the accuracy of systems that were previously displayed.

As mentioned in the first chapter, this work focuses on extracting the suspicious region with finding the useful features, so were chosen region growing and multi-thresholding algorithms for detecting and extracting the abnormality then collection of the photometric and texture features were extracted to build the classifier. To facilitate the work of the segmentation algorithms the nature of the abnormality contrast has been exploited by using unsharp mask and histogram modification enhancement algorithms which was helpful to get the seed point which is required for the work of region growing algorithm.

The used techniques and algorithms in the proposed system are mentioned below:

1. The mammography image label is removed by using global threshold followed by applying morphological techniques.
2. Chest muscle is removed using geometrical operations.
3. Enhance the ROI was done by using nonlinear and linear operators.
4. Segmentation process is done using two techniques seed-based region growing that worked efficiently better than multi thresholding OTSU's method.
5. 15 Haralick texture and photometric features were extracted then just 7 of them were selected depending on the best in the distinction between normality and abnormality of the breast tissue. Probability density function is proposed to select un-overlapping features.
6. Classification process passes through two steps : training and testing, for training process artificial neural network was suggested and error classification rate registered 0% . Testing step is done using 39 mammograms each image was divided into (16×16) tested parts to determine the type of each part if it was normal or contains cancer.

The proposed sysetem has achieved a sensitivity of **98.3134** %, a correctnes of **74.997** % , a specificity of **92.28** % and an accuracy of **93.8059** %.

7.2 Future Works

Future plan should focus on working towards develop segmentation algorithms used to be able to detect and extract the cancerous regions even if there is more than one suspected region in one image. Features space of the masses and calcifications is extensive which opens the door to improve a classifier to determine the type of lesion as a benign or a malignant tumor, so the expected result would be e.g. malignant mass, benign calcification or normal. In the following are some suggestions for later work:

- Improving the performance of the segmentation algorithms through developing its mechanisms to choose the threshold depending on the statical information of the image histogram.
- Extracting the features of the abnormality more than what have been extracted and using advanced algorithms and more complexity to select the most appropriate features.
- A high sensitivity classifier should be designed to accurately determine abnormal area as a benign or malignant tumor.
- Using Discrete Wavelet Transform the mathematical tool to segment classification.
- Decreasing the rate of FP errors to be able to call a segment function to extract the suspicious area once it gets a result denotes to abnormal class.
- Working toward getting the best accuracy rate of the system.

REFERENCES

- [1] American Cancer Society (2003), “Cancer Prevention and Early Detection Facts and Figures”, from <http://www.cancer.org> Last access on 2 June 2013
- [2] American Cancer Society (2003), “Cancer Facts and Figures 2003”.from <http://www.cancer.org> Last access on 15 July 2013
- [3] **POLAKOWSKI WE.** et. al. (1997), “Computer-Aided Breast Cancer Detection and Diagnosis of Masses using Difference of Gaussians and Derivative-based Feature Saliency”, IEEE Transactions on Medical Imaging, Vol. 16(6)
- [4] **GIGER ML.** et al. (1990), “Investigation of Methods for the computerized detection and analysis of mammographic masses”, SPIE, Vol. 1233, 183-184
- [5] **MARKEY MK.** et. al. (2002), “Cluster analysis of BI-RADS descriptions of biopsy-proven breast lesions”, Proceedings of SPIE Vol. 4684
- [6] **ELLIS H.** (2006), Clinical Anatomy. 11th Ed. UK: Blackwell Publishing Ltd
- [7] **ZIMMERMAN B. T.** (2004), “Understanding Breast Cancer Genetics”, United States of America: University Press of Mississippi
- [8] Was available at: www.nccn.com/type-of-cancer/breast-cancer/1350.html, Last access on 8 July 2013
- [9] **SARREGO J.** (2010), “Breast cancer in the young patient”, The American Surgeon 76 (12): 1397–1401
- [10] **GUIDO M.** (2000), “Computer-Aided Detection and Classification of Masses in Digitized”
- [11] American Cancer Society (2012), “Cancer Facts and Figures 2012-2013”, from <http://www.cancer.org> Last access on 10 June 2013

[12] “Breast Imaging Reporting and Data System (BI-RADS)”, from <http://www.acr.org/Quality-Safety/Resources/BIRADS>

[13] **SAMPAT M.** et. al. (2005),”Computer-Aided Detection and Diagnosis in Mammography”, In A. BOVIK, Handbook of Image and Video Processing (pp. 1195-1217), Elsevier Academic Press

[14] **FEIG SA.** et. al. (1995), “Digital Mammography, Computer-Aided Diagnosis and Telemammography”, The Radiologic Clinics of North America, Breast Imaging, Vol 33, 6, 1205-1230

[15] **PHIL E. W.** (1995), “Breast Masses Appropriate Evaluation”, The Radiologic Clinics of North America, Breast Imaging, Vol 33, 6, 1085-1108

[16] Was available at: www.breast-cancer.ca/screening/mammographic-mass-characteristics.html

[17] Was available at: www.healthhype.com /characteristics-of-benign-and-malignant-tumors.html

[18] **MALAGELADA A. O.** (2007),” Automatic mass segmentation in mammographic images”, Universitat de Girona, Girona

[19] **CHRISTOYIANI I.** et. al. (2000), “Fast Detection of Masses in Computer Aided Mammography”, IEEE Signal Processing Magazine, pp. 54-64

[20] University of Washington Medical Center,” Mammography How to prepare for your procedure”, 03/2005 Rev. 05/2005

[21] “Screening and Diagnostic Mammography”, from <http://www.imaginis.com>, Last access on 10 August 2013

[22] **GOTZSCHE P. C., NIELSEN M.** (2009), “Screening for breast cancer with mammography”

[23] **PISANO ED.** et. al. (2004),”Digital Mammography”, Philadelphia: Lippincott Williams & Wilkins.

[24] Was available at: www.radiologyinfo.org/en/info.cfm?pg=mammo

[25] **Y. Faridah** et. al. (2008), "Digital versus screen film mammography: a clinical comparison", University of Malaya, Kuala Lumpur, Malaysia.

[26] **ISHIDA M.** et. al. (1982), "Development of a new digital radiographic image processing system", Proc SPIE 1982;347:42–48.

[27] **LODWICK GS.** et al., "Computer diagnosis of primary bone tumor", Proc SPIE 1963;80:273–275.

[28] **FUJITA H.** et. al. (1987), "Image feature analysis and computer-aided diagnosis in digital radiography", Med Phys. 1987;14:549–556.

[29] **HOFFMANN KR.** et. al. (1986), "Automated tracking of the vascular tree in DSA images using a double-square-box region-of-search algorithm", Proc SPIE. 1986;626:326–333.

[30] **GIGER ML.** et. al. (1995), "Computerized detection of lung nodules in digital chest radiographs"

[31] **CHAN HP.** et. al. (1987), "Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography", Med Phys. 1987;14(4):538.

[32] **NISHIKAWA RM.** et. al. (1995), "Computerized detection of clustered microcalcifications: evaluation of performance on mammograms from multiple centers", University of Chicago, IL 60637, USA.

[33] **ABE H.** et. al. (2003), "Computer-aided diagnosis in chest radiology", results of large-scale observer tests performed at the 1996-2001 RSNA Scientific Assemblies, RadioGraphics, 23:255-265.

[34] Was available at: <http://www.icadmed.com/>. Retrieved Feb. 16, 2012, from <http://cadxmed.com/>

[35] Was available at: <http://cadxmed.com/>. Retrieved Feb. 16, 2012, from <http://cadxmed.com/>

[36] **KRUPINSKI A. E.** (2009), "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future", University of Arizona, Tucson, Arizona.

[37] **TUDDENHAM W.** et. al. (1962), "Visual search, Image organization and Reader error in roentgen diagnosis", Studies of the psychophysiology of roentgen image perception".

[38] **SMITH A.** et. al. (2006), "Lesion Visibility in Low Dose Tomosynthesis. In: Digital mammography", 8th international workshop, IWDM 2006, Manchester, UK, June 18–21, 2006 : proceedings. Astley, S, Brady, M, Rose, C, Zwiggelaar, R (Eds.) (Springer, New York, 2006) pp.160

[39] **RAJESH G.** et. al. (2011), "Histogram Equalization Techniques For Image Enhancement", Haryana, India.

[40] **PIZER S.** et. al. (2005), "Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing".

[41] **THANGAVEL K.** et. al. (2005), "Automatic Detection of Microcalcification in Mammograms", ICGST International Journal on Graphics, Vision and Image Processing, pp. 31-61.

[42] **GONZALEZ R. and WOODS R.** (2002), Digital Image Processing, 2nd Ed, New Jersey: Prentice Hall, Inc.

[43] **SIVARMAKRSHNA R.** et. al. (2000), "Comparing the Performance of Mammographic Enhancement Algorithms", A Preference Study. AJR, 175, pp. 45-51

[44] **Misiti, M.** et al. (2001), "Wavelet Toolbox User's Guide", Version 2.1. The MathWorks, Inc.

[45] **SEZGIN M., SANKUR B.** (2004), "Survey over image thresholding techniques and quantitative performance evaluation", Journal of Electronic Imaging 13 (1): 146–165. doi:10.1117/1.1631315

[46] **OTSU N.** (1979), "A threshold selection method from gray-level histograms", IEEE Trans. Sys., Man., Cyber. 9 (1): 62–66. doi:10.1109/TSMC.1979.4310076.

[47] **SAMETI M.** et. al. (2000),” A method for detection of malignant masses in digitized mammograms using a fuzzy segmentation algorithm”, Proceedings of the 19th International Conference IEEE/MBS, 2000, pp. 513–516

[48] **LI H.** et. al. (2001),”Computerized radiographic mass detection-part II: lesion site selection by morphological enhancement and contextual segmentation”, IEEE Trans. Med. Imaging 20 (4) (2001) 302–313

[49] **CHENG H.D.** et. al.(2005),”Approaches for automated detection and classification of masses in mammograms”, Utah State University, Logan, UT 84322-4205, USA

[50] **MASALA G.** (2005), ”Pattern Recognition Techniques Applied to Biomedical Patterns”, International Journal of Biological and Life Sciences, pp.47-55.

[51] **KARSSE-MEIJER N.** (1991),”A stochastic model for automated detection of calcifications in digital mammograms”, in Proc. 12th Int. Conf. Information Processing Medical Imaging, Wye, U.K., pp. 227-238.

[52] **KARSSE-MEIJER N.** (1993), “Recognition of clustered microcalcification using a random field model, biomedical image processing and biomedical visualization”, In SPIE Proc., Vol. 1905, San Jose, CA, pp. 776-786.

[53] **KARSSE-MEIJER N.** (1993), ”Adaptive noise equalization and recognition of microcalcification clusters in mammograms”, Vol. 7, No. 6, pp. 1357-1376.

[54] **CLARKSON E., BARRETT H. H.**, “Statistical Decision Theory and Tumor Detection”, University of Arizona, Tucson, Arizona.

[55] **STEPHEN M.** et. al.,”Display, Including Enhancement, of Two-Dimensional Images”, University of North Carolina, Chapel Hill, North Carolina.

[56] **MEKLE R.** et. al., "Evaluation of a Multiscale Enhancement Protocol for Digital Mammography", Columbia University, New York.

[57] **KARSSEMEIJER N.**, "Detection of Masses in Mammograms, University Medical Center", Nijmegen, Netherlands.

[58] **RANGARAJ M.** et. al., "Region-Based Adaptive Contrast Enhancement", University of Calgary, Alberta, Canada.

[59] **WANG Z., TAO J.** (2006), "A Fast Implementation of Adaptive Histogram Equalization", School of Information Engineering, University of Science and Technology Beijing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing / China 100080.

[60] **YOSHIDA H.** et. al. (1994), "Automated detection of clustered microcalcifications in digital mammograms using wavelet transform techniques", Proc. SPIE, Vol. 2167, Newport Beach, CA, pp. 868-886.

[61] **YOSHIDA H.** et. al. (1995), "Optimizing wavelet transform based on supervised learning for detection of microcalcifications in digital mammograms", In Proc. IEEE Int. Conf. Image Processing, Vol. 3, Washington, DC, pp. 152-155.

[62] **Nor Ashidi Mat Isa, Ting Shyue Siong** (2012), "Automatic Segmentation and Detection of Mass in Digital Mammograms", Imaging and Intelligent Systems Research Team (ISRT), School of Electrical and Electronic Engineering, Engineering Campus, University Sains Malaysia, 14300 Nibong Tebal, Penang, MALAYSIA.

[63] **Ping-Sung Liao** et. al. (2001), "A Fast Algorithm for Multilevel Thresholding", Department of Electrical Engineering National Cheng Kung University Tainan, 701 Taiwan.

[64] **PRADEEP N.** (2012),"Segmentation and Feature Extraction of Tumors from Digital Mammograms", Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, INDIA. Vol 3, No.4.

[65] **EL-NAQA I.** et. al. (2002), "A Support Vector machine for detection of microcalcifications", IEEE transactions on medical imaging, Vol. 21, No. 12.

[66] **NETSCH T.** (1996),"A scale-space approach for the detection of clustered microcalcifications in digital mammograms", Proc. 3rd Int. Workshop Digital Mammography, Chicago, IL, pp. 301-306.

[67] **MARTINS L. de O.** et. al. (2009), "Detection of Masses in Digital Mammograms using K-means and Support Vector Machine", Federal University of Maranhao , Brazil, Electronic Letters on Computer Vision and Image Analysis 8(2):39-50.

[68] **SAMITI M.** et. al. (2009),"Image Feature Extraction in the Last Screening Mammograms Prior to Detection of Breast Cancer", Vol. 3, NO. 1.

[69] **MAT N. A.** et. al. (2012)," Automatic Segmentation and Detection of Mass in Digital Mammograms", University Sains Malaysia 14300 Nibong, MALAYSIA.

[70] **THANGAVEL K.** et. al.(2005),"Automatic Detection of Microcalcification in Mammograms",ICGST International Journal on Graphics, Vision and Image Processing, pp. 31-61.

[71] **PIZER SM.** (1981), "An automatic intensity mapping for the display of CT scans and other images", *Medical Image Processing: Proceedings of the VIIth International Meeting on Information Processing in Medical Imaging*, Stanford University, 276–309.

[72] **ZIMMERMAN JB.** (1985), "Effectiveness of adaptive contrast enhancement", PhD dissertation, Department of Computer Science, The University of North Carolina at Chapel Hill.

[73] **STEFANOU H.** et. al. (2005), "Wavelet-based Mammographic Enhancement", University of Athens, Greece.

[74] **RIVERA A. R.** et. al. (2012), "Content-Aware Dark Image Enhancement Through Channel Division", *IEEE Transactions On Image Processing*, Vol. 21, No. 9.

[75] **KUMAR H.** et. al. (2012), "Enhancement Of Mammographic Images using Morphology and Wavelet", Bangalore, India.

[76] **SCHARCANSKI J., JUNG C. R.** (2006), "Denoising and enhancing digital mammographic images for visual screening," *Computer Medical Imaging and Graphics*, vol. 30, no. 4, 2006, pp. 243–254.

[77] **INDRA K. M.** et. al. (2011) "Automated digital mammogram segmentation for detection of abnormal masses using binary homogeneity enhancement algorithm", Dept. of Computer Science & Engineering, University of Calcutta, Kolkata, India.

[78] **PAPADPOULOS A.** et. al. (2008), "Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques", Department of Medical Physics, Medical School, University of Ioannina, GR 45110 Ioannina, Greece.

[79] **JHA., R. K.** et. al. (2012), “Noise-induced Contrast Enhancement of Dark Images using Non-dynamic Stochastic Resonance”, 978-1-4673-0816-8/12S.

[80] **JUNG S. W.** et. al. (2012), “Sharpness Enhancement of Stereo Images Using Binocular Just-Noticeable Difference”, IEEE Transactions On Image Processing, Vol. 21, No. 3.

[81] **FANG X.** et. al. (2011), “A Method to Improve the Image Enhancement Result based on Image Fusion”, 978-1-61284-774-0/11.

[82] **HASIKIN K., ISA N. A.**(2012),“ Enhancement of the low contrast image using fuzzy set theory”,14th International Conference on Modelling and Simulation.

[83] **YASMIN M.** et. al. (2012), “ Brain Image Enhancement - A Survey”, World Applied Sciences Journal 17 (9): 1192-1204, 2012,ISSN 1818-4952© IDOSI Publications.

[84] **ZAILANI M. N.** et. al. (2008), ”Mammographic Image Contrast Enhancement through the Use of Moving Contrast Sweep”, School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Seri Ampangan, 14300 Nibong Tebal, Pulau Pinang, Malaysia.

[85] **LEINER B.J.** et. al. (2008),”Microcalcifications Detection System through Discrete Wavelet Analysis and Image Enhancement Techniques”, Popular Univ. of Cesar, Valledupar , Lorena.

[86] **KEKRE H. B., M. G. SAYLEE** (2010), ”Image Segmentation using Extended Edge Operator for mammographic Images”, NMIMS University, Mumbai-56, India.

[87] **ZOU F.** et. al. (2008), "Gradient Vector Flow Field and Mass Region Extraction in Digital Mammograms", In: 21st IEEE International Symposium on Computer-Based Medical Systems, CMBS 2008, Jyväskylä, June 17-19, pp. 41–43.

[88] **JUE W., ALBERT C.S.** (2005), "A segmentation method using compound Markov random fields based on a general boundary model", Bioengineering Program and Department of Computer Science, School of Engineering, The Hong Kong University of Science and Technology, Hong Kong.

[89] **SAHEB B., DR.K.SATYA P.** (2009), "Automatic Detection of Breast Cancer Mass In Mammograms Using Morphological Operators and Fuzzy C –Means Clustering"

[90] **YOUNESI F.** et. al. (2007), "Computer-aided mass detection on digitized mammograms using adaptive thresholding and fuzzy entropy", Department of Medical Physics & Biomedical Engineering, School of Medicine, Medical Sciences / University of Tehran, Tehran-Iran.

[91] **DOMINGUEZ A.R.** et. al. (2007), "Enhanced Multi-Level Thresholding Segmentation and Rank Based Region Selection for Detection of Masses in Mammograms", In: IEEE International Conference on Acoustics, Speech and Signal Processing 2007, ICASSP 2007, Honolulu, HI, April 15-20, pp. 449–452.

[92] **VARELA C.** et. al. (2007), "Computerized Detection of Breast Masses in Digitized Mammograms", Computers in Biology and Medicine 37, 214–226.

[93] **YUAN Y.** et. al. (2008), "Correlative Feature Analysis of FFDM Images.", In: Giger, M.L., Karssemeijer, N. (eds.) Proc. of SPIE Medical Imaging 2008: Computer-Aided Diagnosis, Vol. 6915.

[94] TIMP S., KARSSEMEIJER N. (2006), "Interval Change Analysis to Improve Computer Aided Detection in Mammography", *Medical Image Analysis* 10, 82–95

[95] TIMP S., KARSSEMEIJER N. (2007), "Temporal Change Analysis for Characterization of Mass Lesions in Mammography", *IEEE Transactions on Medical Imaging* 26 (7), 945–953.

[96] KIM S., YOON, S. (2008), "Mass Lesions Classification in Digital Mammography using Optimal Subset of BI-RADS and Gray Level Features", In: 6th International Special Topic Conference on ITAB, 2007, Tokyo, pp. 99–102. IEEE, Los Alamitos.

[97] NANDI R.J. et. al. (2006), "Genetic Programming and Feature Selection for Classification of Breast Masses in Mammograms", In: Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, USA.

[98] HUPSE R., KARSSEMEIJER N. (2008), "Feature Selection for Computer-Aided Detection: Comparing Different Selection Criteria", Vol. 6915, 6915 691503-1.

[99] LI H. et. al. (2008), "Performance of CADx on a Large Clinical Database of FFDM Images", Vol. 5116, pp. 510–514.

[100] FAUCI F. et. al. (2004), "Mammogram Segmentation by Contour Searching and Massive Lesion Classification with Neural Network", In: 2004 IEEE Nuclear Science Symposium Conference Record, Rome, Italy, October 16–22, Vol. 5, pp. 2695–2699.

[101] BALL J.E., BRUCE L.M. (2007), "Digital Mammogram Spiculated Mass Detection and Spicule Segmentation using Level Sets", In: Proceedings of the 29th Annual International Conference of the IEEE EMBS, Cité Internationale, Lyon, France, August 23-26, pp.4979–4984.

[102] **MU T.** et. al. (2007), “Strict 2-Surface Proximal Classifier with Application to Breast Cancer Detection in Mammograms:”, In: IEEE ICASSP 2007, pp. II 477–480 .

[103] **LI H.D** et. al. (1995), ”Markov Random Field for Tumor Detection in Digital Mammography”, IEEE Transactions on Medical Imaging 14(3), 565–576.

[104] **KRISHNAPURAM B.** et. al. (2008), “Multiple-Instance Learning Improves CAD Detection of Masses in Digital Mammography”, In: Krupinski, E.A. (ed.) IWDM 2008. LNCS, vol. 5116, pp. 350–357.

[105] **WEST D.** et. al. (2005),”Mangiameli, P., Rampal, R., West, V.: Ensemble Strategies for a Medical Diagnostic Decision Support System: a Breast Cancer Diagnosis Application”, European Journal of Operational Research 162, 532–551.

[106] **SANKAR D., THOMAS T.** (2007), ”Fractal Modeling of Mammograms based on Mean and Variance for the Detection of Microcalcifications”, In: Proceedings of the 2007 International Conference on Computational Intelligence and Multimedia Applications, Sivakasi, India, December 2007, pp. 334–338.

[107] **ELTER M., HELD C.** (2008), ”Semiautomatic segmentation for the computer aided diagnosis of clustered microcalcifications”, In: Proc. SPIE, San Diego, CA, USA, February 2008, Vol. 6915, 691524-691524-8.

[108] **HEATH M.** et. al. (2001),” The Digital Database for Screening Mammography”, In: Proceedings of the 5th International Workshop on Digital Mammography, Canada, June 11-14, pp. 212–218. Medical Physics Publishing.

[109] **LOPEZ-ALIGUE F.J.** et. al. (2007), "Detection of Microcalcifications in Digital Mammograms", In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS.Lyon, France, August 22-26, pp. 3906–3909.

[110] **THANGAVEL K., KARNAN M.** (2005), "Computer Aided Diagnosis in Digital Mammograms", Detection of Microcalcifications by Meta Heuristic Algorithms. GVIP Journal 5(7), 41–55.

[111] **LI H. et. al.** (2001), " Computerized Radiographic Mass Detection C Part I: Lesion Site Selection by Morphological Enhancement and Contextual Segmentation", IEEE Transactions on Medical Imaging 20(4), 289–301.

[112] **BALL J.E., BRUCE L.M.** (2007), "Digital Mammographic Computer Aided Diagnosis (CAD) using Adaptive Level Set Segmentation", In: Proceedings of the 29th Annual International Conference of the IEEE EMBS, Cité Internationale, Lyon, France, August 23- 26, pp. 4973–4978.

[113] **EONARDO O. M.** et. al. (2009), "Detection of Masses in Digital Mammograms using K-means and Support Vector Machine", Federal University of Maranhão, Av. dos Portugueses, SN, Campus do Bacanga, 65085-580, São Luís, MA, Brazil.

[114] **ELTONSY N.** et. al. (2007), "A concentric morphology model for the detection of masses in mammography", Medical Imaging, IEEE Transactions on 26 (6) (2007) 880–889.

[115] **DOMÍNGUEZ A. R., NANDI A. K.** (2009), "Toward breast cancer diagnosis based on automated segmentation of masses in mammograms", Pattern Recognition 1138 – 1148.

[116] **BELLOTTI R.** (2006), "A completely automated CAD system for mass detection in a large mammographic database", *Medical Physics* 33 (2006) 3066.

[117] **CHAN H.P. J.** et. al. (2005), "Computer-aided Detection System for Breast Masses on Digital Tomosynthesis Mammograms", Preliminary Experience, *Radiology* (2005) 2373041657.

[118] **MARTINS L. O.** et. al. (2009), "Detection of breast masses in mammogram images using growing neural gas algorithm and ripley's k function", *J. Signal Process. Syst.* 55 (1-3) (2009) 77–90.

[119] **CAMPANINI R.** et. al. (2004), "A novel featureless approach to mass detection in digital mammograms based on support vector machines", *Physics in Medicine and Biology* 49 (2004) 961–975.

[120] **MARTINS L. O.** et. al. (2006), "Classification of Normal, Benign and Malignant Tissues using Co-Occurrence Matrix and Bayesian Neural Network in Mammographic Images", *Proceedings of the Ninth Brazilian Symposium on Neural Networks* (2006) 479–486.

[121] **COSTA D. D.** et. al. (2007), "Independent component analysis in breast tissues mammograms images classification using lda and svm.", *Information Technology Applications in Biomedicine - ITAB2007 - Tokyo. Conference on 6th International Special Topic* (2007) 231–234.

[122] **CAMPOS L. F. A.** et. al. (2005), "Diagnosis of breast cancer in digital mammograms using independent component analysis and neural networks", *XI Bero-American Conference on Pattern Recognition, Havana, Lecture notes in computer science* 3773 (2005) 460–469.

[123] **JIANG J.** et al. (2005), "Integration of fuzzy logic and structure tensor towards mammogram contrast enhancement", Computer Medical Imaging and Graphics, Vol. 29, no. 1, 2005, pp. 83–90.

[124] **BALL J.E., BRUCE L.M.** (2007), "Digital Mammogram Spiculated Mass Detection and Spicule Segmentation using Level Sets", In: Proceedings of the 29th Annual International Conference of the IEEE EMBS, Cité Internationale, Lyon, France, August 23-26, pp.4979–4984.

[124] **THANGAVEL K.** (2005), "Automatic Detection of Microcalcification in Mammograms", ICGST International Journal on Graphics, Vision and Image Processing, pp. 31-61.

[125] **WANGA Z.** et. al. (2013), "Image fusion scheme for differential phase contrast mammography".

[126] **AL-ASADI A. and Al-SAADi A. K.** (2012), "Computer Aided Diagnosis For Microcalcification In Breast Mammograms", Basrah University, Basrah, Iraq.

[127] **JUN W.** et. al. (2008), "Computer Aided Detection Systems for Breast Masses: Comparison of Performances on Full-Field Digital Mammograms and Digitized Screen-Film Mammograms".

[128] Was available at: http://cdiok.com/?page_id=12

[129] Was available at: <http://radiology.rsna.org/cgi/content/abstract/2373041657v1>

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, name: ALSARORI, FAOZIA ALI SAIF

Nationality: YEMENI

Date and place of birth: 1978, YEMEN / TAIZ

Marital state: Single

Phone: +90 544 483 95 77 +966 502 624 291

Email: eng.fooz_alsourory@hotmail.com, c1171007@student.cankaya.edu.tr

EDUCATION

| Degree | Institution | Year of Graduation |
|--------|--------------------------------------------------------------|--------------------|
| MS | Çankaya Univ. Graduate School of Natural And Applied Science | 2013 |
| BS | University of Science and Technology | 2004 |

WORKEXPERIENCE

| Year | Place | Position |
|------------|-------------------|-------------------------|
| 2005- 2010 | Sana'a University | Teaching Assistant |
| 2005- 2010 | Many | Computer Skills Trainer |

FOREIGN LANGUAGES

Arabic, English and Turkish

HOBBIES

Reading, Writing prose, Traveling, Meeting new people, Being with family and friends, Computing, Swimming , Yoga, Listening to music and Collecting Souvenirs.