

COMPUTER AIDED DIAGNOSIS OF DIGITAL MAMMOGRAMS

by

Eng. Wael Abdel-Rahman Mohamed Ahmed

Electrical Engineering Department
High Institute of Technology, Benha University

A Thesis Submitted to the
Faculty of Engineering, Cairo University

In Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

in
SYSTEMS AND BIOMEDICAL ENGINEERING

FACULTY OF ENGINEERING, CAIRO UNIVERSITY

GIZA, EGYPT

DECEMBER 2009

COMPUTER AIDED DIAGNOSIS OF DIGITAL MAMMOGRAMS

by

Eng. Wael Abdel-Rahman Mohamed Ahmed

Electrical Engineering Department

High Institute of Technology, Benha University

A Thesis Submitted to the

Faculty of Engineering, Cairo University

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in

SYSTEMS AND BIOMEDICAL ENGINEERING

Under the Supervision of

Prof. Dr. Yasser Mostafa Kadah Dr. Mohammed Ibrahim Ismail Owis

Systems and Biomedical Engineering Department

Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY

GIZA, EGYPT

DECEMBER 2009

COMPUTER AIDED DIAGNOSIS OF DIGITAL MAMMOGRAMS

by

Eng. Wael Abdel-Rahman Mohamed Ahmed

Electrical Engineering Department

High Institute of Technology, Benha University

A Thesis Submitted to the

Faculty of Engineering, Cairo University

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in

SYSTEMS AND BIOMEDICAL ENGINEERING

Approved by the

Examining Committee:

Prof. Dr. Yasser Mostafa Kadah, Main Thesis Advisor

Prof. Dr. Abdalla Sayed Ahmed Mohamed, Member

Prof. Dr. Samia Abdel-Razik Mashali, Member

FACULTY OF ENGINEERING, CAIRO UNIVERSITY

GIZA, EGYPT

DECEMBER 2009

Table of Contents

LIST OF FIGURES	VIII
LIST OF TABLES	XI
LIST OF ABBREVIATIONS	XII
ACKNOWLEDGMENT	XIV
ABSTRACT	XV
1. INTRODUCTION.....	1
1.1 OVERVIEW OF THE THESIS	1
1.2 OBJECTIVES OF THE THESIS	3
1.3 ORGANIZATION OF THE THESIS.....	3
2. BREAST CANCER AND CAD OVERVIEW	5
2.1 INTRODUCTION.....	5
2.2 BREAST CANCER STATISTICS.....	6
2.3 EARLY DETECTION OF CANCER	9
2.4 MAMMOGRAPHIC ABNORMALITIES	10
2.4.1 Asymmetry	10
2.4.2 Architectural Distortion.....	10
2.4.3 Calcification.....	11
2.4.4 Mass.....	14
2.5 COMPUTER-AIDED DETECTION AND DIAGNOSIS FOR MAMMOGRAPHY ...	16
2.5.1 Background:.....	16
2.5.2 Detection versus Diagnosis.....	17
2.6 ROC AND FROC FOR CAD SYSTEMS PERFORMANCE EVALUATION	18
2.7 IMAGING MODALITIES	20
2.7.1 Mammography.....	20
2.7.2 Ultrasonography	20
2.7.3 Magnetic Resonance Imaging.....	21

2.7.4 Other tests used for breast cancer	21
2.8 COMMERCIAL CAD SYSTEMS IN MAMMOGRAPHY	22
3. CAD FOR MASS DETECTION	25
3.1 INTRODUCTION.....	25
3.2 REVIEW OF MASS DETECTION.....	25
3.3 FRACTAL BASED CAD SYSTEM.....	28
3.3.1 Fractal Analysis	29
3.4 METHODOLOGY	31
3.4.1 DDSM Mammographic database.	31
3.4.2 Extraction of ROI.	38
3.4.3 Feature extraction.	39
3.4.4 Feature selection.	47
3.4.5 Classification	48
3.4.5.1 Minimum Distance Classifier	49
3.4.5.2 Voting K-Nearest Neighbor (K-NN) classifier.....	49
3.4.6 Evaluation for the Classification	50
3.5 RESULTS.....	53
3.6 SUMMARY	54
4. FRACTAL MODELING FOR MICROCALCIFICATIONS	
DETECTION.....	56
4.1 REVIEW OF MICROCALCIFICATIONS DETECTION	56
4.2 THEORETICAL BACKGROUND	62
4.2.1 What is “Fractals”?	62
4.2.2 Self-similarity	63
4.2.3 Fractal Dimension.....	63
4.2.4 Why do scientists study fractals?.....	64
4.2.5 Fractal Modeling.....	64
4.2.6 IFS Coding Steps.....	66
4.3 ALGORITHM IMPLEMENTATION	66
4.3.1 Fractal Modeling:.....	68

4.3.2 Microcalcifications enhancement:.....	68
4.3.3 MIAS database:	69
4.3.4 ROI Selection:	71
4.3.5 Features Extraction	72
4.3.6 Classification	72
4.4 RESULTS & DISCUSSIONS.....	79
4.5 SUMMARY	80
5. UNSUPERVISED CLUSTERING/BICLUSTERING FOR ABNORMALITIES DETECTION	81
5.1 INTRODUCTION.....	81
5.2 REVIEW	82
5.3 PREPROCESSING	83
5.4 FEATURE EXTRACTION	83
5.4.1 Wavelet features	84
5.4.2 First order statistical features.....	84
5.4.3 Second order statistical features.....	85
5.4.4 Shape features.....	85
5.4.5 Fractal dimension features.....	85
5.5 FEATURE MATRIX VISUALIZATION	87
5.6 UNSUPERVISED CLUSTERING/BICLUSTERING TECHNIQUES	87
5.6.1 Clustering.....	87
5.6.1.1 Hierarchical clustering	89
5.6.1.2 Partitional clustering	89
5.6.2 Biclustering.....	92
5.6.2.1 Binary inclusion-maximal (Bimax) Algorithm.....	94
5.6.2.2 Cheng and Church (CC) algorithm	96
5.6.2.3 Order Preserving Sub-Matrix (OPSM) algorithm	96
5.6.2.4 Statistical Algorithmic Method for Bicluster Analysis (SAMBA)	97
5.6.2.5 Biclustering Toolboxes	98

5.7 RESULTS & DISCUSSIONS.....	102
5.7.1 Unsupervised Clustering Results.....	102
5.7.2 Biclustering techniques results	106
5.8 SUMMARY	117
6. CONCLUSIONS AND FUTURE WORK	118
REFERENCES	121

LIST OF FIGURES

Fig. 2.1: Incidence of breast cancer in selected countries in the eastern mediterranean region.....	7
Fig. 2.2: 10 leading causes of cancer deaths in Egypt according to the World Health Organization in 2005	8
Fig. 2.3: 10 leading causes of cancer deaths in Egypt according to the National Cancer Institute (NCI) in 2002-2004 period.....	8
Fig. 2.4: Types of breast cancer	11
Fig. 2.5: Mass descriptors for shape and margin.	16
Fig. 2.6: The role of computer-aided interpretation in breast cancer screening .	17
Fig. 2.7: ROC curve	19
Fig. 2.8: FROC curve	19
Fig. 3.1: Block diagram of the CAD system.....	29
Fig. 3.2: An example of the files in the directory for any provided case in the DDSM database	32
Fig. 3.3: An example of the ".ics" files	33
Fig. 3.4: Cancerous case left CC.....	34
Fig. 3.5: Cancerous case left MLO	34
Fig. 3.6: Normal case right CC	35
Fig. 3.7: Normal case right MLO.....	35
Fig. 3.8: An example of the ".OVERLAY" files	36
Fig. 3.9: The chain code for abnormalities boundary	37
Fig. 3.10: A ".16_PGM" image.....	38
Fig. 3.11: Acquiring normal regions from cancerous images.....	38
Fig. 3.12: Triangular prism surface area method algorithm	47

Fig. 3.13: Confusion matrix.	51
Fig. 4.1: Two famous examples for fractal images.....	62
Fig. 4.2: Creation of fractal object levels.....	63
Fig. 4.3: Self-similarity property of fractal images.....	63
Fig. 4.4: Fractal IFS coding example	66
Fig. 4.5: Shade and non-shade blocks of images	67
Fig. 4.6: Examples of mammographic images from MIAS database	71
Fig. 4.7: Mammograms obtained from MIAS database and the selected ROI...	71
Fig. 4.8: Classification of two classes using SVM classifier	74
Fig. 4.9: Example of how the fractal coding system works.....	80
Fig. 5.1: The features matrix	86
Fig. 5.2: Illustration of the bimax algorithm.....	95
Fig. 5.3: Graphical user interface of the BicAT software.....	99
Fig. 5.4: Graphical user interface of the EXPANDER software.	100
Fig. 5.5: Graphical user interface of the BicOverlapper software.	102
Fig. 5.6: Feature matrix heatmap for normal and mass images.	103
Fig. 5.7: Feature matrix heatmap for normal and microcalcification images...	103
Fig. 5.8: k-means clustering results for normal and mass images.	104
Fig. 5.9: k-means clustering results for normal and microcalcification.	104
Fig. 5.10: Feature matrix heatmap for mass images (benign & malignant).	105
Fig. 5.11: Feature matrix heatmap for microcalcification images (benign & malignant).	105
Fig. 5.12: Bicluster number 4 obtained from SAMBA algorithm of mass images	107

Fig. 5.13: Bicluster number 2 obtained from SAMBA algorithm of mass and microcalcification images	108
Fig. 5.14: Bicluster number 11 obtained from SAMBA algorithm of mass and microcalcification images	109
Fig. 5.15: Bicluster number 15 obtained from SAMBA algorithm of mass and microcalcification images	109
Fig. 5.16: Bicluster number 1 obtained from SAMBA algorithm of normal and microcalcification images	110
Fig. 5.17: Bicluster number 6 obtained from SAMBA algorithm of normal and microcalcification images	111
Fig. 5.18: Bicluster number 14 obtained from SAMBA algorithm of normal and microcalcification images	112
Fig. 5.19: An example of Overlapper for five biclusters from Bimax biclustering analysis of mass matrix	113
Fig. 5.20: An example of Overlapper for ten biclusters from CC biclustering analysis of mass matrix	114
Fig. 5.21: An example of Overlapper for seventeen biclusters from OPSM biclustering analysis of mass matrix	115
Fig. 5.22: An example of Overlapper for nine biclusters from SAMBA bi-clustering analysis of mass matrix	116

LIST OF TABLES

Table 3.1: Summary of results for mass detection.....	53
Table 4.1: Formulation of kernel functions	78
Table 4.2: Summary of fractal coding results	79
Table 5.1: Summary of the extracted features	86
Table 5.2: Summary of biclustering algorithms.....	94
Table 5.3: Biclustering toolboxes and their programming languages	98
Table 5.4: k-means clustering results for normal, mass and microcalcification	104
Table 5.5: k-means clustering results for benign, and malignant	106
Table 5.6: The number of biclusters produced by biclusters techniques	107

LIST OF ABBREVIATIONS

DCIS	: Ductal Carcinoma In Situ
ACR	: American College of Radiology
MRI	: Magnetic Resonance Imaging
FDA	: Food and Drug Administration
CAD	: Computer-Aided Diagnostic
CC	: Cranial-Caudal
MLO	: Medio-Lateral Oblique
FN	: False Negatives
FP	: False Positives
TP	: True Positive
TN	: True Negative
ROI	: Regions of Interest
2-D	: Two-Dimension
GLCM	: Gray Level Co-occurrence Matrix
ROC	: Receiver Operating Characteristic
FROC	: Free-response Receiver Operating Characteristic
FPpI	: False Positive per Image
ANN	: Artificial Neural Network
DDSM	: Digital Database for Screening Mammography
MIAS	: Mammographic Image Analysis Society
SVM	: Support Vector Machine
K-NN	: K-Nearest Neighbor
MC	: Microcalcification
RVM	: Relevance Vector Machine
LSSVM	: Least Square Support Vector Machine
FD	: Fractal Dimension
PMBC	: Piecewise Modified Box-Counting
PTPSA	: Piecewise Triangular Prism Surface Area
Bimax	: Binary inclusion-maximal

CC : Cheng and Church
OPSM : Order Preserving Sub-Matrix
SAMBA : Statistical Algorithmic Method for Bicluster Analysis
BicAT : Biclustering Analysis Toolbox
ISA : Iterative Signature Algorithm
EXPANDER : EXPression ANalyzer and DisplayER
GEM : Gene Expression Microarray

ACKNOWLEDGMENT

I would first like to thank ALLAH the Beneficent, the Merciful. Praise be to ALLAH, Lord of the Worlds. ALLAH who guides me along the way. And thanks to my supervisors, Prof. Dr. Yasser M. Kadah, and Dr. Mohammed I. Owis for their guidance, inspiration and wisdom during the preparation of this thesis. Also, I appreciate the great aid and support from all the members of the System and Biomedical Engineering Department, Cairo University.

Finally, I am indebted to thank my family for their continuous support, encouragement and understanding. Special thanks go to my wife, my guardian angel, for her support in all my life.

ABSTRACT

Breast cancer is the second cause of fatality among all cancers for women. Since the cause of breast cancer remains unknown, primary prevention becomes impossible. The most familiar breast tumors types are mass and microcalcification and their early detection is one of the key issues for breast cancer control.

Currently, X-ray mammography is the single most effective, low-cost, and highly sensitive technique for detecting small lesions. However, the sensitivity of mammography is highly challenged by the presence of dense breast parenchyma, which deteriorates both detection and characterization tasks. As the consequences of errors in detection or classification are costly, and since mammography alone cannot prove that a suspicious area is tumorous, malignant or benign, the tissue has to be removed for closer examination using breast biopsy techniques. Nevertheless, false-positive detection causes unnecessary biopsy. It has been estimated that only 15–30% of breast biopsy cases are proved to be cancerous. On the other hand, in false-negative detection an actual tumor remains undetected. Retrospective studies have shown that 10–30% of the visible cancers are undetected. Thus, there is a significant necessity for developing methods for automatic detection and classification of suspicious areas in mammograms, as a means of aiding radiologists to improve the efficacy of screening programs and avoid unnecessary biopsies.

Our goals of this thesis is to develop an approach for a Computer-Aided Diagnosis (CAD) system based on supervised classification that can be very helpful for radiologist in diagnosing breast cancers' patterns (mass and microcalcification) in digitized mammograms earlier and faster than typical screening programs. The proposed system has been implemented in four stages: (a) Region of interest (ROI) selection which identifies suspicion regions. (b) Feature extraction stage for locally processed image (ROI) to compute the important features of each breast cancer. (c) Feature selection stage to select the

most effective features that help discriminate between normal and abnormal patterns. (d) Classification stage, which classify between normal and abnormal patterns.

The proposed CAD system based on the fractal properties of the mammographic images. Where, in case of mammograms, when compared with microcalcifications, the breast background tissues have high local self-similarity, which is the basic property of fractal objects. Abnormalities (microcalcifications) may be enhanced by using the fractal modeling, where the enhanced image (from which background structures were removed) may be achieved by subtracting the fractal modeled image from the original one and ignoring the negative values which does not contain any information about spots brighter than background (microcalcifications).

Another new technique for CAD system in digital mammograms based on unsupervised classification and biclustering methods. This technique considers unlabeled data and provides unsupervised classes that give a better insight into classes and their interrelationships; moreover it does not need any training cost and the need to enlarge the mammography images database, thus improving the overall effectiveness of the diagnosis. This technique is also extended to utilize biclustering methods, which how the biclusters will match up with the known pathologies and allow for definition of unsupervised clusters of both pathologies and which features are relevant to these pathologies. The proposed system was shown to have the large potential for breast cancer diagnostic in digital mammograms and provide more flexibility, and hence better diagnostic accuracy, than the commonly used feature selection strategies.

CHAPTER 1

INTRODUCTION

1.1 Overview of the Thesis

Breast cancer is the uncontrolled growth of abnormal cells in the breast that starts in the breast, usually in the inner lining of the milk ducts or lobules. There are different types of breast cancer, with different stages (spread), aggressiveness, and genetic makeup. With best treatment, 10-year disease-free survival varies from 98% to 10%. Treatment includes surgery, drugs (hormone therapy and chemotherapy), and radiation. Breast cancer is the most common cancer and continues to be a significant public health problem among women around the world. Primary prevention seems impossible since the cause of this disease still remains unknown [1]. It is believed that the most promising way to decrease the number of patient suffering from the disease is by early detection. The earlier breast cancer is detected, the better the chances that treatment will work and the better a proper treatment options can be provided.

Mammography continues to be the standard screening tool for breast cancer detection resulting in at least a 30% reduction in breast cancer deaths [2]; however, not all breast cancer can be detected by mammograms such as microcalcification, masses, architectural distortion, asymmetry between breasts, breast edema and lymphadenopathy, the interpretations of their presence are very difficult because of its morphological features [3] and the sensitivity of mammography is highly challenged by the presence of dense breast parenchyma, which deteriorates both detection and characterization tasks [4]. As the consequences of errors in detection or classification are costly, and since mammography alone cannot prove that a suspicious area is tumorous, malignant or benign, the tissue has to be removed for closer examination using breast biopsy techniques.

Among the various types of breast abnormalities which are visible in mammograms, clustered microcalcifications (or ‘calcifications’) and mass

lesions, distortion in breast architecture, and asymmetry between breasts are the most dangerous ones. Masses and clustered microcalcifications often characterize early breast cancer that is detectable in mammograms before a woman or the physician can sense them. Masses appear as dense regions of varying sizes and properties and can be characterized as circumscribed, spiculated, or ill defined. On the other hand, microcalcifications appear as small bright arbitrarily shaped regions on the large variety of breast texture background. Finally, asymmetry, and architectural distortion are also very important and abnormalities are difficult to detect. The great variability of the mass appearance along with the other abnormalities in digital mammograms is the main obstacle of building a unified mass detection method [5].

Second reading of mammogram images would help the radiologist make his final decision. But; it may not be feasible to routinely perform a second reading for mammographic film by a radiologist due to financial, technical, and logistical restraints. Another method was performed using breast biopsy techniques. Nevertheless, false-positive detection causes unnecessary biopsy. It has been estimated that only 15–30% of breast biopsy cases are proved to be cancerous [5]. On the other hand, in false-negative detection an actual tumor remains undetected. Retrospective studies [5]-[7] have shown that 10–30% of the visible cancers are undetected. Thus, there is a significant necessity for developing methods for automatic classification of suspicious areas in mammograms, as a means of aiding radiologists to improve the efficacy of screening programs and avoid unnecessary biopsies.

Therefore, efforts were made to develop a computer-aided detection (CAD) system. CAD can be defined as a diagnosis made to improve radiologists' performance by indicating the sites of potential abnormalities, to reduce the number of missed lesions, and/or by providing quantitative analysis of specific regions in an image to improve diagnosis [8].

1.2 Objectives of the Thesis

The objective of this thesis is to address the needs of screening algorithms that are designed to aid in detecting and classifying abnormalities in digital mammograms. Specifically, this thesis seeks to provide and improve tools that are considered essential to the construction of a comprehensive screening system or to develop a Computer-Aided Diagnosis (CAD) system.

The proposed CAD system based on the fractal properties of the mammographic images. Where, in case of mammograms, when compared with microcalcifications, the breast background tissues have high local self-similarity, which is the basic property of fractal objects. Abnormalities (microcalcifications) may be enhanced by using the fractal modeling, where the enhanced image (from which background structures were removed) may be achieved by subtracting the fractal modeled image from the original one and ignoring the negative values which does not contain any information about spots brighter than background (microcalcifications). To detect the abnormalities in the mammograms we extracted an effective set of features from the image and after we used them in detection and classification of breast cancer patterns by using supervised classifiers, as well as unsupervised clustering methods; these techniques consider unlabeled data and provide unsupervised classes that give a better insight into classes and their interrelationships, thus improving the overall effectiveness of the diagnosis, and biclustering algorithms, which allow how the biclusters will match up with the known pathologies and which features are relevant to these pathologies. The proposed system was shown to have the large potential for breast cancer diagnosis in digital mammograms.

1.3 Organization of the Thesis

This thesis has been organized as follows:

Chapter 1: Provides an introduction and summary of the thesis as it contains a brief summary of the thesis and the purpose of it and presents how it is organized

Chapter 2: Contain a brief summary of breast cancer, and different types of it and also some statistics and existing technologies to detect this type of cancer, this chapter also contains an explanation of the use of X-ray mammography, as well as a brief on the use of computers to detect breast cancer with an indication of some commercial systems that already exist for this Purpose.

Chapter 3: Describes the main components of a proposed system for diagnosis using the computer to detect breast cancer with mention of some ways and researches that are used in the development of the computer-aided diagnosis system. This part focuses on the detection of the presence of masses in digital mammograms.

Chapter 4: This section provides a new method proposed for microcalcification detection based on the self-similarity property in those images, which is the basic property of the fractal coding algorithm provided. This chapter begins by citing some of the methods and techniques used to detect this kind followed by a detailed explanation of the new method used and some of the results reached by the researcher.

Chapter 5: Presents a new methodology for CAD system for breast cancer in digital mammograms using unsupervised clustering and biclustering algorithms.

Chapter 6: this chapter provides the conclusions drawn up from the thesis. It describes the main outcome of this thesis, and what more can be done in the future.

CHAPTER 2

BREAST CANCER AND CAD OVERVIEW

This chapter provides an introduction to breast cancer and a description of the screening methods and procedures of mammography. Examples of mammograms are presented with details. In addition, mammogram abnormalities are shown. This discussion is intended to provide sufficient background information and to demonstrate the need for computer aided diagnosis algorithms in breast cancer detection and classification.

2.1 Introduction

Breast cancer is one of the most significant public health problems in the world. It is a leading cause of fatality among all cancers for women in the 35 to 55 age group. Until now there is no known way to prevent breast cancer but the earlier the cancer is detected, the higher the chance of survival for patients. Mammography is the most effective method that is used in the early detection of breast cancer [9], [10].

It may not be feasible to routinely perform a second reading by a radiologist due to financial, technical, and logistical restraints. Therefore, efforts were made to develop a computer-aided detection (CAD) system [6], [7]. CAD can be defined as a diagnosis made to improve radiologists' performance by indicating the sites of potential abnormalities, to reduce the number of missed lesions, and/or by providing quantitative analysis of specific regions in an image to improve diagnosis.

Recently, computer-aided diagnosis (CAD) has become a part of the routine clinical work for detection of breast cancer on mammograms at many screening sites and hospitals in many countries. This seems to indicate that CAD is beginning to be applied widely in the detection and differential diagnosis of many different types of abnormalities in medical images obtained in various examinations by use of different imaging modalities. In fact, CAD has become

one of the major research subjects in medical imaging and diagnostic radiology. Although early attempts at computerized analysis of medical images were made in the 1960s, serious and systematic investigation on CAD began in the 1980s with a fundamental change in the concept for utilization of the computer output, from automated computer diagnosis to computer-aided diagnosis [11].

With CAD, radiologists use the computer output as a “second opinion,” and radiologists make the final decisions. Therefore, for some clinical cases in which radiologists are confident about their judgments, radiologists may agree with the computer output, or disagree and then disregard the computer. However, for cases in which radiologists are less confident, it is expected that the final decision can be improved by use of the computer output. This improvement is possible, of course, only when the computer result is correct. The higher the performance of the computer, the better the overall effect on the final diagnosis. However, the performance level of the computer does not have to be equal to or higher than that of radiologists. With CAD, the potential gain is due to the synergistic effect obtained by combining the radiologist’s competence and the computer’s capability. Because of these multiplicative benefits, the current CAD has become widely used in practical clinical situations.

2.2 Breast Cancer Statistics

Cancer is an important factor in the global burden of disease. The estimated number of new cases each year is expected to rise from 10 million in 2002 to 15 million by 2025, with 60% of those cases occurring in developing countries. Breast cancer is the most common cancer in women in the Eastern Mediterranean Region and the leading cause of cancer mortality worldwide. There is geographic variation, with the standardized age incidence rate being lower in developing than industrialized countries [12].

During the past two decades significant demographic changes have taken place in the Eastern Mediterranean Region. The progressive decline in the crude death rate, increasing life expectancy, urbanization and changes in lifestyle

associated with economic transition have resulted in an increase in non-communicable diseases.

There is now sufficient evidence to indicate that cancer is becoming a major health concern for many countries within the Eastern Mediterranean Region, although there is considerable variation in the types and incidence of cancers, mostly related to age distribution, and environmental and lifestyle changes. Among cancers in the female population of the Region, breast and, in a few countries cervical, cancers lead in the incidence of mortality and morbidity.

It is important to have accurate and updated census data on cancer-specific mortality and incidence. There are no significant data to indicate the incidence of breast cancer based on geographical distribution, but the age-standardized incidence of breast cancer is 12–50 per 100 000 women, with the lowest incidence in the Islamic Republic of Iran and Pakistan. A higher incidence of breast cancer (50/100 000) is seen in Middle Eastern and North African countries. However, the relative frequency of breast cancer in the majority of the countries in the Region is between 15% and 25% of all cancers diagnosed see Fig. 2.1 [12].

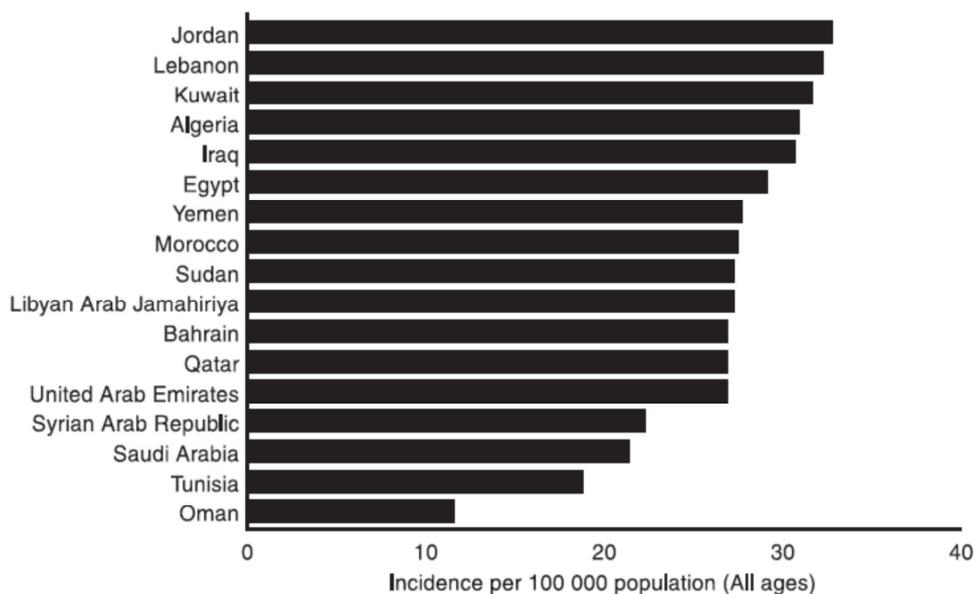


Fig. 2.1: Incidence of breast cancer in selected countries in the Eastern Mediterranean Region and Algeria [12].

According to the World Health Organization, breast cancer is the leading cause of cancer deaths in women in Egypt [13]. In 2005 cancer killed approximately 42,000 people in Egypt 31,000 of those people were under the age of 70. The 10 leading causes of cancer deaths in Egypt are shown in Fig. 2.2 and Fig. 2.3 according to the World Health Organization and the National Cancer Institute [14] respectively.

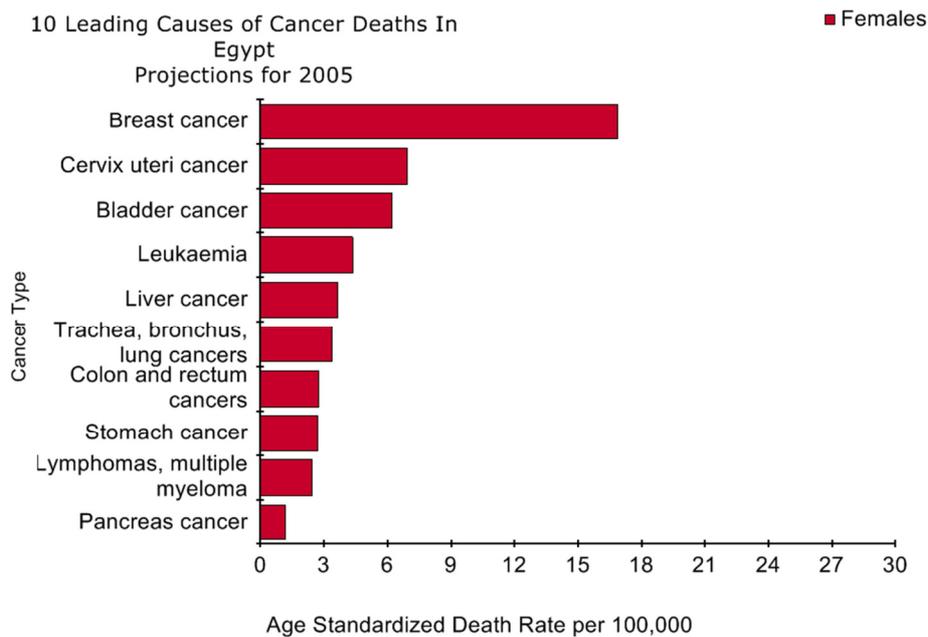


Fig. 2.2: 10 leading causes of cancer deaths in Egypt according to the World Health Organization in 2005 [13].

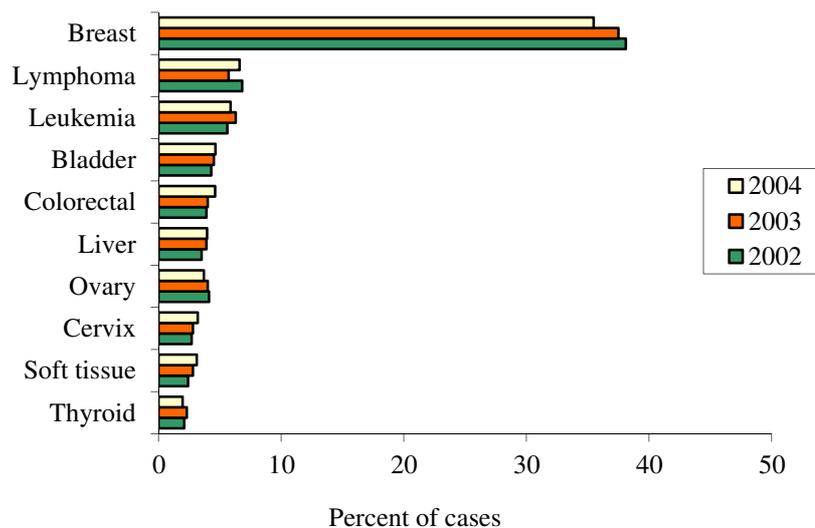


Fig. 2.3: 10 leading causes of cancer deaths in Egypt according to the National Cancer Institute (NCI) in 2002-2004 period [14].

2.3 Early detection of cancer

Survival rates are significantly higher when the cancer is detected at an early stage [15]-[17]. The 5-year relative survival for female breast cancer patients has improved from 63% in the early 1960s to 89% today. The survival rate for women diagnosed with localized breast cancer (malignant cancer that has not spread to lymph nodes or other locations outside the breast) is 98% [18]. Clearly, detecting breast cancer at an early stage is critical to patient care.

The most common and effective early-detection tool currently available to clinicians is screening mammography. In fact, half of the cancers detected in screening mammography are impalpable. Studies have shown that mammography is the only screening program proven to reduce mortality [18]. Mammography is also inexpensive and widely available.

The early detection of breast cancer is the key to successful treatment. The primary means of screening for breast cancer is by means of mammography [19]. If cancer is detected a woman is usually required to undergo further testing which may include:

- ✓ An ultrasound scan of the breast
- ✓ Fine core needle aspiration – using a local anesthetic, cells are drawn up through a needle that is inserted through the skin of the breast into the suspicious lesion
- ✓ Core biopsy – using a local or general anesthetic, a sample of tissue is taken from the suspicious area of the breast
- ✓ Diagnostic open biopsy – a diagnostic (surgical) biopsy performed with a needle localization technique

When breast cancer is confirmed, treatment involves management of the breast and systemic therapy. Management of the breast involves either removal of the lump (lumpectomy), normally followed by radiation therapy to the breast, or removal of the entire breast (mastectomy). Systemic therapy involves such techniques as chemotherapy or Tamoxifen (a drug used to treat breast cancer) [19].

In a screening mammographic examination, the breast is compressed before imaging. There are normally two views examined: craniocaudal (CC), which is from the top down, and mediolateral-oblique (MLO), which is from the side. These views normally allow a radiologist to localize a mass to a certain region of the breast. To ensure high contrast at a small dose, the tube setting for mammography is between twenty-three and twenty-eight kilovolts peak. Contrast is also increased by the use of low-ratio grids. Molybdenum is commonly used as the x-ray target and filter.

2.4 Mammographic Abnormalities

Mammography is used to detect a number of features that may indicate a potential clinical problem, which include asymmetries between the breasts, architectural distortion, calcifications and masses [3], [19].

2.4.1 Asymmetry

Breast asymmetry exhibits as breast tissue that is greater in volume or denser in one breast than the other. This may be the result of either a greater volume of fibroglandular tissue on one side, or asymmetrically dense breast tissue. The latter is a term reserved to denote the broad regions of dense breast tissue that do not form masses, but are distinctly different from the corresponding contralateral regions of tissue. The morphology of the two regions is similar except that there is an increase in the tissue density in the mammogram involved. Variations may be the result of natural differences between corresponding left and right breasts or decreased density in one of the mammograms as a result of the surgical removal of breast tissue. The vasculature of the breast is generally symmetrical in size and distribution; therefore an asymmetrically large vein may also indicate the presence of an abnormality see Fig. 2.4.e [20].

2.4.2 Architectural Distortion

The structures of the breast, comprising the glandular tissue, i.e. lobules, ductules, lobes and ducts converge toward the nipple. Disturbances in this

symmetrical flow, i.e. pulling of structures toward a point eccentric from the nipple, are the sign of a potential abnormality.

An architectural distortion is defined as follows: “The normal architecture is distorted with no definite mass visible. This includes spiculations radiating from a point, and focal retraction or distortion of the edge of the parenchyma” see Fig. 2.4.f [20].

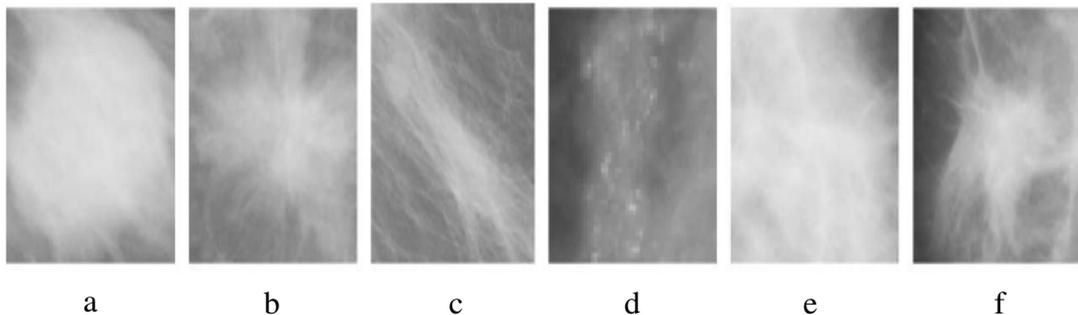


Fig. 2.4: Types of breast cancer, (a) circumscribed mass, (b) speculated mass, (c) ill-defined mass, (d) microcalcification, (e) asymmetry, and (f) architectural distortion [5].

2.4.3 Calcification

Calcifications are small mineral (calcium) deposits within the breast that appear as localized high-intensity regions (spots) in the mammogram. There are two types of calcifications: microcalcifications and macrocalcifications:

- Macrocalcifications are coarse (larger), scattered calcium deposits that are most likely changes in the breasts caused by aging of the breast arteries, old injuries, or inflammation. These deposits are related to non-cancerous conditions and usually do not require a biopsy. Macrocalcifications are found in about half the women over 50, and in 1 of 10 women under 50.
- Microcalcifications may be isolated, appear in clusters, or found embedded in a mass. Individual microcalcifications typically range in size from 0.1-1.0 mm with an average diameter of about 0.5 mm.

Microcalcifications seen on a mammogram are more of a cause for concern, but still usually do not mean that cancer is present. The shape and layout of microcalcifications help the radiologist judge how likely it is that cancer is present. In most cases, the presence of microcalcifications does not mean a biopsy is needed. But if the microcalcifications have a suspicious look and pattern, the radiologist may recommend a biopsy.

A cluster is typically defined to be at least three microcalcifications within a 1cm^2 region; the clusters are important cues for the mammography in determining if the reading is suspicious. About 30-50 % of non-palpable cancers are initially detected due to the presence of microcalcifications clusters [21]. Similarly, in a large majority of the ductal carcinoma in situ (DCIS) cancers, calcification clusters are present [22].

Most breast calcifications are benign. The term microcalcification is often used for calcifications found with malignancy, which are usually smaller, more numerous, clustered, and variously shaped (rods, branches, teardrops). Calcifications associated with benign conditions are usually larger, fewer in number, widely dispersed and round. These are termed macro-calcifications. In the middle are hard-to-tell calcifications that are often labeled indeterminate. The number of calcifications that make up a cluster can be used as an indicator of benignity and malignancy. While the actual number itself is arbitrary, a minimum number of either four, five or six calcifications per cluster are considered to be of significance. The morphology of calcifications is considered to be the most important indicator in differentiating benign from malignant. As discussed earlier, round and oval shaped calcifications are more likely to be benign. Those associated with malignant processes resemble small fragments of broken glass and are rarely rounded or smooth.

Calcifications are analyzed according to their size, shape, number, and distribution. The general rule is that larger, round or oval shaped calcifications uniform in size has a higher probability of being associated with a benign process and smaller, irregular, polymorphic, branching calcifications heterogeneous in size and morphology are more often associated with a

malignant process. Certain calcification patterns are almost always pathognomic of a benign process, and in such cases no further analysis is needed. In the majority of cases, however, a pattern of calcification deposition is inconclusive and may be attributable to either a benign or malignant process. Needless to say, these cases require additional evaluation such as using magnification mammography to further elucidate the calcifications' morphology and distribution [23].

Size: Generally speaking, microcalcifications are associated with a malignant process and macrocalcifications are associated with a benign process. The problem with this general rule is that there is no fine line of measurement that could enable one to distinguish between micro and macro. All calcifications start out imperceptibly small and radiographically invisible. Most radiologists place calcifications 0.5 mm or less to have a high probability of association with cancer; and calcifications of 2.0 mm or larger are typical of a benign process. The smallest visible calcifications on a mammogram is approximately 0.2 - 0.3 mm [23].

Number: The number of calcifications that make up a cluster has been used as an indicator of benignity and malignancy. While the actual number itself is arbitrary, radiologists tend to agree that the minimum number of calcifications be either four, five, or six to be of significance. Any number of calcifications less than four will rarely lead to the detection of breast cancer in and of itself. Again, as with all criteria in mammographic analysis, no number is absolute and two or three calcifications may merit greater suspicion if they exhibit worrisome morphologies [23].

Morphology: The morphology of calcifications is considered to be the most important indicator in differentiating benign from malignant. As noted earlier, round and oval shaped calcifications that are also uniform in shape and size are more likely to be on the benign end of the spectrum. Calcifications that are irregular in shape and size fall closer to the malignant end of the spectrum. It has been described that calcifications associated with a malignant process resemble small fragments of broken glass and are rarely round or smooth [23].

2.4.4 Mass

Masses are three-dimensional lesions which may represent a localizing sign of breast cancer. A mass is a group of cells clustered together more densely than the surrounding tissue. A (non-cancerous) cyst may appear as a mass in a mammographic film. Masses can be caused by benign breast conditions or by breast cancer. The similarity in intensities with the normal tissue and in morphology with other normal textures in the breast makes it more difficult to detect masses compared with calcifications [21]. They are characterized by their location, size, shape, margin characteristics, x-ray attenuation, effect on surrounding tissue, and other associated findings like architectural distortion, associated calcifications and skin changes [24]. Depending on the morphologic criteria of the mass, the likelihood of malignancy can be established. These categories help radiologists to precisely describe masses found in mammograms and to classify masses as benign or potentially malignant:

Location: The location of the mass may be established from the physical examination if the mass is palpable. Otherwise, its location can be determined from several different mammographic views. It is important to realize that the mass seen on a mammogram may not correspond to a palpable lump. Because breast cancer tends to develop in the peripheral zone of the breast's parenchymal cone, a mass' location can raise suspicion of malignancy.

Size: Size alone does not predict malignancy. Nonetheless, the size of a malignant mass is indicative of its progression. Needless to say, the objective of mammography is to detect breast cancer in its earliest stage of development.

Shape: A mass shape may have one of five characteristics: Round, Oval, Lobular, Irregular, and Architectural distortion. The descriptions are fairly self-explanatory, and a schematic picture of each shape is shown in Fig. 2.5. Architectural distortion is not technically a mass since there is no definite mass visible. It can be identified by distortion in the normal breast architecture, including spiculations radiating from a point and focal retraction or distortion of the parenchyma edge. Architectural distortion can also be an associated finding of a mass [25].

Margin: The margin is the border of a mass, and it should be examined carefully, sometimes using magnification view for clarity. It is one of the most important criteria in determining whether the mass is likely to be benign or malignant. There are five types of margins: Circumscribed, Obscured, Micro-lobulated, Ill-defined, and Spiculated shown in Fig. 2.5. Circumscribed margins are well defined and sharply demarcated with an abrupt transition between the lesion and the surrounding tissue. Micro-lobulated margins have small undulating circles along the edge of the mass. Obscured margins are hidden by superimposed or adjacent normal tissue. Ill-defined margins are poorly defined and scattered. Spiculated margins are marked by radiating thin lines. If there is no visible mass, the basic description of architectural distortion with spiculation as a modifier is used [25].

Most benign masses are well circumscribed, compact, and roughly circular or elliptical [26]. Malignant lesions usually have a blurred boundary, an irregular appearance, and sometimes are surrounded by a radiating pattern of linear spicules. However, some benign lesions may have a spiculated appearance or blurred periphery.

X-ray attenuation: X-ray attenuation is a description of the density of the mass. Generally speaking, breast cancer often appears denser (whiter) than the surrounding normal breast parenchyma.

Effect on surrounding tissues and associated finding: These are descriptions associated with the mass such as architectural distortion, enlarged duct, skin changes, nipple and areolar abnormalities, etc.

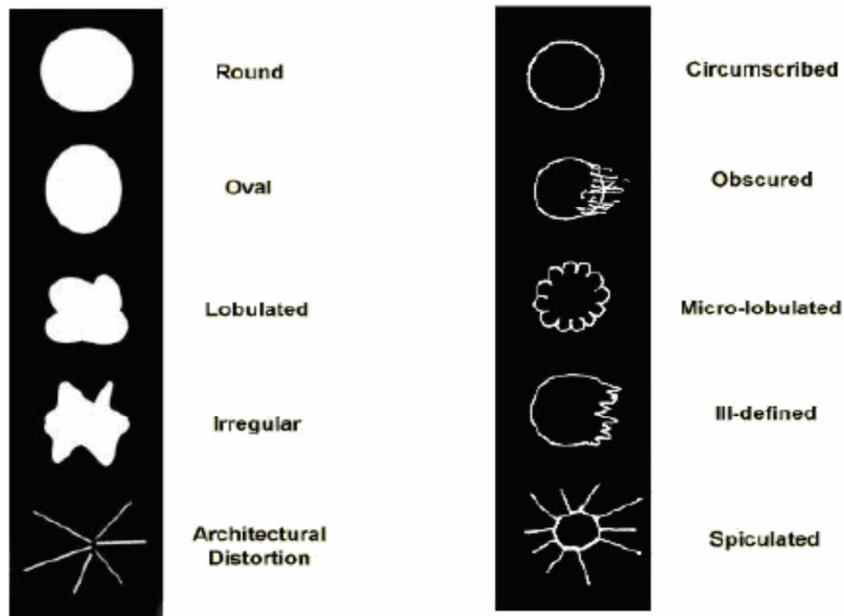


Fig. 2.5: Mass descriptors for shape (left), and margin (right).

2.5 Computer-aided Detection and Diagnosis for Mammography

2.5.1 Background:

To aid radiologists in the interpretation of mammograms, research has been directed towards developing computer-aided detection and computer-aided diagnosis tools. Mammograms are read more accurately when read by more than one radiologist; unfortunately, having multiple radiologists read the same mammograms is neither time- nor cost-efficient. CAD systems have been demonstrated to serve as a reliable, accurate, and efficient second reader to aid radiologists [27]-[38].

Radiologists are typically proficient at extracting features from mammograms; however, not all radiologists are equal in how well they combine these features to make an accurate diagnosis. CAD algorithms have shown promise at merging image- and radiologist-derived features into accurate decisions [39]. Fig. 2.6 outlines the role CAD plays in the overall context of breast cancer screening.

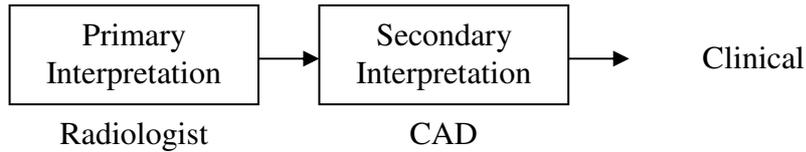


Fig. 2.6: The role of computer-aided interpretation in breast cancer screening

2.5.2 Detection versus Diagnosis

When first introduced, CAD was an acronym for computer-aided diagnosis. In the current literature, the same acronym is used to describe computer-aided detection algorithms as well. The dual meaning of CAD is quite descriptive of the evolving nature of this field. Some CAD systems perform only the detection task; that is, they search through medical images and identify regions that contain specific abnormalities - they do not distinguish the presence or absence of malignant disease. Other algorithms do not perform the detection task; they are designed to classify manually-identified lesions as benign or malignant.

Systems designed to perform detection and diagnosis, both utilize two general operations: feature extraction and classification. In detection algorithms, there are additional sections pre-pended to the process: image preprocessing and filtration (pattern matching) to identify the suspicious areas.

Similarly, classification mechanisms can be the same for detection and diagnosis algorithms, yet their goals are different. In detection algorithms, the classification stage differentiates between abnormalities and normal tissue. In the case of screening mammography, there are many more examples of normal tissue than abnormalities. The main purpose of the classification stage in a detection algorithm is to reduce false positive detections (i.e., identify normal regions as being abnormal) without sacrificing the detection of abnormal lesions. The performance metrics for detection algorithms are sensitivity and false positives per image (FPpI), where:

$$sensitivity = \frac{\text{number of abnormal objects correctly identified}}{\text{total number of abnormal objects}}, \quad (2.1)$$

and

$$FPpI = \frac{\text{total number of false positives}}{\text{total number of images}}. \quad (2.2)$$

The above definition of sensitivity measures the proportion of abnormal lesions correctly identified; i.e., the sensitivity is measured on a per-lesion basis. In most instances, CAD systems that measure sensitivity on a per-lesion basis require that the lesion be detected in either the CC or MLO view image, but not necessarily both. In some CAD systems, however, the sensitivity is defined on a per-case basis: no matter how many lesions may be present in a single case, the detection of any of the lesions counts as detection for the sensitivity calculation.

Diagnosis-driven classification algorithms decide whether a lesion is benign or malignant. In this setting, because there is no detection component, FPpI is not usually a metric of concern. In classification algorithms, the relevant metrics are sensitivity and specificity, where:

$$\text{specificity} = \frac{\text{number of normal objects correctly identified}}{\text{total number of abnormal objects}}. \quad (2.3)$$

The CAD algorithms that merge detection and diagnosis must therefore be concerned with both false positive reduction and diagnostic performance.

2.6 ROC and FROC for CAD Systems Performance

Evaluation

After calculating sensitivity, specificity, and FPpI, a plot of sensitivity versus 1-specificity is called a Receiver Operating Characteristic (ROC) curve and this is generally used to report the performance of the diagnosis algorithm. An example of an ROC curve is shown in Fig. 2.7. It is important to note that the ROC methodology can be correctly applied in classification tasks where localization of the abnormality is not an issue like in the diagnosis task described above. However, for tasks where localization is an important issue the ROC

methodology has some inherent problems as it does not require correct localization of the abnormality. Also the ROC does not apply to situations where the radiologist has to detect and localize multiple lesions on the same image. For these situations the FROC curve should be used to report performance. The Free-Response Receiver Operating Characteristic (FROC) plot is a plot of sensitivity versus FPpI, and this is generally used to report the performance of the detection algorithm as seen in Fig. 2.8.

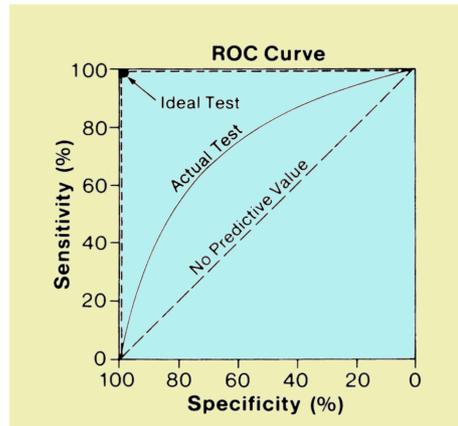


Fig. 2.7: In a ROC curve, sensitivity is plotted on the y-axis and 1-Specificity or False Positive Fraction FPF is plotted along the x-axis. The dotted line in the ROC curve represents chance performance.

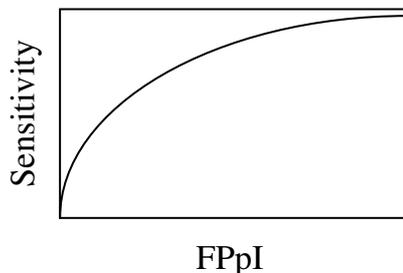


Fig. 2.8: In a FROC curve, sensitivity is plotted on the y-axis and the number of FPpI is plotted along the x-axis.

Until recently, FROC analysis has been limited by the fact that the statistical analysis of FROC curves was less developed than that of traditional Receiver Operating Characteristic (ROC). Major advances have recently been made in FROC analysis, particularly by Chakraborty *et al.* [40]. However, despite the consistent use of evaluation methods in the literature, direct comparison of systems for detecting mammographic abnormalities is difficult because few studies have been reported on a common database.

2.7 Imaging modalities

The evaluation of breast cancer includes at least one imaging study [41]. What are the modalities for imaging breast cancer and when are they used?

Breast cancer is one of the most common conditions among women in the United States. The importance of screening cannot be stressed enough. Through clinical evaluation, imaging studies, and biopsies as needed, the chances of detecting breast cancer early are greater. Many topics related to breast cancer deserve consideration. Among them are the methods of breast cancer imaging.

2.7.1 Mammography

A mammogram is essentially an x-ray of the breast [41], [42], used for screening breast cancer in asymptomatic women and diagnosing breast cancer in women who have symptoms. During the test, the breast is compressed in order to minimize x-ray scatter and maximize image quality. This may be uncomfortable but not necessarily painful. From there, x-ray images are shot at different angles.

The radiologist reviews the images to look for abnormalities like masses, densities, and structural irregularities. Calcifications, or soft tissue hardening with calcium deposits, are especially important. They are often an early sign of breast cancer, especially if the calcifications are small (microcalcifications) or irregularly shaped. Calcifications appear bright on x-ray imaging, one reason why mammograms are the standard tool for breast cancer screening.

The study does have some limitations. Imaging is more difficult with breasts that are dense or breasts in younger women. Breasts with implants or significant surgical scars are also difficult to visualize on mammography. Nevertheless, mammography is recommended for breast cancer screening starting at age 40 and for diagnosing suspected breast cancer as indicated.

2.7.2 Ultrasonography

An ultrasound of the breast is currently used as a diagnostic tool [41], [43]. If a physician notes a lump or other suspicious finding on a clinical breast exam, he

or she may evaluate further with an ultrasound. This can tell if the abnormality is a hollow cyst or something solid and if it has malignant characteristics like irregular shape and calcifications. Ultrasound is also used as an imaging guide during a needle biopsy of a suspicious breast mass.

Ultrasound as a means of screening breast cancer is under investigation. Challenges exist that hinder the acceptance of ultrasound screening. The technique may require proficient skill and optimal imaging. There is also a risk of missing microcalcifications that can show up on mammography but not on ultrasound. Until studies demonstrate equivalent or greater efficacy than mammography, ultrasound is not recommended as a breast cancer screening tool.

2.7.3 Magnetic Resonance Imaging

MRI of the breast has been explored and improved over the years [41], [44]. It has the advantages of flexible angles of visualization and not using ionizing radiation. However, it is an expensive modality with questionable screening capabilities. It can still help diagnose breast cancer, but usually in conjunction with mammography and not alone. MRI has high sensitivity approaching 98%, but it has moderately low specificity. MRIs may depict many abnormalities that are later proved not to be cancer. Like breast ultrasound, MRI for screening breast cancer is being researched in ongoing studies.

2.7.4 Other tests used for breast cancer

2.7.4.1 Electrical impedance imaging (T-scan)

Electrical impedance imaging scans the breast for electrical conductivity, based on the idea that breast cancer cells conduct electricity better. It involves passing a very small electrical current through the body and detecting it on the skin of the breast with a small probe (similar to an ultrasound probe). The test does not use radiation and does not require breast compression. This test has received approval by the US Food and Drug Administration to be used as a

diagnostic aid to mammography. However, it has not undergone enough clinical testing to recommend its use in breast cancer screening.

2.7.4.2 Nuclear medicine imaging (scintimammography)

Although not indicated as a screening procedure for the detection of breast cancer, scintimammography may play a useful and significant role in various specific clinical indications, as in cases of non-diagnostic or difficult mammography and in the evaluation of high-risk patients, tumor response to chemotherapy, and metastatic involvement of auxiliary lymph nodes.

2.7.4.3 Other tests under investigation include the following:

- Thermography (thermal imaging) and computerized thermal imaging: These depend on mapping heat radiating from the breast, with the assumption that cancerous tissue produces more heat than normal breast tissue. It is not approved of as a screening tool for breast cancer.
- Computed tomography laser mammograms: This is an experimental test that uses a laser to produce a 3-dimensional view of the breast. It has not yet been approved by the Food and Drug Administration for clinical use.

Finally, we can state that, Mammography continues to be the standard screening tool for breast cancer, but that does not mean it will forever be this way. If future research demonstrates superior imaging methods of screening and diagnosing breast cancer, women can expect a dramatic change in their preventive care.

2.8 Commercial CAD systems in mammography

The practice of mammography is regulated in United States by the Food and Drug Administration (FDA) under the authority of the Mammography Quality and Standards Act of 1992 [45]. So far, the FDA has approved only three commercially available CAD systems to aid radiologists in detecting mammographic abnormalities [4].

ImageChecker (R2 Technology Inc., Los Altos, CA) was the first commercial CAD system approved by the FDA. This device is designed to search for all types of signs that may be associated with breast cancer. The detection accuracy

of microcalcifications was reported as 98.5% sensitivity at 0.74 false positive detections per case (set of four images). The detection accuracy of masses was reported as 85.7% at 1.32 false positive marks per case [46].

MammoReader (Intelligent Systems Software Inc., Clearwater, FL) was designed to detect primary signs of breast cancer in mammogram images including clusters of microcalcifications, well- and ill-defined masses, spiculated lesions, architectural distortions, and asymmetric densities. The reported overall sensitivity was 89.3% (91 % in cases where microcalcifications were the only sign of cancer and 87.4% in the remaining cases where malignant masses were present). The system made (on average) 1.53 true positive marks and 2.32 false positive marks per case among cancer cases and 3.32 false positive marks among cases without cancer [47].

SecondLook (CADx Inc., Nashua, NH) was the third commercial device to receive an FDA approval. The system was designed to mark areas of a mammogram that are indicative of cancer. The sensitivity of the system was reported to be 85% for screening-detected cancers (combination of masses and microcalcification clusters) [48].

In brief, there are several companies developing commercial CAD systems and software to provide radiologists with the technology to help them interpret mammogram Films. *Imagechecker* CAD System made by R2 Technology, *MammoReader* CAD system made by ISSI Inc., and *Second Look* made by CADx Inc. (FDA approved). Also there are *Mammex Tr* made by Scanis Inc., and *ImageClear* made by Titan Systems Corp.'s DBA Systems Division. Most of the systems mentioned above use neural networks to achieve their goals. However, since these commercial systems are proprietary, very little information about the methodologies or algorithms they use is made public in spite of very lengthy brochures and publications discussing the benefits of their products.

A recent major study by Dr. Matthew Gromet of the Breast Imaging Section of Charlotte Radiology, compared the recall rate, sensitivity, positive predictive value, and cancer detection rate for single reading with CAD versus double reading without CAD. Dr. Gromet found that a single reader with CAD had a

statistically significant increase in sensitivity (11%) and a smaller increase in recall rate (4%), when compared to a single reader without CAD assistance. The study also found that single reading with CAD review, when compared with independent double reading, resulted in a not statistically significant increase in sensitivity but with a statistically significant lower recall rate. With manpower constraints limiting the use of double reading, Dr. Gromet concludes that “CAD appears to be an effective alternative that provides similar, and potentially greater, benefits.”[49]

CHAPTER 3

CAD FOR MASS DETECTION

3.1 Introduction

In this chapter, we propose a CAD system for detecting masses in the digitized mammograms. This study is done through two main phases; the training phase and the testing phase. First in the training phase, the system is trained how to differentiate between normal and cancerous cases by using predefined normal and cancerous images. Then in the testing phase, we test the performance of the system by entering a test image to compute the correctness degree of the system decision.

3.2 Review of mass detection

A mass is defined as a space-occupying lesion seen in at least two different projections [50]. Radiologists characterize masses by their shape and margin properties. A number of researchers have worked on methods for detecting masses in mammograms. Circumscribed masses usually have variable sizes with normal dense tissue causing difficulties for mass detection. Masses with spiculated margins have a very high likelihood of malignancy and thus some methods have been developed specifically for the detection of spiculated masses. A spiculated mass is characterized by lines radiating from the margins of a mass [51]. However, since not all malignant masses are spiculated, the detection of nonspiculated masses is also important. Most mass detection algorithms consist of two stages: (1) detection of suspicious regions on the mammogram and (2) classification of suspicious regions as mass or normal tissue.

In [52], Pohlman *et al.* presented an adaptive region growing technique to segment masses from normal background. It achieved 97% detection sensitivity for a set of 51 mammograms. A fuzzy region growing method for segmenting breast masses was proposed by Guliato *et al.* in [53], [54]. Petrick *et al.* [55], [56] used a two-stage adaptive density-weighted contrast enhancement filter in

combination with a Difference-of-Gaussian (DoG) edge detector for the detection of masses. Mudigonda *et al.* [57] segmented the breast mass portions by establishing intensity links from the central portions of masses into the surrounding areas. Features based on the flow orientation in an adaptive ribbon of pixels around the mass boundaries were used to separate mass regions from normal breast regions. The methods yielded a mass versus normal tissue classification accuracy represented as an area (A_z) of 0.87 under the receiver operating characteristics (ROCs) curve with a dataset of 56 images including 30 benign, 13 malignant, and 13 normal cases selected from the mini Mammographic Image Analysis Society (MIAS) database [58]. A sensitivity of 81% at 2.2 false positives per image was obtained. The detected mass regions (13 malignant and 19 benign) were further classified as benign and malignant by using five features based on gray-level co-occurrence matrices (GCMs). The classification of benign vs. malignant yielded a performance of $A_z = 0.79$. In [59], Catarious *et al.* proposed a method where the initial mass was segmented from a difference of Gaussian (DoG) filtered images through multi-level thresholding. Features including shape, fractal dimension, the output from a Laguerre-Gauss (LG) Channelized Hotelling observer (CHO) were used to reduce the false positive rate. It achieved a sensitivity of 88%. Using the selected features, the false positives per image were reduced from 20 to 5 with no loss in sensitivity.

Other techniques have been investigated for the detection of masses in screening mammograms, such as neural networks by Christoyianni *et al.* [60] and Lo *et al.* [61], genetic algorithm by Xu *et al.* [62], support vector machines by Chu *et al.* [63], wavelet packets by Zhang [64], texture analysis Kwok *et al.* [65], and graph techniques by Li *et al.* [66].

Kegelmeyer *et al.* [67] developed a method to detect spiculated masses using a set of 5 features for each pixel. They used the standard deviation of a local edge orientation histogram (ALOE) and the output of four spatial filters which are a subset of Laws texture features [68], [69]. Karssemeijer *et al.* [70] detected stellate distortions by a statistical analysis of a map of pixel orientations. They

grouped suspicious regions and discarded regions which were smaller than 500 pixels. Li *et al.*[71] developed a method for lesion site selection using morphological enhancement and stochastic model-based segmentation technique. A finite generalized Gaussian mixture distribution was used to model histograms of mammograms.

Liu and Delp [72] pointed out that in general it is difficult to estimate the size of the neighborhood that should be used to compute the local features of speculated masses. Small masses may be missed if the neighborhood is too large and parts of large masses may be missed if the neighborhood is too small. To address this problem they developed a multi-resolution algorithm for the detection of spiculated masses. They generated a multi-resolution representation of a mammogram using the Discrete Wavelet Transform. They extracted four features at each resolution for each pixel. Pixels were then classified using a binary classification tree.

Li *et al.* [73] developed a two-step for detection of masses. In the first step, adaptive gray level thresholding was used to obtain an initial segmentation of suspicious regions. In the second stage, a fuzzy binary decision tree was used to classify the segmented regions as masses or normal tissue using features based on shape, region size and contrast. Matsubara *et al.* [74] developed an adaptive thresholding technique for the detection of masses. They employed histogram analysis techniques to divide mammograms into 3 categories ranging from fatty to dense tissue. Potential masses were detected using multiple threshold values based on the category of the mammogram. A number of features such as circularity, area, and standard deviation were used to reduce the number of false positives.

Petrick *et al.* [55] developed a two-stage algorithm for the enhancement of suspicious objects. In the first stage they proposed an adaptive Density Weighted Contrast Enhancement filter (DWCE) to enhance objects and suppress background structures. The DWCE filter and a simple edge detector (Laplacian of Gaussian) were used to extract ROIs containing potential masses. In the second stage, the DWCE was re-applied to the ROI. They further improved the

detection algorithm by adding an object-based region-growing algorithm to it [75].

Polakowski *et al.* [76] used a single Difference of Gaussian (DoG) filter to detect masses. The DoG filter was designed to match masses which were approximately 1 cm in diameter. ROIs were selected from the filtered image. Kobatake *et al.* [77] modeled masses as rounded convex regions, and based on this idea developed an "iris filter" to enhance and detect masses. Brzakovic *et al.* [78] used a two stage multi-resolution approach for detection of masses. Qian *et al.* [79] developed a multi-resolution and multi-orientation wavelet transform for the detection of masses and spiculation analysis. They observed that traditional wavelet transforms cannot extract directional information which is crucial for a spiculation detection task and thus, they introduced a Directional Wavelet Transform.

Zhang *et al.* [80] noted that the presence of spiculated lesions led to changes in the local mammographic texture. They proposed that such a change could be detected in the Hough domain, which is computed using the Hough transform. They partitioned an image into overlapping ROIs and computed the Hough transform for each ROI. The Hough domain of each ROI was thresholded to detect local changes in mammographic texture and to determine the presence or absence of a spiculated mass. Brake *et al.* [81] defined a number of features to discriminate between lesions and normal tissue that were designed to capture image characteristics like intensity, iso-density, location and contrast.

3.3 Fractal based CAD system

The main objective of our work is to clarify the usefulness of using the fractal features (as a texture scale-invariant features) to classify normal and cancerous images. We showed that in three steps, in the first two steps we did not use any preprocessing techniques such as smoothing, edge sharpening, or wavelet decomposition. We just dealt with the mammograms as raw data without any alteration in it. The first step was classification of images without fractal features. Then, in the next step we classify with all features including the

fractal texture features (but without any enhancement). The third step we included the fractal features but after preprocessing which is the use of the median filter. Fig. 3.1 shows the block diagram for the proposed CAD system.

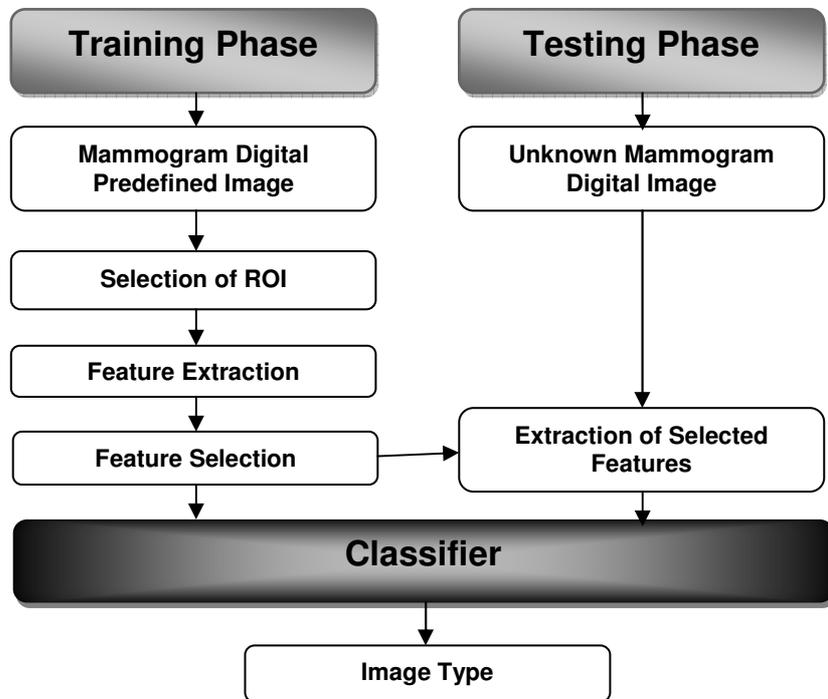


Fig. 3.1: Block diagram of the CAD system.

3.3.1 Fractal Analysis

Traditionally the Euclidean objects [82], such as lines, planes, and circles etc., have used as the basis of the intuitive understanding of the geometry of nature. However, most nature objects do not resemble Euclidean objects. Fractal geometry made it possible to model nature objects to a better description in many conditions. The concept of fractal was first introduced by Mandelbrot [83]. The main distinct difference between Euclidean and fractal geometry is that of self-similarity which is described by nonuniform scaling. In theory, shapes of fractal objects keep invariant under successive magnifying or shrinking the objects. It is known that the texture is a problem of scale, and the texture description is scale dependent. Hence, using fractal geometry can overcome the scale problem of texture. Because the concept of fractal dimension is an indicator of the surface roughness, people usually describe texture as fine,

coarse, grained, and smooth, etc. Hence, it implies that fractal-based texture analysis is a correlation between texture coarseness.

A variety of procedures, including box-counting, fractal Brownian motion [84], [85], and fractal interpolation function system [82], have been proposed for estimating the fractal dimension of images. The fractional Brownian motion model with gray-scale variation [84], [85] has been shown promise in the medical image texture. The Brownian motion curve concepts can be extended to the fractional Brownian motion curve $I(x)$, and $|I(x_2)-I(x_1)|$ have a mean value proportional to $|x_2-x_1|^H$. Thus, in the fractal Brownian motion there is only one parameter of interest, H , or the Hurst coefficient, which can be described as texture features when we applied to classify breast tumor images. Considering the topological dimension T_d , for images, $T_d=3$, the fractal dimension D can be estimated from the Hurst coefficient $H = T_d - D$.

For the medical images, the fractal dimension can be estimated from the above relationship. For an $M \times M$ image I , the implementation of estimation fractal dimension [84] can be defined as

$$di(k) = \frac{\sum_{x_1=0}^{M-1} \sum_{y_1=0}^{M-1} \sum_{x_2=0}^{M-1} \sum_{y_2=0}^{M-1} |I(x_2, y_2) - I(x_1, y_1)|}{Pn(k)}, \quad (3.1)$$

where $Pn(k)$ is total number of pixel pairs with distance

$$\Delta r_k = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (3.2)$$

And

$$f(k) = \log(di(k)) - \log(di(1)), \quad (3.3)$$

for $k = 1, 2, \dots, n$.

The vector $[di(1), di(2), \dots, di(n)]$ is called the multiscale intensity difference (MSID) vector and the vector $[f(1), f(2), \dots, f(n)]$ is called the fractional Brownian motion feature (FBM) vector. Fractal dimension D is then derived from the values of the Hurst coefficients. A small value of the fractal dimension, i.e. a large value of H , represents a fine texture, while a large fractal dimension corresponds to a coarse texture. However, the drawback of the Eq. (1) tends to be time consume, this algorithm needs M^4 operations for estimating the fractal dimension of an image of size $M \times M$. Thus, Chen *et al.*, [85] proposed another modified method using the gray level differences between the pixel pairs with horizontal, vertical, diagonal, and asymmetric-diagonal directions. The $di(k)$ is redefined as

$$di(k) = \left(\begin{array}{l} \sum_{x=0}^{M-1} \sum_{y=0}^{M-k-1} |I(x, y) - I(x, y+k)| / M(M-k) + \sum_{y=0}^{M-1} \sum_{x=0}^{M-k-1} |I(x, y) - I(x+k, y)| / M(M-k) \\ + \sum_{x=0}^{M-k-1} \sum_{y=0}^{M-k-1} |I(x, y) - I(x+, y+k)| / (M-k)^2 + \sum_{x=0}^{M-k-1} \sum_{y=0}^{M-k-1} |I(x, M-y) - I(x+k, M-(y+k))| / (M-k)^2 \end{array} \right) / 4 \quad (3.4)$$

As we all know, the fractal analysis is sensitive to noise; hence a filter (like median filter, mean filter, or morphological operations such as erosion and dilation) is needed to eliminate the noise from the image to be analyzed. Also a histogram equalization [86] may be used to make the images at comparable gray levels (as they came from different scanners), also to enhance the image contrast.

3.4 Methodology

3.4.1 DDSM Mammographic database.

The data collection that was used in our experiments was taken from the digital database for screening mammography (DDSM) distributed by University of South Florida [87]. The DDSM is a resource for use by the mammographic image analysis research community. Primary support for this project was a grant from the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command. The database contains approximately 2620 studies.

Each study includes two images of each breast, along with some associated patient information (e.g. age at time of study) and image information (e.g. scanner, spatial resolution). Images containing suspicious areas have been associated with information about the locations and types of suspicious regions.

Each case contains between 6 and 10 files. These are an ".ics" file, an overview "16-bit PGM" file, four image files that are compressed with lossless JPEG encoding and zero to four overlay files. Normal cases will not have any overlay files. Fig. 3.2 shows an example of the files in the directory for a case in the DDSM database.

```
B-3024-1.ics
B_3024_1.RIGHT_CC.OVERLAY
B_3024_1.RIGHT_MLO.OVERLAY
B_3024_1.LEFT_CC.LJPEG
B_3024_1.LEFT_CC.OVERLAY
B_3024_1.LEFT_MLO.LJPEG
B_3024_1.LEFT_MLO.OVERLAY
B_3024_1.RIGHT_CC.LJPEG
B_3024_1.RIGHT_MLO.LJPEG
TAPE_B_3024_1.COMB.16_PGM
```

Fig. 3.2: An example of the files in the directory for any provided case in the DDSM database [87].

Each type of these files is described as follows.

3.4.1.1 Description of ".ics" files

The ".ics" file provides information about a case as a whole. In ASCII format, it lists important information such as the date of the study, the patients' age, the date of digitization of the films, the type of digitizer used and a list of the image files.

The size of each image file, number of bits per pixel, the scanning resolution (in microns) and information on the existence or lack of an overlay file for each image is provided. As you can see in Fig. 3.3, all four images have overlays and these files are listed in Fig. 3.2. If the image description lines had "NON-OVERLAY" instead of "OVERLAY" then the images would not have overlay files.

```
ics_version 1.0
filename B-3024-1
DATE_OF_STUDY 2 7 1995
PATIENT_AGE 42
FILM
FILM_TYPE REGULAR
DENSITY 4
DATE_DIGITIZED 7 22 1997
DIGITIZER LUMISYS
SELECTED
LEFT_CC LINES 4696 PIXELS_PER_LINE 3024 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
LEFT_MLO LINES 4688 PIXELS_PER_LINE 3048 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
RIGHT_CC LINES 4624 PIXELS_PER_LINE 3056 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
RIGHT_MLO LINES 4664 PIXELS_PER_LINE 3120 BITS_PER_PIXEL 12 RESOLUTION 50
OVERLAY
```

Fig. 3.3: An example of the ".ics" files [87].

3.4.1.2 Description of ".LJPEG" files

The images have all been stored in a format using LOSSLESS JPEG compression. Even with the compression, each image file is very large because the films were scanned with a resolution between 42 and 100 microns. Once uncompressed, each image file contains only raw pixel values. Because there is no "header information" in the file, the size of each image must be obtained from the ".ics" file.

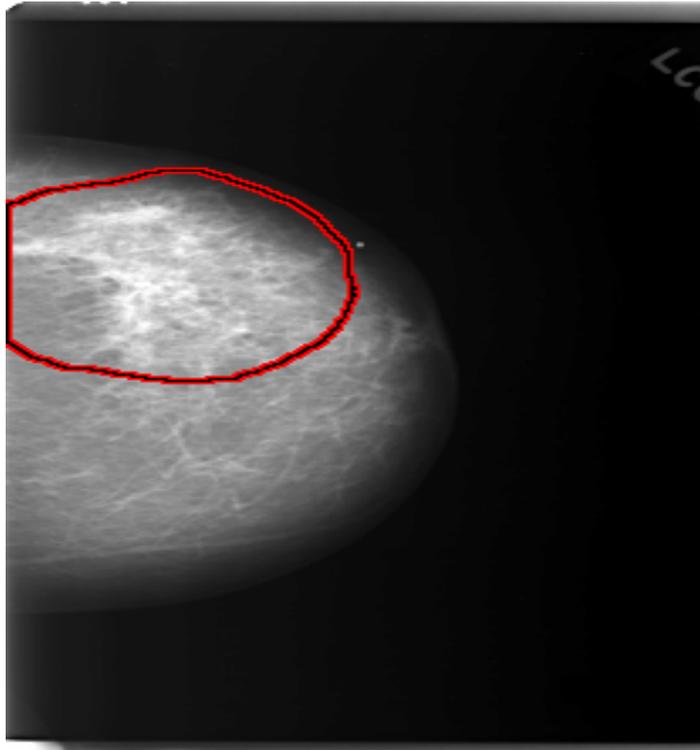


Fig. 3.4: Cancerous case left CC [87].



Fig. 3.5: Cancerous case left MLO [87].

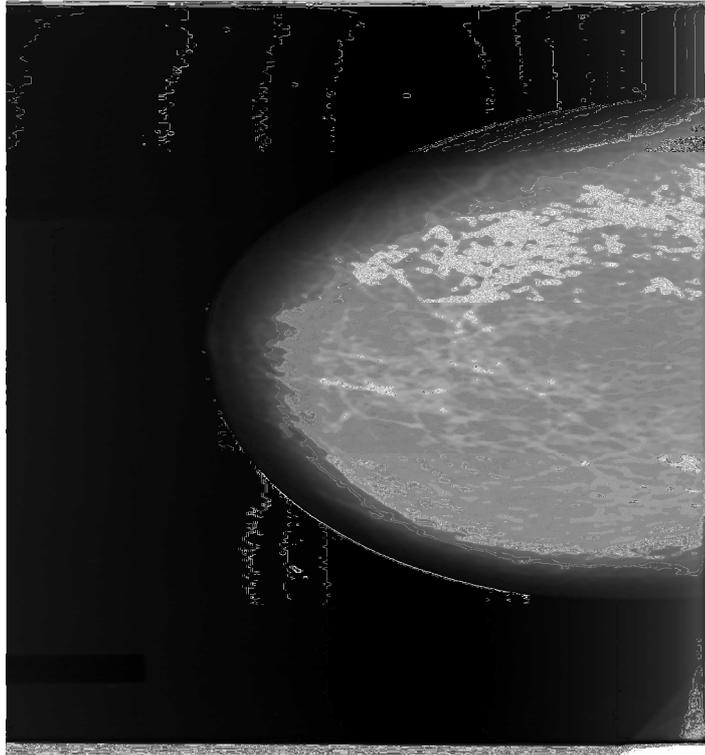


Fig. 3.6: Normal case right CC [87].

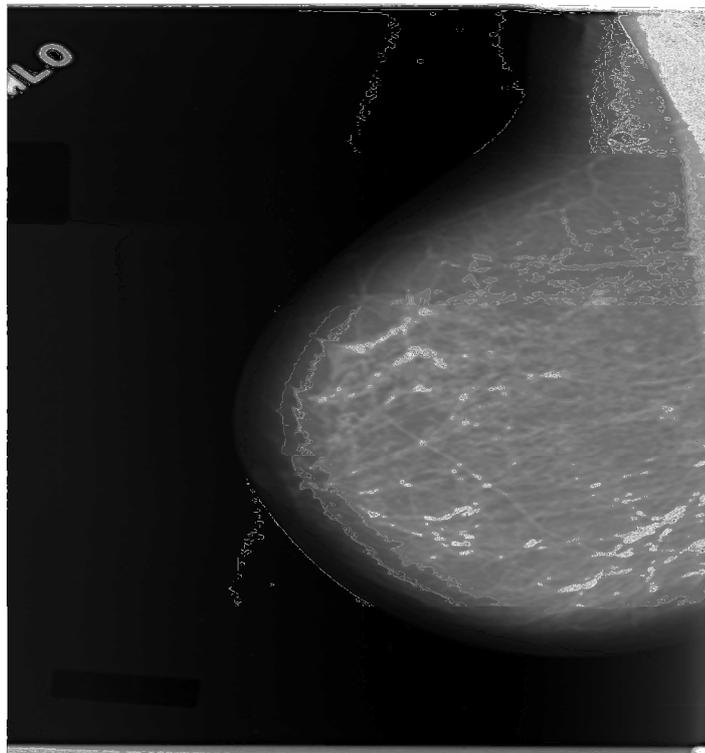


Fig. 3.7: Normal case right MLO [87]

3.4.1.3 Description of ".OVERLAY" files

```
TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE CALCIFICATION TYPE PLEOMORPHIC-
FINE_LINEAR_BRANCHING DISTRIBUTION REGIONAL
ASSESSMENT 5
SUBTLETY 4
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 4
BOUNDARY
8 1368 4 4 4 4 4 4 4 4 2 2 2 2 2 2 2 2 ... 0 0 0 0 0 0 0 0 0 1 #
    CORE
168 1824 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 ... 1 0 1 1 0 1 1 0 1 1 #
    CORE
384 1848 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 ... 0 0 0 0 0 0 0 0 0 0 #
    CORE
368 2192 6 6 6 6 6 6 6 6 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 0 #
```

Fig. 3.8: An example of the ".OVERLAY" files [87].

Abnormal cases have between one and four overlay files depending on the number of images in which the radiologist marked any abnormalities. You can know which images have overlay files by looking in the ".ics" file. Each image that has "OVERLAY" at the end of the line (not "NON-OVERLAY") will have an overlay file. Each overlay file may specify multiple abnormalities, so the first line of the file gives the total number of abnormalities. In the case of multiple abnormalities, each abnormality is then listed one after another.

Each abnormality has information on the lesion type, the assessment, the subtlety, the pathology and at least one outline. The assessment code is a value from 1 to 5. The subtlety rating is a value from 1 to 5, where 1 means "x1," 2 means "x2," 3 means "x3," 4 means "x4," and 5 means "x5." The lesion type,

assessment, and subtlety are specified by an experienced radiologist. Similarly, the outlines for the suspicious regions are derived from markings made on the film by an experienced radiologist. In some cases there is more than one outline for the same abnormality. In these situations the "TOTAL_OUTLINES" number is more than one. Fig. 3.8 shows an example of this. The first boundary will contain all of the other boundaries, and all boundaries after the first one will begin with the work "CORE".

Each boundary is specified as a chain code. This chain code is found on the line after the keyword "BOUNDARY" or "CORE" as discussed above. The first two values of each chain code are the starting column and row of the chain code that order. Following these two numbers, the chain code is given and a "#" character indicates the end of the chain code. The numbers correspond to the directions as follows:

	X→		
	7	0	1
Y ↓	6	X	2
	5	4	3

Chain code	0	1	2	3	4	5	6	7
X Coordinate	0	1	1	1	0	-1	-1	-1
Y Coordinate	-1	-1	0	1	1	1	0	-1

Fig. 3.9: The chain code for abnormalities boundary [87].

3.4.1.4 A description of ".16_PGM" files

The ".16_PGM" files are concatenated sub-sampled images. They are stored in 16-bit PGM (Portable Gray Map) format. A small version of the result of displaying the image in that way is displayed in Fig. 3.10 since global histogram equalization does not provide very good results; the quality of the displayed image may be poor even though the file contains pretty good information. The purpose of these files is just to provide a "quick" look at the images.

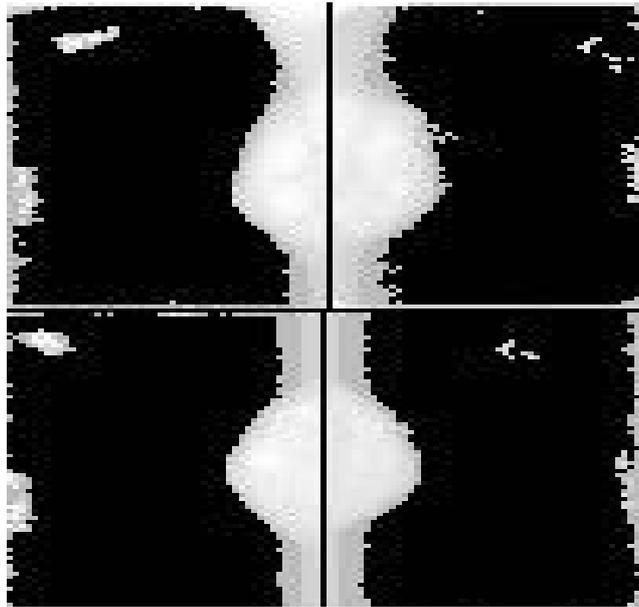


Fig. 3.10: A ".16_PGM" image [87].

3.4.2 Extraction of ROI.

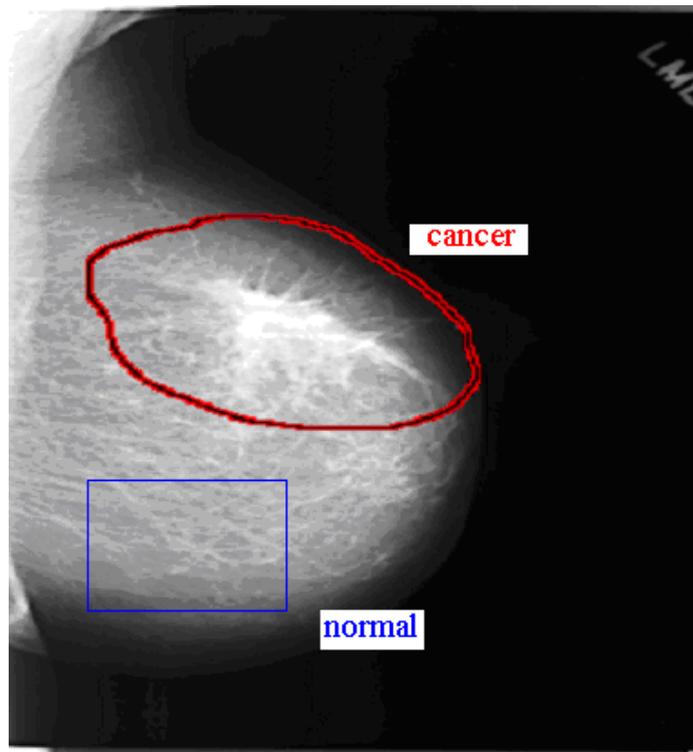


Fig. 3.11: Acquiring normal regions from cancerous images [87].

Using the contour supplied by the DDSM for each mammogram, we extracted the ROI of size 512 x 512 pixels with mass centered in the window.

We have used 21 cases. These cases are digitized by the LUMISYS digitizer. They are cancerous cases. We got normal images from the cancerous images by taking regions away from the cancer region as shown in Fig. 3.11. From each image we took from one to three normal regions depending on the available normal space in the image. This resulted in 55 cancerous and 70 normal ROIs.

3.4.3 Feature extraction.

A typical mammogram contains a great amount of heterogeneous information that depicts different tissues, vessels, ducts, chest skin, breast edge, the film, and the X-ray machine characteristics. In order to build a robust CAD system that correctly classifies normal and abnormal regions of mammograms, we have to present all the available information that exists in mammograms to the diagnostic system so that it can easily discriminate between the normal and the abnormal tissue. However, the use of all the heterogeneous information, results to high dimensioned feature vectors that degrade the diagnostic accuracy of the utilized systems significantly as well as increase their computational complexity and calculation time. Therefore, reliable feature vectors should be considered that reduce the amount of irrelevant information thus producing robust Mammographic descriptors of compact size. In our approach, we examined a set of 35 features were applied to the ROI using a window of size 64x64 pixels with 64 pixels shift, (i.e. no overlap).

3.4.3.1 Mean value of gray levels

It is represent the summation of gray level values of pixels divided by their number. It represents the average gray level in the image. Mathematical equation that describe mean [88]:

$$Mean = \mu = \sum_{i=1}^N ih(i). \quad (3.5)$$

Assuming that $h(i)$ is a normalized histogram value for a gray-level i and N denotes number of gray levels in the image.

3.4.3.2 Standard deviation of gray levels

Standard deviation measure hue spread out the value in a data set with respect to the mean. It is measure of average difference between the values of the data in the data set. If the data points are all similar, then the standard deviation will be low (closer to zero). If the data points are highly variable, then standard deviation will be high. The mathematical equation is as follow [88]:

$$Std = \sigma = \sqrt{\sum_{i=1}^N (i - \mu)^2 h(i)}. \quad (3.6)$$

Assuming that $h(i)$ is a normalized histogram value for a gray-level i , μ is mean, and N denotes number of gray levels in the image.

3.4.3.3 Variance

Measure the dispersion of a set of data points around their mean value (a more credible feature, measure region "roughness"). The mathematical definition of variance is as follow [88]:

$$Variance = \sigma^2 = \sum_{i=1}^N (i - \mu)^2 h(i). \quad (3.7)$$

Assuming that $h(i)$ is a normalized histogram value for a gray-level i , μ is mean, and N denotes number of gray levels in the image.

3.4.3.4 Skewness

It is the third central moment [88], [90]. Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. It is impact on our project is that: if the skewness is positive, this region of interest (ROI) will have a mean gray level that is greater than the normal mean value and this ROI will have be brighter than the normal regions. If the skewness is negative, this ROI will have

a mean gray level that is less than the normal mean value and this ROI will have be darker than the normal regions (are the pixel intensities usually darker/lighter than average?). The mathematical equation for skewness is as follow:

$$Skewness = \sigma^{-3} \sum_{i=1}^N (i - \mu)^3 h(i). \quad (3.8)$$

Assuming that $h(i)$ is a normalized histogram value for a gray-level i , μ is mean, and N denotes number of gray levels in the image.

3.4.3.5 Kurtosis

Kurtosis is a measurement of how outlier-prone distribution is [88], [90]. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than normal distribution have kurtosis greater than 3; distributions that are less outlier-prone than normal distribution have kurtosis less than 3 (how "uniform" is the grey level distribution?). The mathematical equation for kurtosis is as follow:

$$Kurtosis = \sigma^{-4} \sum_{i=1}^N [(i - \mu)^4 h(i)] - 3. \quad (3.9)$$

Assuming that $h(i)$ is a normalized histogram value for a gray-level i , μ is mean, and N denotes number of gray levels in the image.

Kurtosis is a measure of extent of the peak (or the degree of flatness near its center) in a distribution. If the distribution is normal, then this ratio is 3. A ratio greater than 3 indicates more values in the neighborhood of the mean (is more peaked than the normal distribution). If the ratio is less than 3, then it is an indication that the curve is flatter than the normal [90].

3.4.3.6 Entropy

Entropy is a statistical measure of randomness that can be used to characterize the texture of the image (how normal/non-normal is the grey level distribution?). The mathematical equation for entropy is as follow [88]:

$$Entropy = -\sum (p * \log(p)), \quad (3.10)$$

where p is the probability of pixels value.

3.4.3.7 Spreadness

This feature shows the degree of spread of the shape around the centered intuitively, i.e. measure the circularity of ROI [77]. Mathematical equation of spreadness is as follow:

$$Spreadness = \frac{\sum_i \sum_j S(i, j)(i - i_0)^2 + \sum_i \sum_j S(i, j)(j - j_0)^2}{\sum_i \sum_j S(i, j)}, \quad (3.11)$$

where (i_0, j_0) is the center of gravity of the region S and the sum is taken within the region.

3.4.3.8 Percentile and Cumulative Frequency Graph

Cumulative relative frequency, or cumulative percentage, gives the percentage of quantities have a measurement less than or equal to the upper boundary of the class interval, and provides a class of important statistics known as percentiles or percentile scores [90]. The 90th percentile, for example is the numerical value that exceeds 90% of the values in the data set and is exceeded by only 10% of them. Or, as another example, the 80th percentile is the numerical value that exceeds 80% of the values in the data set and is exceeded by only 20% of them, and so on. The 50th percentile is commonly called the median.

$$Percentile K\% = [k] \left\{ k : \left(\sum_{i=1}^k h(i) \right) \left(\sum_{i=1}^N h(i) \right)^{-1} = K\% \right\}. \quad (3.12)$$

Assuming that $h(i)$ is a normalized histogram value for a gray-level i and N denotes number of gray levels in the image. We are extracted nine percentiles features from 10^{th} percentile to 90^{th} percentile.

3.4.3.9 Range

It is the difference between the maximum and the minimum of a sample. The range is an easily-calculated estimate of the spread of the values in a data set.

3.4.3.10 Interquartile Range (IQR)

It is the difference between the 75th and the 25th percentiles of the values in a data set. The IQR is a robust estimate of the spread of the data, since changes in the upper and lower 25% of the data do not affect it. If there are outliers in the data, then the IQR is more representative than the standard deviation as an estimate of the spread of the data.

3.4.3.11 Average Absolute Deviation (AAD).

The average absolute deviation or simply average deviation of a data set is the average of the absolute deviations and is a summary statistic of statistical dispersion or variability. It is also called the mean absolute deviation, but this is easily confused with the median absolute deviation.

The average absolute deviation of a set $\{x_1, x_2, \dots, x_n\}$ is

$$Energy = \sum_{i=1}^N \sum_{j=1}^N p(i, j)^2. \quad (3.13)$$

The choice of measure of central tendency, $m(X)$, has a marked effect on the value of the average deviation.

3.4.3.12 Seven invariant moments

These features a set of moments is invariant to translation, rotation, and scale change [89]. The central moments in digital form are defined as:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y). \quad (3.14.a)$$

The central moments of order up to 3 are given in the below:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}},$$

$$\mu_{00} = \sum_x \sum_y (x - \bar{x})^0 (y - \bar{y})^0 f(x, y) = \sum_x \sum_y f(x, y) = m_{00}. \quad (3.14.b)$$

$$\mu_{10} = \sum_x \sum_y (x - \bar{x})^1 (y - \bar{y})^0 f(x, y) = m_{10} - \frac{m_{10}}{m_{00}} m_{00} = 0. \quad (3.14.c)$$

$$\mu_{01} = \sum_x \sum_y (x - \bar{x})^0 (y - \bar{y})^1 f(x, y) = m_{01} - \frac{m_{01}}{m_{00}} m_{00} = 0. \quad (3.14.d)$$

$$\mu_{11} = \sum_x \sum_y (x - \bar{x})^1 (y - \bar{y})^1 f(x, y) = m_{11} - \frac{m_{10}}{m_{00}} m_{01} = m_{11} - \bar{y} m_{01} = m_{11} - \bar{x} m_{10}. \quad (3.14.e)$$

$$\begin{aligned} \mu_{20} &= \sum_x \sum_y (x - \bar{x})^2 (y - \bar{y})^0 f(x, y) = m_{20} - \frac{2m_{10}^2}{m_{00}} + \frac{m_{10}^2}{m_{00}} \\ &= m_{20} - \frac{m_{10}^2}{m_{00}} = m_{20} - \bar{x} m_{10}. \end{aligned} \quad (3.14.f)$$

$$\mu_{02} = \sum_x \sum_y (x - \bar{x})^0 (y - \bar{y})^2 f(x, y) = m_{02} - \frac{m_{01}^2}{m_{00}} = m_{02} - \bar{y} m_{01}. \quad (3.14.g)$$

$$\mu_{21} = \sum_x \sum_y (x - \bar{x})^2 (y - \bar{y})^1 f(x, y) = m_{21} - 2\bar{x} m_{11} - \bar{y} m_{20} + 2\bar{x}^2 m_{01}. \quad (3.14.h)$$

$$\mu_{12} = \sum_x \sum_y (x - \bar{x})^1 (y - \bar{y})^2 f(x, y) = m_{12} - 2\bar{y} m_{11} - \bar{x} m_{02} + 2\bar{y}^2 m_{10}. \quad (3.14.i)$$

$$\mu_{30} = \sum_x \sum_y (x - \bar{x})^3 (y - \bar{y})^0 f(x, y) = m_{30} - 3\bar{x} m_{20} + 2\bar{x}^2 m_{10}. \quad (3.14.j)$$

$$\mu_{03} = \sum_x \sum_y (x - \bar{x})^0 (y - \bar{y})^3 f(x, y) = m_{03} - 3\bar{y}m_{02} + 2\bar{y}^2 m_{01}. \quad (3.14.k)$$

The normalized central moments, denoted η_{pq} are defined as: $\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}$,

where $\gamma = \frac{p+q}{2} + 1$ for $p+q = 2, 3, \dots$

A set of seven invariant moments can be derived from the second and third moments:

$$\phi_1 = \eta_{20} + \eta_{02}. \quad (3.15)$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2. \quad (3.16)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2. \quad (3.17)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (3\eta_{21} + \eta_{03})^2. \quad (3.18)$$

$$\phi_5 = (\eta_{30} + 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}). \quad (3.19)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}). \quad (3.20)$$

$$\phi_7 = (3\eta_{12} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]. \quad (3.21)$$

3.4.3.13 Fractal Dimension Features

We calculated the fractal dimension with eight fractional Brownian motion feature coefficients as fractal texture features [91]. A fractal is an irregular geometric object with an infinite nesting of structure at all scale. Some of the most important properties of fractals are self-similarity, chaos, and non-integer fractal dimension (FD). The FD offers a quantitative measure of self-similarity

and scaling. The fractal dimension can be defined as the exponent of the number of self-similar pieces (N) with magnification factor ($1/r$) into which a figure may be broken [92]. The equation for FD is as the following:

$$FD = \frac{\ln(N)}{\ln(1/r)}. \quad (3.22)$$

Two methods which had taken to estimate the fractal dimension feature:

The Piecewise Modified Box-Counting (PMBC)

In PMBC method, the image of size $M \times M$ pixels is scaled down to a size $r \times r$ where $M/2 \geq r > 1$ and r is an integer. Consider the image $i(x, y)$ as a 2D plane and the pixel intensity p as the height above a plane. Thus, image $i(x, y)$ is partitioned into grids of size $r \times r$ and on each grid there is a column of boxes of size $r \times r \times p$. Assume that the maximum and minimum gray levels of the image $i(x, y)$ in (I, j) -th grid fall in box number k and l respectively. Then $n_r(I, j) = k - l + 1$ is the contribution of N_r in the (I, j) -th grid. The contributions from all the grids using this equation:

$$N_r = \sum_{i,j} n_r(i, j), \quad (3.23)$$

where N_r is computed for different values of the square of size r . The FD of an image is calculated from the slope of the linear regression line obtained when the horizontal axis and the vertical axis are taken as $\log 1/r$ and $\log N_r$ [92].

The Piecewise Triangular Prism Surface Area (PTPSA)

In the PTPSA method used the grayscale elevation values at the corners of a box at points (A , B , C , and D), and the average value of the corners as center elevation value at point (E) forms four triangular (ABE , BCE , CDE , and DAE) shown in Fig. 3.12. By repeating this calculation for different box size r , the logarithm of surface areas of the top triangular surfaces versus logarithm of the box size is calculated to obtain the slope (FD) [92].

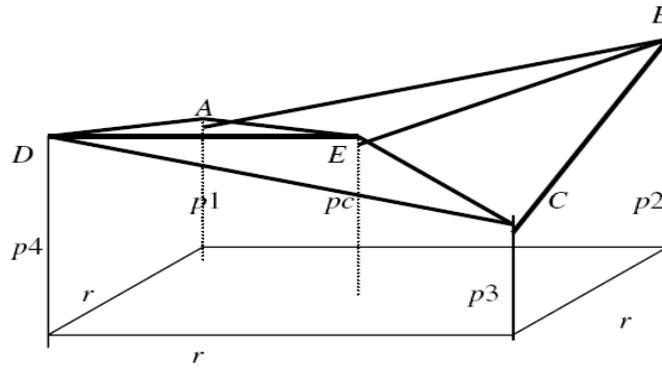


Fig. 3.12: Triangular prism surface area method algorithm: the p_1 , p_2 , p_3 , and p_4 are the grayscale elevation values at the box corners, p_c is the average of corner values. ABE, BCE, CDE, and DAE are four triangles [92].

3.4.4 Feature selection.

After the extraction of the previously mentioned features, we transformed these 2-D matrices to 1-D vectors to form the two clusters; normal cluster and cancerous cluster. This is done by putting the columns one after the other in one vector.

We want to know whether a certain feature can differentiate between normality and cancer or not. This is done through the t-test. The t-test is a built-in function in MATLAB. It takes the two vectors of each feature. We used a significance level α of 0.05. Our null hypothesis is that the two vectors come from the same distribution. The alternative hypothesis is that the two vectors are not from the same distribution and this feature has the ability of discriminating between normal and cancerous breast tissues. The significance level is related to the degree of certainty required in order to reject the null hypothesis in favor of the alternative. For this significance level, the probability of incorrectly rejecting the null hypothesis when it is actually true is 5% [93]-[95].

The t-test computes the p-value which is the probability of observing the given sample result under the assumption that the null hypothesis is true. If the p-value is less than α , then the null hypothesis is rejected. For example, if $\alpha = 0.05$ and the p-value is 0.03, then you reject the null hypothesis. The

converse is not true. If the p-value is greater than α , you have insufficient evidence to reject the null hypothesis.

The two vectors that enter to the t-test are the normal vector and the cancerous vector of each feature. Each vector of a certain feature is formed by concatenating the vectors of this feature from all images in one vector.

First and without any enhancement we applied the t-test on the 35 candidate features. The test indicated that only 26 features can discriminate between the two clusters (spreadness, range, skewness, kurtosis, 5th and 6th invariant moments, 2nd and 5th and 7th fractal coefficients are excluded). All the differentiating features have a p-value that is so close to zero.

Then we tried to enhance the images using the median filter and histogram equalization, and then the 35 features are applied and tested. In this case the test indicated that 31 features can be used to discriminate between the two clusters (range, skewness, 5th and 6th invariant moments are excluded).

Now, we have the proper features for classification. So, we can proceed further in our work toward mass detection. In the next stages, we are going to form the classes, build the classifiers and test the system.

3.4.5 Classification

First, we will construct the clusters, then learn the system and test it. The classification process is divided into the training phase and the testing phase. In the training phase, known data are given and the features are calculated by the processing which precedes classification. Separately, the data on a candidate region which has already been decided as a tumor or as normal are given, and the classifier is trained. We used the training set for this phase which consists of 35 cancerous ROIs and 40 normal ROIs. In the testing phase, unknown data are given and the classification is performed using the classifier after training. Breast cancer image diagnosis assistance is the task in the recognition phase. We used a testing set for this phase which consisted of 20 cancerous ROIs and 30 normal ROIs.

There are different types of classifiers. We used the minimum distance and the Voting K-Nearest Neighbor (K-NN) classifiers for their simplicity.

3.4.5.1 Minimum Distance Classifier

A pattern classifier employing linear discriminant functions is termed a linear machine (Nilsson, 1965), an important special case of which is the minimum-distance classifier or nearest-neighbor rule. Suppose we are given a set of prototype points p_1, \dots, p_n , one for each of the n classes w_1, \dots, w_n . The minimum-distance classifier assigns a pattern x to the class w_i associated with the nearest point w_i . For each point, the squared Euclidean distance is

$$d_{\min} = \min \|x - p_i\|, \quad (3.24)$$

and the minimum value indicates that this pattern x belongs to the associated class w_i that resulted in this minimum distance.

In case of using many features at the same time to make pattern recognition,

x will be a vector $\begin{matrix} x \\ - \end{matrix}$ and p_i will be a vector $\begin{matrix} p \\ - i \end{matrix}$ and the minimum norm between the pattern $\begin{matrix} x \\ - \end{matrix}$ and the center means of the features of each class indicates the class of the pattern.

In this algorithm, a test sample is classified by assigning it to the class which has the nearest mean vector.

3.4.5.2 Voting K-Nearest Neighbor (K-NN) classifier

The Voting k-Nearest Neighbor (k-NN) classifier is nonparametric technique, it assigns a test sample to the class of the majority of its K-neighbors; that is, assuming that the number of voting neighbors is $k = k_1 + k_2 + k_3$ (where k_i is the number of samples from class i in the k-sample neighborhood of the test sample), the test sample is assigned to class m if $k_m = \max \{k_i, i=1, 2, 3\}$ [96]. The algorithm used to implement this technique is described below:

1. Obtain and store the distances between the parameter sets of the test sample and all the samples in the designed set. Distance can be a measure of Euclidean Distance (ED).
2. Sort the obtained distance values in ascending order.
3. Consider the subset of the first k distances in the sorted array; i.e., the k nearest neighbors. Knowing the class membership of each of these samples, assign the test sample to the majority class in this subset if it exists, otherwise the result is considered inconclusive.
4. Estimate the error rate by comparing the classification results with actual class membership. Treat the special case of inconclusive decisions individually as a spread entity (i.e., neither an error nor a correct decision) and obtained its rate of occurrence.

3.4.6 Evaluation for the Classification

The supervised classifier used with label or known data and the classification process by using supervised classifier is divided into the training phase and the testing phase. Through the training phase, the system was trained how to differentiate between normal and cancerous cases is learned by using known normal and cancerous images. In the testing phase, the performance of the system is test by entering a test image to compute the correctness degree of the system decision by using unknown normal and cancerous images.

Since there is a limited amount of available data in training, it is very important to test the system with extra data. However, it is an issue how to use this limited amount of data in both training and testing. More data used in training lead to better system designs, whereas more data used in testing lead to more reliable evaluation of the system.

Evaluating the system according to the success obtained on the training set brings the risk of memorization of data and obtaining over-optimistic error rates. To circumvent the memorization problem, the system should be evaluated on a separate data set that is not used in training the system. For that, one approach is

to split the data into two disjoint sets and use these sets to train and test the system. In the case that it is not feasible to use a significant portion of the data as the testing set, k -fold cross-validation can be used. This approach randomly partitions the data set into k groups. Then, it uses $(k - 1)$ groups to train the system and uses the remaining group to estimate an error rate. This procedure is repeated k times such that each group is used for testing the system. Leave-one-out is a special case of the k -fold cross validation where k is selected to be the size of the data; therefore only a single sample is used to estimate the error rate in each step.

Since the testing stage should measure how well the system will work on unknown samples in the future, the test set should also consist of the samples that are independent from those used in the training. We measured, quantitatively, the detection performance of the classifiers by computing the sensitivity and specificity on the data. Before definition of sensitivity and specificity; Fig. 3.13 displays the confusion matrix, which illustrates the four possible outcomes of an evaluation. The results of a test performed with perfect sensitivity and specificity will all be either TP or TN and never FP or FN [97]-[102].

		Identified	
		Yes	No
Actual	Yes	TP	FN
	No	FP	TN

Fig. 3.13: Confusion matrix.

Usually, an image region can be called cancerous (positive) or normal (negative), and a decision for a detection result can be either correct (true) or incorrect (false). A decision for a detection result, therefore, will be one of four possible categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). FN and FP are two kinds of errors.

- True positive (TP): the diagnostic system yields positive test result for the sample and the sample actually has the disease,
- False positive (FP): the diagnostic system yields positive test result for the sample but the sample does not actually have the disease,
- True negative (TN): the diagnostic system yields negative test result for the sample and the sample does not actually have the disease,
- False negative (FN): the diagnostic system yields negative test result for the sample but the sample actually has the disease.

By using the number of samples that fall into these categories, sensitivity and specificity are defined to assess the success of the diagnostic system. Sensitivity is the conditional probability of detecting cancer while there is really cancer in the image.

$$Sensitivity = \frac{TP}{TP + FN}. \quad (3.25)$$

Specificity is the conditional probability of detecting normal breast while the true state of the breast is normal.

$$Specificity = \frac{TN}{TN + FP}. \quad (3.26)$$

In the terms of the false-negative rate and the false-positive rate:

$$Sensitivity = 1 - \text{false negative rate}, \quad (3.27)$$

$$Specificity = 1 - \text{false positive rate}. \quad (3.28)$$

False negative rate: the probability that the classification result indicates a normal breast while the true diagnosis is indeed a breast disease (i.e. positive).

This case should be completely avoided since it represents a danger to the patient.

False positive rate: the probability that the classification result indicates a breast disease while the true diagnosis is indeed a normal breast (i.e. negative).

3.5 Results

Results differed by applying different type of classifiers due to the fact that each classifier has its own method for the formulation of the normal and cancerous clusters upon which it decides whether a test ROI is considered cancerous or normal.

Each classification method was adopted to verify the classification results. The images are divided into the training set and the test set. The training set is used to build the classifier model and the test set is used to verify the trained classifier model (Note that: the cases in the test set are not used to train the classifier model). Table 3.1 illustrates the results obtained using each of these classifiers in the three cases. Case (A) is obtained without enhancement and without the fractal features, case (B) is obtained without enhancement but with fractal features, and case (C) is obtained with fractal features after preprocessing to enhance the images.

Table 3.1: Summary of results for mass detection

		Minimum distance classifier		Voting K-Nearest Neighbor (K-NN)			
				K = 1		K = 3	
		Train	Test	Train	Test	Train	Test
A	Sensitiv	80%	75%	100	50%	91.4	40%
	specific	72.5	70%	100	73.3	87.5	76.7
B	Sensitiv	80%	80%	100	90%	94.3	90%
	specific	75%	63.3	85%	76.7	70%	70%
C	Sensitiv	91.4	85%	100	100	100	100
	specific	85%	80%	100	90%	95%	90%

Evaluating the results obtained, it's found that the best results for both the training set and testing set is obtained when using K-NN classifier especially with K=1 and when using fractal features and enhanced images.

Why Manual ROI Selection?

We were entering the whole image to the system. The result was very strange, where all the images were classified as cancerous images. By tracing the classification process, we found that there are two things that lead to this result. They are the background and the led that is used to protect the patient from the X-rays through the screening procedure. The problem of the background is solved by checking the gray level values of the ROI that we take through a window. If the minimum value is zero, we classify this ROI as normal. This solution is not an accurate solution as the window can include a part from the cancer.

The problem of the led is that it has large gray level values. Solution of this problem can be solved by checking the gray level value in the window of the ROI under study and to classify it as normal if the average gray level is greater than a threshold value that we assign to the led.

We ended with entering a ROI manually instead of entering the whole image.

Study of Changing the Window Size and Shift

Through all the previous steps, we used a window size of 64x64 and a shift of 64. We want to study the effect of changing the window size and the shift. First, we used a shift of 32 and changed the window size as 32, 64 and 96. Second, we fixed the window size of 64 and changed the shift as 16, 32 and 64.

All of this is done only with the preprocessed images and using the fractal features and the K-NN (with K=1) as a classifier. The results of this study were very close to the previous results mentioned in table 3.1.

3.6 Summary

In this chapter, a computer-aided diagnosis system for mass detection on mammographic images based on statistical features, invariant features and fractal analysis is proposed. The input image is the ROI subimage containing the lesion pre-selected by a physician. The fractal analysis is applied to obtain the fractal texture features in order to classifying the test cases into normal and

cancerous. From the experimental results, we can conclude that the fractal analysis is useful to represent the texture information of breast lesions.

CHAPTER 4

FRACTAL MODELING FOR MICROCALCIFICATIONS DETECTION

4.1 Review of microcalcifications detection

Microcalcifications are small calcium deposits that form in the breast as a result of benign or malignant processes. Mammographically, they appear as bright white spots of various sizes and shapes. The important characteristics of microcalcifications are their size, shape or morphology, number and distribution.

Microcalcifications are considered to be important signs of breast cancer. It has been reported that 30–50% of breast cancers detected radiographically demonstrate microcalcifications on mammograms, and 60–80% of breast carcinomas reveal microcalcifications upon histologic examinations. The high correlation between the presence of microcalcifications and the presence of breast cancers indicates that accurate detection of microcalcifications will improve the efficacy of mammography as a diagnostic procedure. The task of detection of microcalcifications for the diagnosis of breast cancer is a difficult one. Dense breasts, improper technical factors, or simple oversight by radiologists may contribute to the failure of detecting microcalcifications.

Given a mammogram, there are three major problems in analyzing and detecting microcalcifications [10].

- ☑ Microcalcifications are very small. On mammograms, they appear as tiny objects which can be described as granular, linear, or irregular. According to the literature, the sizes of microcalcifications are from 0.1–1.0 mm, and the average diameter is about 0.3 mm. Small ones (ranging 0.1–0.2 mm) can hardly be seen on the mammogram due to their superimposition on the breast parenchymal textures and noise.
- ☑ Microcalcifications often appear in an inhomogeneous background describing the structure of the breast tissue. Some parts of the

background, such as dense tissue, may be brighter than the microcalcifications in the fatty part of the breast.

- ☑ Some microcalcifications have low contrast to the background. In other words, the intensity and size of the microcalcifications can be very close to noise or the inhomogeneous background.

On the other hand, the high degree of localization of microcalcifications makes them somewhat easier to model (they are "impulse-like"), and indeed, a number of robust methods have been developed for the detection of microcalcifications and a great deal of success has been achieved with these methods. The detection performance of current commercial systems is reported at 95% sensitivity at less than 1 False Positive per Image (FPpI) [46].

A number of different approaches have been applied for the detection of microcalcifications. Microcalcifications represent high spatial frequencies in the image. Thus, one approach to the microcalcification detection task is to localize the high spatial frequencies of the image. The Wavelet transform is an optimal tool for such a task, as compared to other transforms such as the Fourier transform which only gives information on the frequency content and cannot spatially localize the frequencies. Thus, a number of authors have used wavelet transforms for the detection of microcalcifications [103]-[107]. In some of these methods, the image is first processed by a sub-band decomposition filter-bank. The coefficients in the sub-band images, which correspond to high spatial frequencies, are selectively weighted to enhance the microcalcifications. A new image with enhanced microcalcifications is created with the inverse Wavelet transform. The microcalcifications are then detected using global and local thresholds. Finally, the individual microcalcifications are then grouped together to detect clusters. A summary of some methods, which have used the wavelet transform, is given below.

Strickland and Hahn [103] proposed a method using undecimated bi-orthogonal wavelet transforms and sub-band weighting to detect and segment clustered microcalcifications. Yoshida *et al.* [104] used undecimated wavelet

transforms and supervised learning for microcalcification detection. Zhang *et al.* developed a method [105] to optimize the weights at individual scales of the wavelet decomposition. Qian *et al.* [106] used a tree structured wavelet transform for multi-resolution decomposition and selective reconstruction of sub-images to segment microcalcifications. They used a non-linear filter for suppressing image noise.

In most of the methods, the detection is carried out in the spatial domain. However, Gurcan *et al.* [107] performed the detection in the sub-band image domain. The key aspect of their method was that microcalcifications would produce outliers in the high-pass and band-pass sub-images. Thus, the symmetry of the distribution of the band-pass and high-pass image coefficients is altered in regions containing the microcalcifications. The changes in the distribution were captured by computing the skewness and kurtosis of the distribution.

Another reason why wavelets have been so effective is that microcalcifications appear as small bright dots on the mammogram and can be viewed as point discontinuities. Recently, mathematicians have argued that wavelets have finite square supports and are ideal for capturing point discontinuities, but not edges [108]. This fact intuitively explains the tremendous success of wavelet transform based methods in the detection of microcalcifications as well as why they have been less successful for the detection of masses.

In addition to wavelets, other multi-scale methods have been investigated. Netsch and Peitgen [109] proposed a multi-scale detection method based on the Laplacian of Gaussian filter and a mathematical model. They used scale-space signatures obtained from Laplacian filtering for the detection of clustered microcalcifications. Other non-wavelet based methods have also been developed for the detection of microcalcifications. These methods generally try to make maximum use of the fact that microcalcifications have much higher intensity values than the surrounding tissue in a mammogram. These methods are more likely to fail when the microcalcifications are present in dense background tissue.

Chan *et al.* [110] employed a difference-image processing technique to detect microcalcifications. In this methodology, they computed "signal enhanced" and "signal suppressed" images and subtracted these to obtain a difference image. Global and local level thresholding was then used to extract potential microcalcifications. In a later study [111], they incorporated an artificial neural network in order to reduce the number of False Positive FP clusters per image.

Davis and Dance [112] used local area thresholding to detect microcalcifications. Although they showed that this method was successful on a small test set, in general picking a threshold that will work successfully on a large set of images is extremely difficult.

Nishikawa *et al.* [113] combined the difference image technique with morphologic erosion filters and gray level thresholding techniques to extract microcalcifications. To reduce the number of false positives, Zhang *et al.* [80] applied a shift-invariant artificial neural network. Zheng *et al.* [114] developed a multistage algorithm including Gaussian filtering, nonlinear global thresholding for microcalcification detection. They used a mixed feature-based neural network for detection.

A number of authors have focused on developing techniques to reduce the FPs. The main aim here is to classify ROIs as either containing microcalcifications (positive ROI) or normal tissue (negative ROI). Various schemes have been developed for this purpose.

In [115] a comparative study of texture-analysis methods is performed for the surrounding region-dependence method, which has been proposed by Kim *et al.*, and conventional texture-analysis methods, such as the spatial gray level dependence method, the gray-level run-length method, and the gray-level difference method. Textural features extracted by these methods are exploited to classify regions of interest (ROI's) into positive ROI's containing clustered microcalcifications and negative ROI's containing normal tissues. A three-layer back-propagation neural network is used as a classifier. The results of the neural network for the texture-analysis methods are evaluated by using a receiver operating-characteristics (ROC) analysis. The surrounding region-dependence

method is shown to be superior to the conventional texture-analysis methods with respect to classification accuracy and computational complexity.

Nagel *et al.* [116] examined three methods of feature analysis, namely, rule based, an artificial neural network (ANN), and a combined method. In an independent database of 50 images, at a sensitivity of 83%, the average number of false positive (FP) detections per image was: 1.9 for rule-based, 1.6 for ANN, and 0.8 for the combined method. The authors demonstrated that the combined method performs best because each of the two stages eliminates different types of false positives.

Other method in microcalcification detection based on the fractal modeling and the fractal properties of the mammographic images where, comparing with microcalcifications, the breast background tissues have high local self-similarity, which is the basic property of fractal objects.

The fractal block coding method mainly exploits the self similarity property of fractals. Fractal image coding was first proposed by Barnsley [117] where, he introduced the notion of fractal image compression in which real world images could be modeled by deterministic fractal objects, which are attractors of sets of two dimensional affine transformations. Deterministic fractals have extremely high visual complexity with very low information content. They have high degree of redundancy such that they can be recursively made of transformed copies of either themselves or parts of themselves. A.E. Jacquin proposed a novel method for image compression [118], [119] by fractal block coding of images. Fractals have been used in a lot of image processing applications, compression segmentation, analysis, restoration, etc [10], [84], [120]-[125].

Li *et al.* [10] compared a fractal-based enhancement method with wavelet and morphological enhancement methods. They used three metrics (Contrast Improvement Index (CII), Peak Signal to Noise Ratio (PSNR), and Average Signal to Noise Ratio (ASNR)) to compare the performance of these three enhancement methods. Authors showed that the fractal approach was the best, compared to the other methods. The noise level in the fractal approach was also lower than the other two methods.

A fractal model has been used to describe the mammographic image, thus, allowing the use of a matched filtering stage to enhance microcalcifications against the background [126]. A region-growing algorithm, coupled with a neural classifier, detects existing lesions. Subsequently, a second fractal model is used to analyze their spatial arrangement so that the presence of microcalcification clusters can be detected and classified. Reported results indicate that fractal models provide an adequate framework for medical image processing; consequently high correct classification rates are achieved.

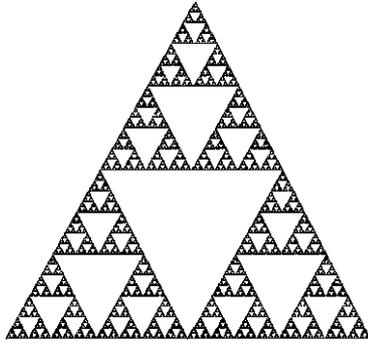
In the conventional method of fractal image coding, all the image blocks are searched for the matching domain for a particular range block. This is an exhaustive search and takes enormous amount of time. To reduce the time, the domain blocks whose mean value classes are the same or adjacent as the class of the range block are used in the searching process [127].

D. Sankar and T. Thomas [128] proposed a method for modeling the breast background tissues using mean and variance approach in the deterministic fractal model. In their study the average correlation between the original and the modeled mammograms were obtained as 0.9740 and the average mean square error was found to be 5.939. The results show that the true positive rate is 82% with an average of 0.214 negative clusters per image for 28 mammograms were obtained.

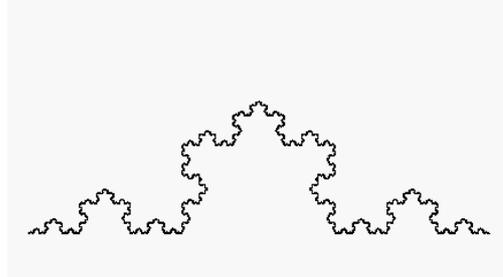
In a later study, a new fast fractal modeling approach for the detection of microcalcifications in mammograms by D. Sankar and T. Thomas [129]. Because modeling using fractal encoding takes a tremendous amount of time, the authors used mean and variance, the dynamic range of the image blocks, and mass center features to reduce the amount of time needed for encoding. This reduced the encoding time by a factor of 3, 89, and 13, respectively, in the three methods with respect to the conventional fractal image coding method with quad tree partitioning. The mammograms obtained from The Mammographic Image Analysis Society database (ground truth available) gave a total detection score of 87.6%, 87.6%, 90.5%, and 87.6%, for the conventional and the proposed three methods, respectively.

4.2 Theoretical Background

4.2.1 What is “Fractals”?



Sierpinski gasket



Koch curve

Fig. 4.1: Two famous examples for fractal images.

Straight lines, squares, triangles and circles are fundamental geometrical objects that we are all familiar with. Graphic designers often put together a number of them to create pretty patterns. Mathematicians on the other hand are more ambitious. They have learned to compose extraordinarily complex but yet highly regular patterns using infinitely many basic objects each of infinitesimal sizes. Fig. 4.1 shows two famous examples: the Sierpinski gasket and the Koch curve. In the 1970's, mathematician Benoit Mandelbrot named these and many other related geometrical objects 'fractals'.

Let us first take a closer look at the Sierpinski gasket. Its construction requires defining a generator, also called the level 1 object shown below. The objects at subsequent levels are then obtained by replacing all solid triangles in the previous level by downsized copies of the generator. To do the drawing by hand, we will find it very tiring after going up a few levels. However, in the very imaginative minds of mathematicians, we can easily go to infinite level and only then we obtain the Sierpinski gasket.

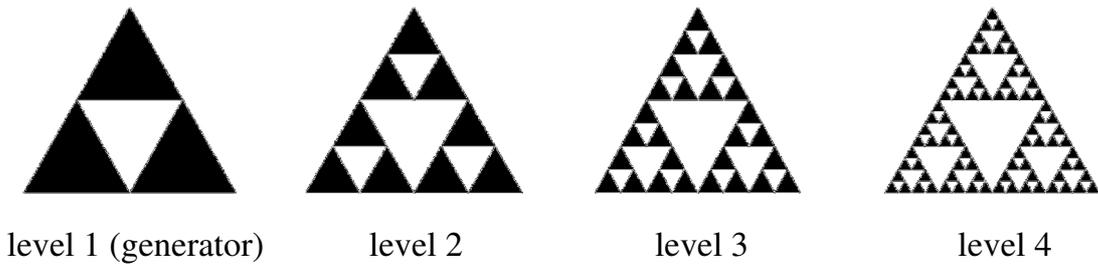


Fig. 4.2: Creation of fractal object levels.

4.2.2 Self-similarity

It is a property of fractal figures which smaller copies of the figure are embedded everywhere inside itself (e.g. Sierpinski gasket) and is an essential property of all types of fractals.

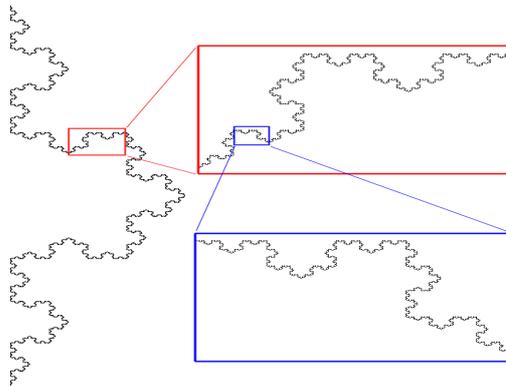


Fig. 4.3: Self-similarity property of fractal images.

4.2.3 Fractal Dimension

A filled triangle is two-dimensional (2-D), while a wire even after being bent into the shape of a triangular frame is still a one-dimensional object (1-D). On the other hand, the Sierpinski gasket is less solid than a filled triangle but much more bulky than a triangular frame. It is for this reason that mathematicians invented a generalization of the concept of dimension, which Mandelbrot called the fractal dimension. Very loosely speaking, it illustrates the degree of "closeness" to normal non-fractal objects with dimension one, two, three, etc. They found that the Sierpinski gasket is best described as having a dimension $\log 3 / \log 2 = 1.585!$ (It's dimension is between 1-D and 2-D).

4.2.4 Why do scientists study fractals?

This is because fractals are not only found in laboratories but are all around us in nature. Famous examples include coastline, mountains, river networks, clouds, blood vessels, broccoli, fern leaves, etc. In case of mammograms, when compared with microcalcifications, the breast background tissues have high local self-similarity, which is the basic property of fractal objects.

4.2.5 Fractal Modeling

Given a complete metric space (X, d) , we can define the metric space $(H(X), h)$, where $H(X)$ is the space of compact subsets of X , and the distance $h : H(X) \times H(X) \rightarrow R$ between two sets A and B is the Hausdorff distance, which is characterized in terms of the metric d . Under these conditions, it can be shown that the metric space $H(X)$ is complete according to the Hausdorff metric [130]. Let $f \in H(X)$ be an original image to be modeled. We wish to find contractive affine map $\tau: H(X) \rightarrow H(X)$, satisfying the requirement

$$\forall f_1, f_2 \in H(X), h(\tau(f_1), \tau(f_2)) \leq s \cdot h(f_1, f_2), \quad (4.1)$$

and such that

$$h(f, \tau(f)) < \delta \quad (4.2)$$

where $s < 1$ and δ is a tolerance which can be set to different values according to different applications. The scalar s is called the contractivity of τ . τ can be a set of contractive mappings τ_i , i.e., $\tau = U_{i=1}^N \tau_i$. According to the deterministic fractal theory, a set of contractive mappings τ_i is the main part of an iterated function system (*IFS*). The definition of *IFS* is given as follows [117].

Definition 1: An iterated function system (*IFS*) consists of a complete metric space (X, d) with a finite set of contraction mappings $\tau_i : X \rightarrow X$, with respective

contractivity factors s_i , for $i = 1, 2, \dots, N$, and its contractivity factor is $s = \max\{s_i : i = 1, 2, \dots, N\}$.

With the definition of *IFS*, one can state the important property of *IFS* in the following theorem.

Theorem 1: (The Collage Theorem) Let (X, d) be a complete metric space. Let $L \in H(X)$ be given, and let $\varepsilon \geq 0$ be given. Choose an *IFS* $\{X; \tau_i\}$ with contractivity factor $0 \leq s < 1$, so that

$$h(L, U_{n=0}^N \tau_n(L)) \leq \varepsilon. \quad (4.3)$$

Then $h(L, A) \leq \varepsilon / (1 - s)$, for all $L \in H(X)$, where A is the attractor of the *IFS*. [10]

The theorem tells us that to find an *IFS* whose attractor is "close to" or "looks like" a given set, one must endeavor to find a set of transformations (contraction mappings on a suitable space within which the given set lies) such that the union, or collage, of the images of the given set under the transformations is near to the given set. Nearness is measured using the Hausdorff metric. The proof of the Collage Theorem can be found in [117].

The Collage Theorem shows that, once an *IFS* is found, i.e., τ is known such that $h(f, \tau(f)) < \delta$ is satisfied, then from any given image f_0 and any positive integer n , one can get

$$h(f, \tau^{on}(f_0)) \leq \frac{1}{1-s} h(f, \tau(f)) + s^n h(f, f_0). \quad (4.4)$$

Since $s < 1$, we see that after a number of iterations, the constructed image $f_n = \tau^{on}(f_0)$ will be close visually to the original image f .

The key point of fractal modeling is to explore the self-similarity property of images. Real world images are seldom self-similar, so it is impossible to find a transformation τ for an entire image. But almost all real images have a local self-

similarity. We can divide the image into n small blocks, and for each block find a corresponding τ_i . So finally, we can define $\tau = U_{i=1}^N \tau_i$.

4.2.6 IFS Coding Steps

1. Let image $I = \text{gray scale} = u(x,y)$.
2. Partition the image into non-overlapping sub-blocks called Ranges (R_i).
3. Partition the image into larger domain blocks (D_i).
4. Search all possible set of D_i and set of transformations W_i and associate with R_i .
5. Pack the fractal codes.

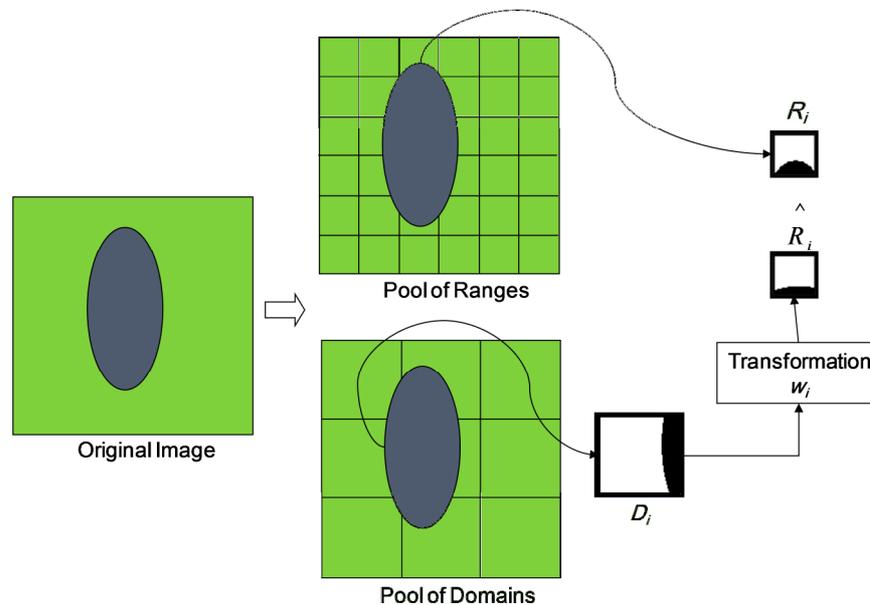


Fig. 4.4: Fractal IFS coding example with range R_i , domain D_i , and transformations W_i (here the transformations are rotation and resizing).

4.3 Algorithm Implementation

Jacquin had classified the image into shade, midrange and edge blocks [118], [119] but D. Sankar and T. Thomas said that the image blocks may be classified into shade and non shade blocks based on their visual perception [130]. In our system we used D. Sankar and T. Thomas method and only the non shade blocks are modeled using the fractal modeling method. Thus, the computation time required in the fractal modeling procedure can be considerably reduced.

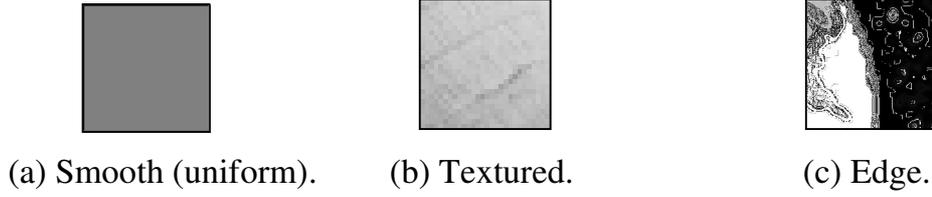


Fig. 4.5: shade and non-shade blocks of images
(a) shade blocks. (b) & (c) Non-Shade blocks.

The image of square size 64x64 is divided into non overlapping range blocks of size 8x8. These range blocks are then classified into shade and non shade blocks. Shade blocks are those blocks that has no major gradients or texture and the gray scale of pixels change slowly or little to human eyes perception. A non shade block has some sudden changes in pixel intensities looking like texture or distinct edges which can be perceived. To classify these blocks the dynamic range (ratio between max. and min. pixel values) of the block is calculated. The block is classified as shade block if the dynamic range is less than 0.05 else it is a non shade block and it has to be modeled by the following procedure.

Here, a mathematical representation for digital gray-level images is introduced. Let $N_1 = [0,1, \dots, M]$, $N_2 = [0,1, \dots, N]$, $N_3 = [0, 1, \dots, L]$, respectively, then for any digital gray-level image $f(k, l)$, we have $(k, l, f(k, l)) \in N_1 \times N_2 \times N_3$. Let D_1, \dots, D_n and R_1, \dots, R_n be subsets of $N_1 \times N_2$, such that $\bigcup_{i=1}^n R_i = N_1 \times N_2$ and $R_i \cap R_j = \emptyset, i \neq j$. We call R_i the range squares, and D_i the domain squares. τ_i can be defined as

$$\tau_i(f(k, l) = s_i \bar{f}(k, l)|_{(k,l) \in D_i} + o_i, \quad (4.5)$$

where s_i is a scaling factor and o_i is an offset factor. The error may be written as:

$$e_i = \sum_k \sum_l (f(k, l) - (s_i \bar{f}(k, l) + o_i))^2. \quad (4.6)$$

The main target in our system is: for each R_i , a $D_i \subset N_1 \times N_2$ and $\tau_i: N_1 \times N_2 \times N_3 \rightarrow N_3$ are sought such that the error is minimized. A value is set for the uniform tolerance $\delta_i = \delta$, and the best D_i is selected such that $e_i < \delta$.

Since we are processing only the non shade blocks, we consider that there is microcalcifications (clusters or some single isolated ones) on the image block above R_i , our intention is to find an area D_i on which the image has a similar structure as on R_i but does not have similar microcalcification patterns. Then when a difference between the original image and modeled image is taken, the microcalcifications will be enhanced. This means that when searching for D_i , the suitable D_i should not cover the region of R_i . In the proposed algorithm, for each given R_i , we constrain the search way of D_i by $R_i \cap D_i = \phi$.

4.3.1 Fractal Modeling:

The fractal modeling may be done via the following steps.

- 1) Choose R_i so that they are a non-overlapping subsquares of size 8x8.
- 2) Perform a search for D_i that satisfy $R_i \cap D_i = \phi$, and $e_i < \delta$ condition is satisfied. If this condition is not satisfied, the domain with minimum error is selected.
- 3) The process is continued until the whole image is modeled.
- 4) Based on the Collage Theorem, the modeled image can be obtained easily by iteration according to τ_i and D_i . The iteration stops when the predetermined tolerance between the original and the modeled image is achieved.

4.3.2 Microcalcifications enhancement:

Microcalcifications may be enhanced by using the fractal modeling in the following manner. Let the original and the modeled images be $f(k,l)$ and $g(k,l)$ respectively. The enhanced image (from which background structures were removed) may be achieved by subtracting the two images and ignoring the

negative values which does not contain any information about spots brighter than background (microcalcifications). It may be written as,

$$f_1(k,l) = \max(0, [f(k,l) - g(k,l)]), \quad (k,l) \in N_1 \times N_2. \quad (4.7)$$

4.3.3 MIAS database:

Due to privacy issues, real medical images are difficult to access for experimentation. The data used in our experiments for this part was taken from the Mammographic Image Analysis Society (MIAS) [58], which is an organization of UK research groups interested in the understanding of mammograms, has produced a digital mammography database.

The images in this database were scanned with a Joyce-Loebl microdensitometer SCANDIG-3, which has a linear response in the optical density range 0-3.2. Each pixel is 8-bits deep and at a resolution of 50 μ m x 50 μ m and clipped/padded so that every image is 1024 pixels x 1024 pixels. The database contains left and right breast images for 161 patients by Medio-Lateral Oblique (MLO) view. It consists of 322 images, which belong to three categories: normal, benign and malignant, which are considered abnormal. There are 208 normal, 63 benign and 51 malignant (abnormal) images. In addition, the abnormal cases are further divided into six categories: circumscribed masses, spiculated masses, microcalcifications, ill-defined masses, architectural distortion and asymmetry. They also include the locations of any abnormalities that may be present. The existing data in the collection consists of the location of the abnormality (like the center of a circle surrounding the tumor), its radius, breast position (left or right), type of breast tissues (fatty, fatty-glandular and dense) and tumor type if exists (benign or malign) see Fig. 4.6.

The MIAS database was associated with information file; this file lists the films in the MIAS database and provides appropriate details as follows:

- 1st column: MIAS database reference number.
- 2nd column: Character of background tissue:

F - Fatty

G - Fatty-glandular

D - Dense-glandular

- 3rd column: Class of abnormality present:

CALC - Calcification

CIRC - Well-defined/circumscribed masses

SPIC - Spiculated masses

MISC - Other, ill-defined masses

ARCH - Architectural distortion

ASYM - Asymmetry

NORM - Normal

- 4th column: Severity of abnormality;

B - Benign

M - Malignant

- 5th, 6th columns: x, y image-coordinates of centre of abnormality.
- 7th column: Approximate radius (in pixels) of a circle enclosing the abnormality.

Also; important notes included in this file were summarized in four points:

1. The list is arranged in pairs of films, where each pair represents the left (even filename numbers) and right mammograms (odd filename numbers) of a single patient.
2. The size of ALL the images is 1024 pixels x 1024 pixels. The images have been centered in the matrix.
3. When calcifications are present, centre locations and radii apply to clusters rather than individual calcifications. Coordinate system origin is the bottom-left corner.
4. In some cases calcifications are widely distributed throughout the image rather than concentrated at a single site. In these cases centre locations and radii are inappropriate and have been omitted.

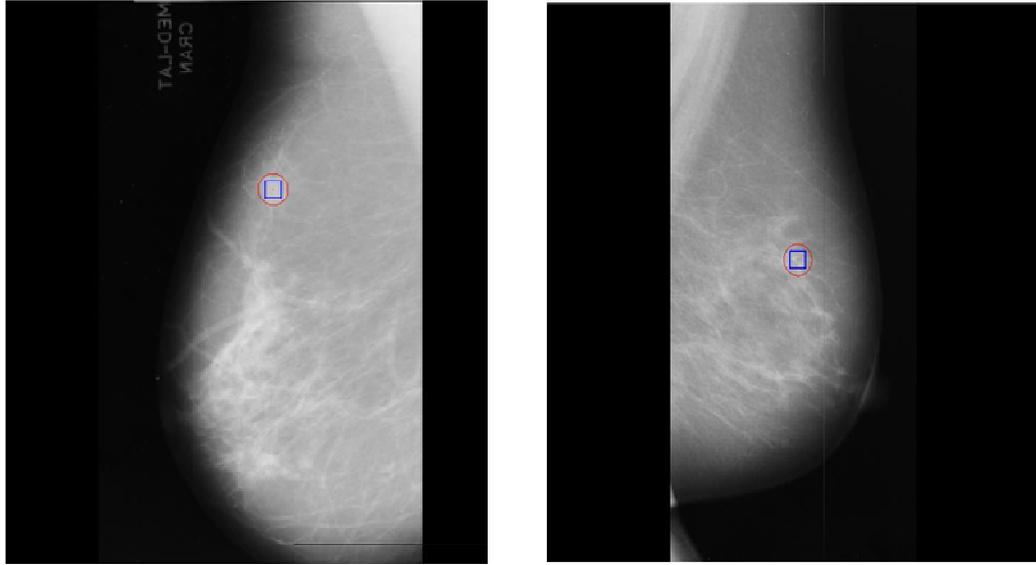


Fig. 4.6: Examples of mammograms images from MIAS database [58].

4.3.4 ROI Selection:

Taking the guidance from the locations of abnormalities (microcalcifications) supplied by the MIAS, the ROI of size 64×64 pixels was extracted with Microcalcifications centered in the sub-image. The ROIs selected were 100 normal and 25 abnormal images (which are all microcalcification images in the database).

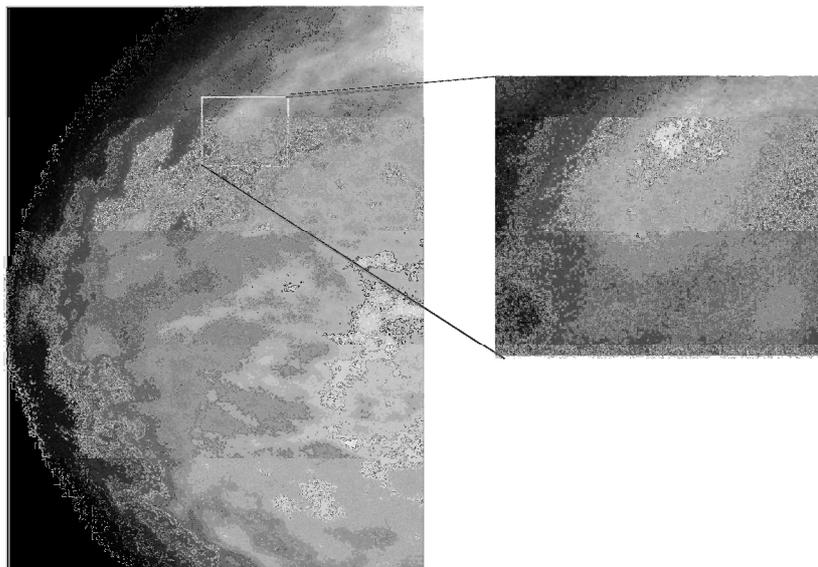


Fig. 4.7: Mammograms obtained from MIAS database and the selected ROI with dimensions of 64 x 64 pixels [58].

4.3.5 Features Extraction

Features are extracted from the original and the enhanced ROIs. We computed the contrast, the peak signal to noise ratio, and the average signal to noise ratio. The contrast C is defined by:

$$C = \frac{f-b}{f+b}, \quad (4.8)$$

where f is the mean gray-level value of a particular object in the image, called the foreground, and b is the mean gray-level value of a surrounding region called background.

The peak and average signal to noise ratio ($PSNR$) & ($ASNR$) are defined as:

$$PSNR = \frac{p-b}{\sigma}, \quad (4.9)$$

$$ASNR = \frac{f-b}{\sigma}, \quad (4.10)$$

where p is the maximum gray-level value of a foreground. And σ is the standard derivation in the background region.

4.3.6 Classification

There are different types of classifiers. Through this part, we used Support Vector Machine (SVM) classifier to classify between normal and abnormal cases. SVM is a learning tool originated in modern statistical learning theory [131]. In recent years, SVM learning has found a wide range of real-world applications, including handwritten digit recognition, object recognition, speaker identification, face detection in images, and text categorization. The formulation of SVM learning is based on the principle of structural risk minimization. Instead of minimizing an objective function based on the training samples [such as Mean Square Error (MSE)], the SVM attempts to minimize a bound on the generalization error (i.e., the error made by the learning machine on test data not

used during training). As a result, an SVM tends to perform well when applied to data outside the training set. Indeed, it has been reported that SVM-based approaches are able to significantly outperform competing methods in many applications [131]. SVM achieves this advantage by focusing on the training examples that are most difficult to classify. These “borderline” training examples are called support vectors.

Classical learning approaches are designed to minimize error on the training data set and it is called the empirical risk minimization (ERM). Those learning methods follow the ERM principle and neural networks are the most common example of ERM. On the other hand, the SVM is based on the structural risk minimization (SRM) principle rooted in the statistical learning theory. It gives better generalization abilities and SRM is achieved through a minimization of the upper bound of the generalization error [132].

SVM has the potential to handle very large feature spaces, because the training of SVM is carried out so that the dimension of classified vectors does not have as distinct an influence on the performance of SVM as it has on the performance of conventional classifier. That is why it is noticed to be especially efficient in large classification problem. This will also benefit in faults classification, because the number of features to be the basis of fault diagnosis may not have to be limited. Also, SVM-based classifier is claimed to have good generalization properties compared to conventional classifiers, because in training SVM classifier the so-called structural misclassification risk is to be minimized, whereas traditional classifiers are usually trained so that the empirical risk is minimized. The performance of SVM in various classification task is reviewed, e.g., in Christiani and Shawe-Taylor [132].

Given data input x_i ($i=1, 2, \dots, M$), M is the number of samples. The samples are assumed have two classes namely positive class and negative class. Each of classes associate with labels be $y_i=1$ for positive class and $y_i=-1$ for negative class, respectively. In the case of linear data, it is possible to determine the hyperplane $f(x) = 0$ that separates the given data

$$f(x) = w^T x + b = \sum_{j=1}^M w_j x_j + b = 0, \quad (4.11)$$

where w is M -dimensional vector and b is a scalar.

The vector w and scalar b are used to define the position of separating hyperplane. The decision function is made using sign $f(x)$ to create separating hyperplane that classify input data in either positive class and negative class. A distinct separating hyperplane should be satisfy the constraints

$$\begin{aligned} f(x_i) &= 1 \quad \text{if } y_i = 1, \\ f(x_i) &= -1 \quad \text{if } y_i = -1, \end{aligned} \quad (4.12)$$

or can be presented in complete equation

$$y_i f(x_i) = y_i (w^T x_i + b) \geq 1 \quad \text{for } i = 1, 2, \dots, M. \quad (4.13)$$

The separating hyperplane that creates the maximum distance between the plane and the nearest data, i.e., the maximum margin, is called the optimal separating hyperplane. An example of the optimal hyperplane of two data sets is presented in Fig. 4.8.

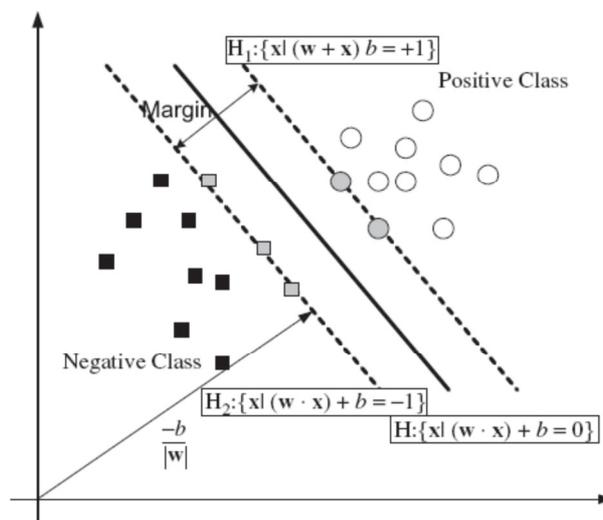


Fig. 4.8: Classification of two classes using SVM classifier [132].

In Fig. 4.8, a series data points for two different classes of data are shown, black squares for negative class and white circles for positive class. The SVM tries to place a linear boundary between the two different classes, and orientate it in such way that the margin represented by the dotted line is maximized. Furthermore, SVM attempts to orient the boundary to ensure that the distance between the boundary and the nearest data point in each class is maximal. Then, the boundary is placed in the middle of this margin between two points. The nearest data points that used to define the margin are called support vectors, represented by the grey circles and squares. When the support vectors have been selected, the rest of the feature set is not required, as the support vectors can contain all the information-based need to define the classifier. From the geometry the geometrical margin is found to be $\|w\|^{-2}$.

Assuming that the data is linearly separable, we seek to find the smallest possible w or maximum separation (margin) between the two classes. This can be formally expressed as a quadratic optimization problem:

$$\min_{w \neq 0, b} \frac{1}{2} \|w\|^2, \quad (4.14)$$

$$\text{Subject to } y_i(w^T x_i + b) \geq 1 \quad \forall i = 1, \dots, M. \quad (4.15)$$

Taking into account the noise with slack variables x_i and the error penalty C , the optimal hyperplane separating the data can be obtained as a solution to the following optimization problem:

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M \xi_i, \quad (4.16)$$

$$\text{Subject to } \begin{cases} y_i(w^T x_i + b) \geq 1 - \xi_i, & i = 1, \dots, M, \\ \xi_i \geq 0, & i = 1, \dots, M, \end{cases} \quad (4.17)$$

where ξ_i is measuring the distance between the margin and the examples x_i that lying on the wrong side of the margin.

The calculation can be simplified by converting the problem with Kuhn–Tucker condition into the equivalent Lagrangian dual problem, which will be

$$\text{Minimize } L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^M \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^M \alpha_i. \quad (4.18)$$

The task is minimizing Eq. (4.18) with respect to w and b , while requiring the derivatives of L to α to vanish. At optimal point, we have the following saddle-point equations:

$$\frac{\partial L}{\partial w} = 0, \quad \frac{\partial L}{\partial b} = 0, \quad (4.19)$$

which replace into form

$$w = \sum_{i=1}^M \alpha_i y_i x_i, \quad \sum_{i=1}^M \alpha_i y_i = 0. \quad (4.20)$$

From Eq. (4.20), we find that w is contained in the subspace spanned by the x_i . Using substitution Eq. (4.20) into Eq. (4.18), we obtain the dual quadratic optimization problem

$$\text{Maximize } L(\alpha) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,l=0}^M \alpha_i \alpha_l y_i y_l x_i \cdot x_l, \quad (4.21)$$

$$\text{subject to } \alpha_i \geq 0, i = 1, \dots, M, \quad \sum_{i=1}^M \alpha_i y_i = 0. \quad (4.22)$$

Thus, by solving the dual optimization problem, one obtains the coefficients α_i which is required to express the w to solve Eq. (4.16). This leads to non-linear decision function.

$$f(x) = \text{sign} \left(\sum_{i,j=1}^M \alpha_i y_i (x_i x_j) + b \right). \quad (4.23)$$

SVM can also be used in non-linear classification tasks with application of kernel functions. The data to be classified is mapped onto a high-dimensional feature space, where the linear classification is possible. Using the non-linear vector function $\Phi(x) = (\varphi_1(x), \dots, \varphi_l)$ to map the n -dimensional input vector x onto l dimensional feature space, the linear decision function in dual form is given by

$$f(x) = \text{sign} \left(\sum_{i,j=1}^M \alpha_i y_i (\Phi^T(x_i) \cdot \Phi(x_j)) + b \right). \quad (4.24)$$

Working in the high-dimensional feature space enables the expression of complex functions, but it also generates the problem. Computational problem occur due to the large vectors and the overfitting also exists due to the high-dimensionality. The latter problem can be solved by using the kernel function. Kernel is a function that returns a dot product of the feature space mappings of the original data points, stated as $K(x_i, x_j) = (\Phi^T(x_i) \cdot \Phi(x_j))$. When applying a kernel function, the learning in the feature space does not require explicit evaluation of Φ and the decision function will be

$$f(x) = \text{sign} \left(\sum_{i,j=1}^M \alpha_i y_i K(x_i, x_j) + b \right). \quad (4.25)$$

Any function that satisfies Mercer's theorem can be used as a kernel function to compute a dot product in feature space [132]. There are different kernel

functions used in SVM, such as linear, polynomial and Gaussian RBF were evaluated and formulated in Table 4.1.

Table 4.1: Formulation of kernel functions

Kernel functions	Equations
Linear Function	$K(x, x_j) = x \cdot x_j$
Polynomial Function	$K(x, x_j) = (x \cdot x_j + r)^d$
Radial Basis Function (RBF)	$K(x, x_j) = \exp\left(-\frac{\ x - x_j\ ^2}{\sigma^2}\right)$

The selection of the appropriate kernel function is very important, since the kernel defines the feature space in which the training set examples will be classified. The definition of legitimate kernel function is given by Mercer's theorem. The function must be continuous and positive definite.

Another SVM classifier is called least square support vector machine (LSSVM). The most important difference between SVM and LSSVM is that LSSVM uses a set of linear equations for training while SVM uses a quadratic optimization problem. While Eq. (4.11) is minimized subject to Eq. (4.13) in standard SVM, in LSSVM Eq. (4.26) is minimized subject to Eq. (4.27).

$$\text{Minimize } \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^M \xi_i^2, \quad (4.26)$$

$$\text{subject to } \begin{cases} y_i (w^T x_i + b) = 1 - \xi_i, & i = 1, \dots, M, \\ \xi_i \geq 0, & i = 1, \dots, M. \end{cases} \quad (4.27)$$

Another difference between SVM and LSSVM is that α_i (Lagrange multipliers) is positive or negative in LSSVM but they must be positive in SVM.

4.4 Results & Discussions

Table 4.2: Summary of fractal coding results

Method	Mammograms	Samples	Sensitivity	Specificity	Average time in minutes
Conventional	Normal	100	-	94%	24.35
	Abnormal	25	92%	-	23.08
Proposed	Normal	100	-	97%	0.82
	Abnormal	25	92%	-	2.21

All results from the proposed system are shown in table 4.2 where we repeated all the work using both the conventional fractal modeling and our proposed system. The encoding time for the conventional method of fractal coding was 23.72 minutes in average, while the proposed method took only 1.51 minutes when encoding normal and abnormal mammograms. Thus a saving of 93.63% of the encoding time is obtained using the proposed fractal modeling method.

It is clear from the table that we measured, quantitatively, the detection performance of the each classifier by computing the sensitivity and specificity on the data. The sensitivity is the conditional probability of detecting a disease while there is in fact a cancerous breast. The specificity is the conditional probability of detecting a normal breast while the breast is indeed normal. The proposed system provided a higher specificity percentage which means saving patients from the unnecessary biopsy.

Fig. 4.9 shows an example of how the system works, where the image is divided into ranges and domains, then the homogeneous ranges are discarded. Only the non-shade blocks are coded with the proposed fractal coding algorithm and the resulted image is provided. Finally the enhanced image is presented with only the microcalcifications in it.

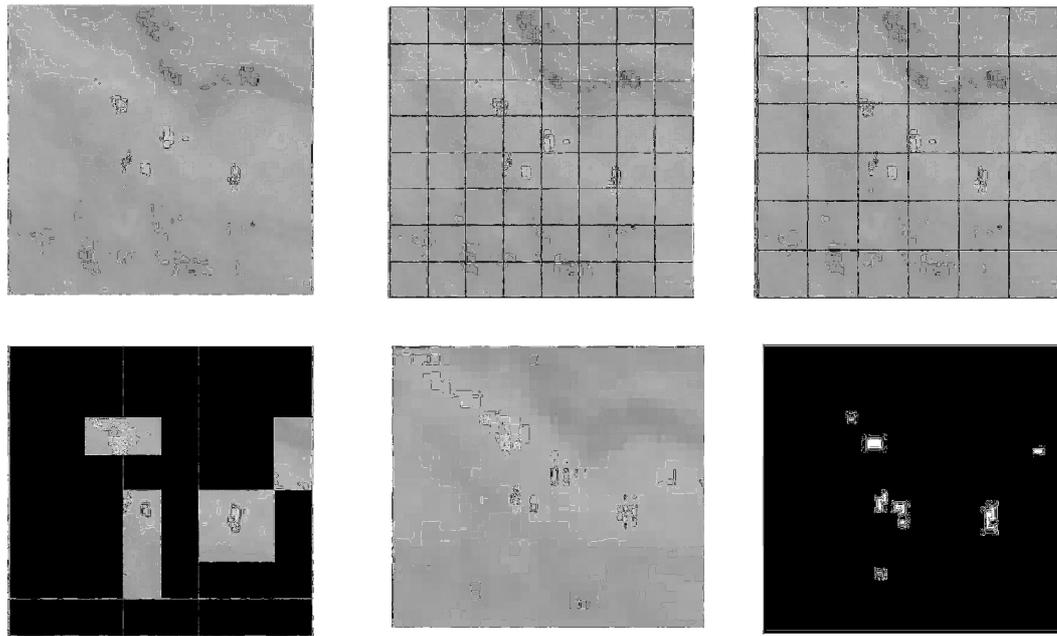


Fig. 4.9: Example of how the system works.

4.5 Summary

In this study, a proposed system for fast fractal modeling of mammograms for microcalcifications detection is presented. The selected ROI is divided into non overlapping range blocks, these blocks are then classified into shade and non shade blocks according to their dynamic range. This system depends on mammographic microcalcification enhancement using the Collage Theorem for fractal modeling of only the non shade blocks.

All results obtained in this study are very encouraging, and indicate that the proposed fractal modeling method is an effective technique to extract mammographic patterns and to enhance microcalcifications embedded in inhomogeneous breast tissues, and this is done faster than the conventional method. Therefore, the proposed method may facilitate the radiologists' diagnosis of breast cancer at an early stage.

CHAPTER 5

UNSUPERVISED CLUSTERING/BICLUSTERING FOR ABNORMALITIES DETECTION

5.1 Introduction

A new methodology for computer aided diagnosis in digital mammography using unsupervised classification and class-dependent feature selection is presented in this part. In spite of the success of the supervised classification methods in improving the overall diagnosis, the labeling of training image samples limits the classification to classes that are deemed independent ignoring the relationships between different pathology types and the progression of each. As a result, images that represent stage of transition between different types are often misclassified. Therefore, a technique that would look at the unlabeled data and provide unsupervised classes would provide an insight into that thus improving the overall effectiveness of the diagnosis.

Moreover, using unsupervised classification methods can be extended to utilize biclustering methods which allow for definition of unsupervised clusters of both pathologies and features. That is, for each pathology type, a particular set of features that can diagnose this particular type are defined. This has potential to provide more flexibility, and hence better diagnostic accuracy, than the commonly used feature selection strategies.

The proposed system consists of four stages: preprocessing, feature extraction, feature matrix visualization, and unsupervised clustering/biclustering. The developed methods are applied to diagnose digital mammographic images from the Mammographic Image Analysis Society (MIAS) database. This system leads to better classification results for the data based on the K-means method. The system has potential in providing more insight into data and show the value for exploratory data analysis methods.

5.2 Review

Several methodologies have been developed in order to improve the interpretation of mammograms. Among those, the incorporation of computer aided diagnosis (CAD) provides a tool that works as a second observer to the radiologist. CAD systems have been demonstrated as effective tools for helping radiologist identify malignancies in mammograms. Various techniques were developed to detect and classify masses and microcalcifications in digital mammograms. Most of these techniques used supervised classification to derive their decision. In spite of the success of these methods in improving the overall diagnosis, the labeling of training image samples limits the classification to classes that are deemed independent ignoring the relationships between different pathology types and the progression of each. As a result, images that represent stage of transition between different types are often misclassified. Therefore, a technique that would look at the unlabeled data and provide unsupervised classes would provide an insight into that thus improving the overall effectiveness of the diagnosis. Moreover, using unsupervised classification methods can be extended to utilize biclustering methods which allow for definition of unsupervised clusters of both pathologies and features [141], [142]. That is, for each pathology type, a particular set of features that can diagnose this particular type are defined. This has potential to provide more flexibility, and hence better diagnostic accuracy, than the commonly used feature selection strategies.

Unsupervised classifiers, such as K-mean clustering, fuzzy C-mean clustering, and self-organizing maps, can be used in the literature to diagnose breast cancer. Kim et al. [133] designed a new type of classifier combining an unsupervised and a supervised model and applied to classification of malignant and benign masses on mammograms. The unsupervised model was based on an Adaptive Resonance Theory (ART2) network that clustered the masses into a number of separate classes. Lee et al. [134] used K-means clustering for classifying unlabeled MRI data. Howard et al. [135] used a SONNET self-organizing to produce a taxonomic organization of the mammography archive in an unsupervised manner. Chen et al. [136] used k-means classifier to classify

breast ultrasound images to benign and malignant. Meyer-Baese et al. [137] introduced automatic lesion segmentation and classification system based on unsupervised clustering and ICA techniques for breast MRI images. None of these methods targeted the development of a CAD system and also biclustering was never considered as a tool for feature selection.

In this chapter, a new CAD system for digital mammograms using unsupervised classification is presented. The potential of using this technique is demonstrated in improving the accuracy of the overall diagnosis. The proposed system consists of four stages: preprocessing, feature extraction, feature matrix visualization, and unsupervised clustering/biclustering.

5.3 Preprocessing

In the preprocessing, the region of interest (ROI) was selected from the digital mammograms images. The data used in our experiments were obtained from the Mammographic Image Analysis Society (MIAS) database [58]. It consists of 322 images belonging to normal, benign and malignant classes. All images were had a resolution of 1024×1024 pixels and 8-bit accuracy (gray level). They also include the locations of any abnormalities that may be present. The existing data consist of the location of the abnormality (like the center of a circle surrounding the tumor), its radius, breast position (left or right), type of breast tissues (fatty, fatty-glandular and dense) and tumor type if exists (benign or malign). Using the locations of abnormalities supplied by the MIAS for each image, a ROI of size 32×32 pixels is extracted with breast cancer centered in the window. We used 100 images for normal cases, 88 images for masses (circumscribed, spiculated, ill-defined, architectural distortion, and asymmetric) cases (51 benign images and 37 malignant images), and 25 images for microcalcification cases (13 benign images and 12 malignant images).

5.4 Feature Extraction

A typical mammogram contains a vast amount of heterogeneous information that depicts different tissues, vessels, ducts, chest skin, breast edge, the film, and

the X-ray machine characteristics. In order to build a robust diagnostic system towards correctly classifying normal and abnormal regions of mammograms and then classify between benign and malignant regions, we have to present all the relevant information in mammograms to the diagnostic system so that it can discriminate between different pathologies effectively. However, the use of all the information results to high dimensionality of feature vectors that degrade the diagnostic accuracy of the utilized systems significantly in addition to sharply increasing their computational complexity. Therefore, a reduced set of reliable features should be considered that summarize only the relevant information. In our approach, we examined an initial set of 224 features obtained from the ROI. These features can be divided into five categories:

5.4.1 Wavelet features (136 features) [3],

Features were extracted from the coefficients that were produced by the wavelet decomposition process on ROI. There are five processing steps in features extraction stage:

- 1) Wavelet decomposition.
- 2) Coefficients extraction.
- 3) Normalization.
- 4) Energy computation.
- 5) Features reduction.

5.4.2 First order statistical features (18 features) [89].

It provides different statistical properties of the intensity histogram of an image. They depend only on individual pixel values.

- Mean
- Standard Deviation
- Kurtosis
- Skewness
- Variance
- Maximum of Gray Level
- Minimum of Gray Level
- Entropy
- Second Central Moment
- 9 Percentiles

5.4.3 Second order statistical features (60 features) [138], [139].

The gray level co-occurrence matrix (GLCM) features. The GLCM is a well-established robust statistical tool for extracting second order texture information from images. Four GLCMs corresponding to four different directions ($\theta=0^\circ$, 45° , 90° and 135°) and one distance ($d=1$ pixel), were computed for each selected ROI region.

Fifteen features were derived from each GLCM, four values were obtained for each feature corresponding to the four matrices, and this corresponds to 60 features matrix.

- Energy
- Contrast
- Homogeneity
- Entropy of GLCM Matrix
- Information Correlation 2
- 1st Order Difference Moment
- 2nd Order Inverse Difference Moment
- Max. of GLCM Matrix
- Correlation
- Sum of Squares
- Sum Average
- Sum Entropy
- Difference Entropy
- Cluster Shade
- Prominence

5.4.4 Shape features (8 features) [89], [77].

The shape features used are the spreadness and the seven invariant moments. Spreadness shows the degree of spread of the shape around the centered intuitively, i.e. measure the circularity of tumor within the ROI. And seven invariant moments which are a set of moments that is invariant to translation, rotation, and scale change.

5.4.5 Fractal dimension features (2 features) [92].

The FD offers a quantitative measure of self-similarity and scaling. In this study two methods were used to estimate the fractal dimension feature, the piecewise modified box-counting (PMBC) method and the piecewise triangular prism surface area (PTPSA) method.

Table 5.1: Summary of the extracted features

1-136 Wavelet Features	171-174 Energy
137 Mean	175-178 Homogeneity
138 Variance	179-182 First order Difference Moment
139 Skewness	183-186 Max of GLCM Matrix
140 Kurtosis	187-190 Prominence
141 Max of Gray Level	191-194 Entropy of GLCM Matrix
142 Entropy	195-198 Second Order Inverse Difference Moment
143 Min of Gray Level	199-202 Cluster Shade
144 Second Central Moment	203-206 Information Correlation 2
145 Standard Deviation	207-210 Sum of Squares
146 Spreadness	211-214 Difference Entropy
147-153 Seven Invariant Moments	215-218 Sum Entropy
154-162 Nine Percentiles	219-222 Sum Average
163-166 Contrast	223 Fractal Dimension By PMBC
167-170 Correlation	224 Fractal Dimension By PTPSA

From all the above features we construct a feature matrix with images as rows and features as columns.

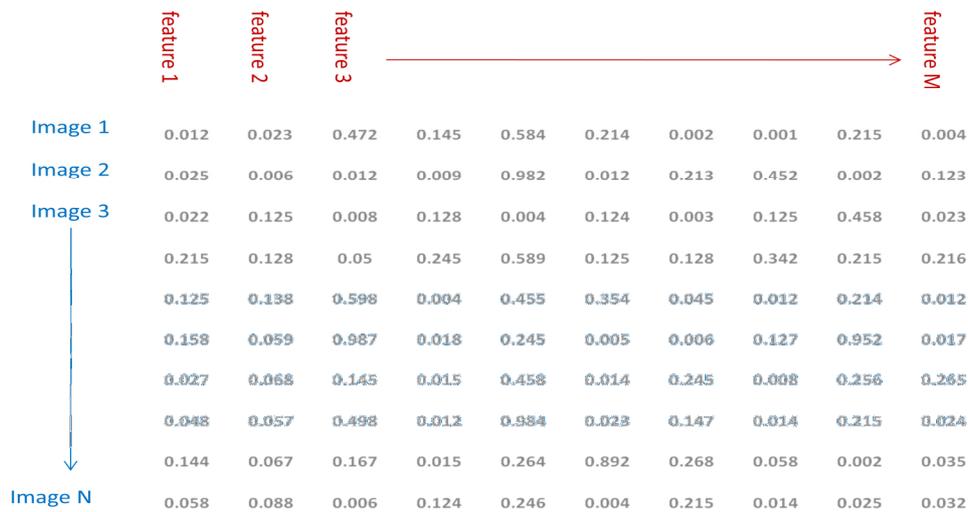


Fig. 5.1: The features matrix

5.5 Feature matrix visualization

In order to address the problem of feature set reduction, the feature matrix is displayed as a heatmap, where feature values are represented by color map [140]. This technique is commonly used in the area of gene expression data analysis where the gene expression matrix is visualized this way. Given the different scales of different features, it is necessary to normalize all features to the same range in order for their range of values to appear properly on the map. The advantages of feature matrix visualization include the ease of data interpretation by visualization rather than figures, gain better understanding of how features perform with different cases.

5.6 Unsupervised Clustering/Biclustering Techniques

5.6.1 Clustering

Clustering [143] is the assignment of objects into groups (called clusters) so that objects from the same cluster are more similar to each other than objects from different clusters. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics.

Clustering analysis is a fundamental but important tool in statistical data analysis. In the past, the clustering techniques have been widely applied in a variety of scientific areas such as pattern recognition, information retrieval, microbiology analysis, and so forth. In its basic form the clustering problem is defined as the problem of finding homogeneous groups of data points in a given data set, each of which is referred to as a cluster.

Clustering techniques or unsupervised classifiers are able to discover clusters inherent in the data. This technique considers unlabeled data and provides unsupervised classes that give a better insight into classes and their interrelationships, thus improving the overall effectiveness of the diagnosis. An unsupervised classifier differs from a supervised classifier in two ways: (1) it

works with unlabeled data, and (2) training of a classification rule with the use of a training data set is not required.

Among those techniques, the k-means clustering algorithm, which is a squared error based clustering algorithm, partitions a given data set into k mutually exclusive clusters such that the sum of the distances between data and the corresponding cluster centroid is minimized. The above distance measure between two data points is taken as a measure of similarity.

Often similarity is assessed according to a distance measure i.e., each clustering problem is based on some kind of “distance” between points. There are two major classes of distance measure "Euclidean and Non-Euclidean". A Euclidean distance is based on the locations of points in such a space. But A Non-Euclidean distance is based on properties of points not their “location” in a space.

Data clustering algorithms can be hierarchical. Hierarchical algorithms find successive clusters using previously established clusters. These algorithms can be either agglomerative ("bottom-up") or divisive ("top-down"). Agglomerative algorithms begin with each element as a separate cluster and merge them into successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters.

Partitional algorithms typically determine all clusters at once, but can also be used as divisive algorithms in the hierarchical clustering.

Density-based clustering algorithms are devised to discover arbitrary-shaped clusters. In this approach, a cluster is regarded as a region in which the density of data objects exceeds a threshold. DBSCAN and OPTICS are two typical algorithms of this kind.

Two-way clustering, co-clustering or biclustering are clustering methods where not only the objects are clustered but also the features of the objects, i.e., if the data is represented in a data matrix, the rows and columns are clustered simultaneously.

Many clustering algorithms require specification of the number of clusters to produce in the input data set, prior to execution of the algorithm.

5.6.1.1 Hierarchical clustering

Hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure called a dendrogram. The root of the tree consists of a single cluster containing all observations, and the leaves correspond to individual observations.

Algorithms for hierarchical clustering are generally either agglomerative, in which one starts at the leaves and successively merges clusters together; or divisive, in which one starts at the root and recursively splits the clusters.

Any valid metric may be used as a measure of similarity between pairs of observations. The choice of which clusters to merge or split is determined by a linkage criteria, which is a function of the pair-wise distances between observations. Cutting the tree at a given height will give a clustering at a selected precision

5.6.1.2 Partitional clustering

- ***k*-means clustering**

The *k*-means algorithm [144] assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster — that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster.

Mathematically, given a set of data vectors $X = [x_1, \dots, x_n]$ where n is the number of observations, the *k*-means clustering algorithm groups the data into k clusters with the aim at minimizing an objective function, a squared error function. Therefore, the *k*-means clustering algorithm is an iterative algorithm that finds a suitable partition which minimizes the sum squared error. The algorithm begins with the initialization of k cluster centroids. Different approaches in initialization have been suggested. A simple method is to initialize the problem by randomly select k data points from the given data. The remaining data points are classified into the k clusters by distance. The centroids are then updated by computing the centroids in the k clusters.

Example: The data set has three dimensions and the cluster has two points: $X = (x_1, x_2, x_3)$ and $Y = (y_1, y_2, y_3)$. Then the centroid Z becomes $Z = (z_1, z_2, z_3)$, where $z_1 = (x_1 + y_1)/2$ and $z_2 = (x_2 + y_2)/2$ and $z_3 = (x_3 + y_3)/2$.

The algorithm steps are:

- Choose the number of clusters, k .
- Randomly generate k clusters and determine the cluster centers, or directly generate k random points as cluster centers.
- Assign each point to the nearest cluster center.
- Re-compute the new cluster centers.
- Repeat the two previous steps until some convergence criterion is achieved.

The main advantages of this algorithm are its simplicity and speed which allows it to run on large datasets [145]. Its disadvantage is that it does not yield the same result with each run, since the resulting clusters depend on the initial random assignments. It minimizes intra-cluster variance, but does not ensure that the result has a global minimum of variance. Another disadvantage is the requirement for the concept of a mean to be definable which is not always the case. For such datasets the k -medoids variant is appropriate. Other popular variants of K -means include the Fast Genetic K -means Algorithm (FGKA) [146] and the Incremental Genetic K -means Algorithm (IGKA) [147]. In this work, the K -Means was performed by using Expander toolbox [140].

- **Fuzzy c -means clustering**

In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the center of cluster. For each point x we have a coefficient giving the degree of being in the k^{th} cluster $u_k(k)$. Usually, the sum of those coefficients for any given x is defined to be 1:

$$\forall x \left(\sum_{k=1}^{\text{num. clusters}} u_k(x) = 1 \right). \quad (5.1)$$

With fuzzy c -means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$\text{center}_k = \frac{\sum_x u_k(x)^m x}{\sum_x u_k(x)^m}. \quad (5.2)$$

The degree of belonging is related to the inverse of the distance to the cluster center:

$$u_k(x) = \frac{1}{d(\text{center}_k, x)}, \quad (5.3)$$

then the coefficients are normalized and fuzzyfied with a real parameter $m > 1$ so that their sum is 1. So

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(\text{center}_k, x)}{d(\text{center}_j, x)} \right)^{2/(m-1)}}. \quad (5.4)$$

For m equal to 2, this is equivalent to normalizing the coefficient linearly to make their sum 1. When m is close to 1, then cluster center closest to the point is given much more weight than the others, and the algorithm is similar to k -means. The fuzzy c -means algorithm is very similar to the k -means algorithm [148] in:

- Choose a number of clusters.
- Assign randomly to each point coefficients for being in the clusters.
- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than ε , the given sensitivity threshold) :

- Compute the centroid for each cluster, using the formula above.
- For each point, compute its coefficients of being in the clusters, using the formula above.

The algorithm minimizes intra-cluster variance as well, but has the same problems as k -means, the minimum is a local minimum, and the results depend on the initial choice of weights. The expectation-maximization algorithm is a more statistically formalized method which includes some of these ideas: partial membership in classes. It has better convergence properties and is in general preferred to fuzzy-c-means.

The limitation of clustering algorithms is that all features are given equal weights in the computation of image similarity. However, some features do not contribute information and instead increase the amount of cluster ambiguity. Moreover, this technique assigns each image to a single cluster, whereas images at different degrees of disease progression may be mid-way between clusters (e.g., normal tissue turning into tumor). An effective means for dealing with such overlaps is through the use of biclustering methods.

5.6.2 Biclustering

Biclustering, co-clustering, or two-mode clustering [149], [150] is a data mining technique which allows simultaneous clustering of the rows and columns of a matrix. Given a set of m rows in n columns (i.e., an $m \times n$ matrix), the biclustering algorithm generates biclusters - a subset of rows which exhibit similar behavior across a subset of columns, or vice versa.

The difference between clustering and biclustering methods is that clustering methods can be applied to either the rows or the columns of the feature matrix, separately whereas biclustering methods, perform clustering in the two dimensions simultaneously. This means that clustering methods derive a global model while biclustering algorithms produce a more effective local model [151].

When clustering algorithms are applied on feature matrix, each image in a given image cluster is defined using all the features. However, each image in a

bicluster is selected using only a subset of the features. The goal of biclustering techniques is thus to identify subgroups of images and subgroups of features, by performing simultaneous clustering of both rows and columns of the feature matrix, where the images exhibit highly correlated activities for every feature, instead of clustering these two dimensions separately. As a result, biclustering was used in this work to find a set of the images participating in a common pathology of interest while defining a subset of features that best describe this pathology. We can then conclude that, unlike clustering algorithms, biclustering algorithms identify groups of images that show similar activity patterns under a specific subset of the features.

There are many biclustering algorithms developed mainly for bioinformatics, including: Block clustering, CTWC, ITWC, δ -bicluster, δ -pCluster, δ -pattern, FLOC, OPC, Plaid Model, OPSMs, Gibbs, SAMBA, Robust Biclustering Algorithm (RoBA), Crossing Minimization, cMonkey [152], PRMs, DCC and LEB (Localize and Extract Biclusters). Biclustering algorithms have also been proposed and used in other application fields under the names co-clustering, biodimensional clustering, and subspace clustering [151].

There are many biclustering algorithms, which differ in its approaches, time complicity and prediction ability. Table 5.2 summarizes these algorithms,

In this work, we applied the biclustering techniques to show how the biclusters will match up with the known pathologies and to find the features relevant to these pathologies. We used four prominent biclustering techniques which used for gene expression matrix; the Binary inclusion-maximal (Bimax) algorithm [153], the Cheng and Church (CC) algorithm [154], the Order Preserving Sub-Matrix (OPSM) algorithm [155], and the Statistical Algorithmic Method for Bicluster Analysis (SAMBA) [142]. The Bimax, CC, And OPSM techniques were performed by using the BicAT toolbox [141] and the SAMBA technique was performed by using Expander Toolbox [140]. After computing the bicluster files, BicOverlapper toolbox [157] was used to pictorially illustrate the found biclusters and to visualize the interaction between the biclusters

Table 5.2: Summary of biclustering algorithms

Algorithm	Author	Approach	Time Complicity	Prediction ability
Bivisu/ <i>pClusters</i>	Kin-On Cheng <i>et al.</i> , 2008 Haixun Wang, 2002	<i>Exhaustive Bicluster Enumeration</i>	$O(m^2n \log m)$	Coherent values
RMSBE	Xiaowen Liu and Lusheng Wang, 2006	<i>Greedy Iterative Search</i>	$O((n + m)^2)$	Coherent values
Bimax	Preli <i>et al.</i> , 2006	Divide -and-Conquer	$O(n m \beta \log \beta)$	Coherent values
ROBA	Alain B. Tchagang and Ahmed H. Tewfik, 2005	Matrix algebra	$O(n \times m \times L \times Nb)$	Coherent Evolution
x-motif	<u>Murali and Kasif, 2003</u>	<i>Greedy Iterative Search</i>		Coherent Evolution
SAMBA	<u>Tanay <i>et al.</i>, 2002</u>	<i>Exhaustive Bicluster Enumeration</i>	$O(n2^d)$	Coherent Evolution
OPSM	<u>Ben-Dor <i>et al.</i>, 2002</u>	<i>Greedy Iterative Search</i>		Coherent Evolution
Plaid	<u>Laura Lazzeroni and Art Owen, 2000</u>	<i>Distribution Parameter Identification</i>		Coherent values
ISA	<u>Ihmels <i>et al.</i>, 2002</u>	iterative signature algorithm		Coherent values
CC / δ <i>biclusters</i>	<u>Cheng and Church, 2000</u>	<i>Greedy Iterative Search</i>		Coherent values

5.6.2.1 Binary inclusion-maximal (Bimax) Algorithm

The idea behind the Bimax algorithm, which is illustrated in Fig. 5.2, is to partition feature matrix E into three submatrices, one of which contains only 0-cells and therefore can be disregarded in the following. The algorithm is then

recursively applied to the remaining two submatrices U and V ; the recursion ends if the current matrix represents a bicluster, i.e. contains only 1s. If U and V do not share any rows and columns of E , i.e. G_W is empty; the two matrices can be processed independently from each other. However, if U and V have a set G_W of rows in common as shown in Fig. 5.2, special care is necessary to only generate those biclusters in V that share at least one common column with C_V [153].

To divide the input matrix into two smaller, possibly overlapping submatrices U and V , first the set of columns is divided into two subsets C_U and C_V , here by taking the first row as a template. Afterwards, the rows of E are resorted: first come all images that respond only to features given by C_U , then those images that respond to features in C_U and in C_V and finally the images that respond to features in C_V only. The corresponding sets of images G_U , G_W and G_V then define in combination with C_U and C_V the resulting submatrices U and V which are decomposed recursively [153].

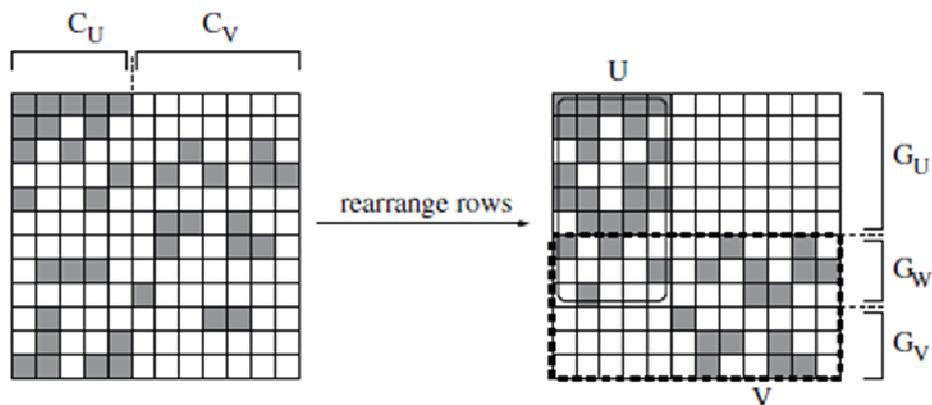


Fig. 5.2: Illustration of the Bimax algorithm [153].

The Bimax algorithm finds subgroups in a binary matrix where all entries are one. The algorithm iterates the following two steps:

1. Rearrange the rows and columns to concentrate ones in the upper right of the matrix.
2. Divide the matrix into two submatrices.

Whenever in one of the submatrices only ones are found, this submatrix is returned. In order to get satisfying results the method has to be restarted several times with different starting points.

5.6.2.2 Cheng and Church (CC) algorithm

The CC method implements the algorithm by Cheng and Church [154]. Starting from an adjusted matrix, where normalization or simple standardization preprocessing is suggested, they define a score as:

$$H(I, J) = \frac{1}{\|I\| \|J\|} \sum_{i \in I, j \in J} (a_{ij} - a_{iJ} - a_{Ij} + a_{IJ})^2, \quad (5.5)$$

where $a_{i,j}$ is the entries value of feature matrix, $a_{i,J}$ is the mean of row i , $a_{I,j}$ is the mean of column j , and $a_{I,J}$ is the overall mean. They call a subgroup a bicluster if the score is below a level α and above a δ -fraction of the whole data.

The algorithm itself has three major steps:

1. Deleting rows and columns with a score larger than *alpha* times the matrix score.
2. Deleting rows and columns with largest scores.
3. Adding Rows or Columns until *alpha* level is reached.

These steps are repeated until a maximum number of biclusters is reached or no bicluster is found. The result is constant bicluster where all $a_{i,j}$ are nearly on the same level. Choosing an appropriate preprocessing method is essential for good solutions.

5.6.2.3 Order Preserving Sub-Matrix (OPSM) algorithm

Recently, Order-Preserving Sub-Matrix (OPSM)), by Ben-Dor *et al.* [155] have been introduced and accepted as a biologically meaningful pattern based subspace cluster model. An OPSM, essentially a sub-space cluster, is a subset of rows and columns in a data matrix where all the rows induce the same linear ordering of the columns. An OPSM cluster may arise when the expression levels

of the co-regulated genes rise and fall synchronously in response to a sequence of environment stimuli. Discovery of significant OPSMs can play an essential role in inferring gene regulatory networks. The OPSM cluster model focuses on the relative order of columns rather than the uniformity of actual values in data matrix. By sorting the row vectors and replacing the entries with their corresponding column labels, the data matrix can be transformed into a sequence database, and OPSM mining is reduced to a special case of the sequential pattern mining problem with some unique properties. In particular, the sequence database is extremely dense since each column label appears exactly once (assuming no missing values) in each sequence. A sequential pattern uniquely specifies an OPSM cluster, with all the supporting sequences as the cluster contents. The number of supporting sequences is the support for the pattern.

Each OPSM represents a subset of genes identically ordered among a subset of experiment conditions in a gene Micro-array dataset. Since this problem is NP-hard, they proposed a probabilistic model to mine an OPSM from a random matrix. The local patterns found by this algorithm seem to be significant. A drawback of this algorithm is that only one cluster can be found at a time and the result is very sensitive to input parameters and initial seeds [155].

5.6.2.4 Statistical Algorithmic Method for Bicluster Analysis (SAMBA)

The SAMBA is algorithm for biclustering (Statistical Algorithmic Method for Bicluster Analysis) used to find high quality and distinct biclusters. It detects significant biclusters in a large expression dataset, using a graph theoretic approach coupled with statistical modeling of the data [142].

The SAMBA algorithm works as follows: It forms a bipartite graph according to the gene expression matrix, and calculates vertex pair weights according to a specific weighting scheme. This scheme is based on probabilistic modeling of the data. Then graph theoretic techniques are used to derive scoring schemes for identifying significant subgraphs. Each subgraph corresponds to a bicluster, and the SAMBA algorithm aims to find the k “heaviest” biclusters.

More precisely, the expression matrix is transformed into a bipartite graph $G=(U,V,E)$, where U corresponds to the conditions and V corresponds to the

genes, and $(u, v) \in E$ if v responds to a condition u . The weight of a subgraph $H=(U_I, V_I, E_I)$ is the sum of its gene-condition pairs weights, edges or non-edges.

The weight of an edge (u, v) is $\log \frac{p_c}{p_{u,v}}$, where $p_{u,v}$ is the fraction of bipartite graphs with degree sequence identical to G that contains the edge (u, v) , and p_c results from an alternative model that assumes that each edge in a true bicluster occurs with a constant probability. Similarly, the weight of every non-edge (u, v) is set to be $\log \frac{1-p_c}{1-p_{u,v}}$, and the weight of E is then:

$$\sum_{(u,v) \in E_1} \log \frac{p_c}{p_{u,v}} + \sum_{(u,v) \in (U_1 \times V_1) / E_1} \log \frac{1-p_c}{1-p_{u,v}}. \quad (5.6)$$

Under this scoring scheme, the weight of a subgraph is the log-likelihood ratio of that bicluster, so we need to discover the k heaviest subgraphs of G ., SAMBA employs a heuristic search for such subgraphs. The implementation of the algorithm is part of the EXPANDER platform [140].

5.6.2.5 Biclustering Toolboxes

Table 5.3: Biclustering toolboxes and their programming languages

Toolbox	Programming language
Expander	JAVA
ISA	MATLAB
BicAT	JAVA
Bivisu	MATLAB
Cmonkey	Java
Plaid	C
MSBE	JAVA

A. *BicAT Toolbox*

The Biclustering Analysis Toolbox (BicAT) is a software platform for clustering-based data analysis that integrates various biclustering and clustering techniques in terms of a common graphical user interface. Furthermore, BicAT provides different facilities for data preparation, inspection, and postprocessing such as discretization, filtering of biclusters according to specific criteria, or

gene pair analysis for constructing gene interconnection graphs; see Fig. 5.3. The possibility to use different biclustering algorithms inside a single graphical tool allows the user to compare clustering results and choose the algorithm that best fits a specific biological scenario. The toolbox is described in the context of gene expression analysis, but is also applicable to other types of data, e.g., data from proteomics or synthetic lethal experiments. BicAT toolbox was programmed using the Java programming language.

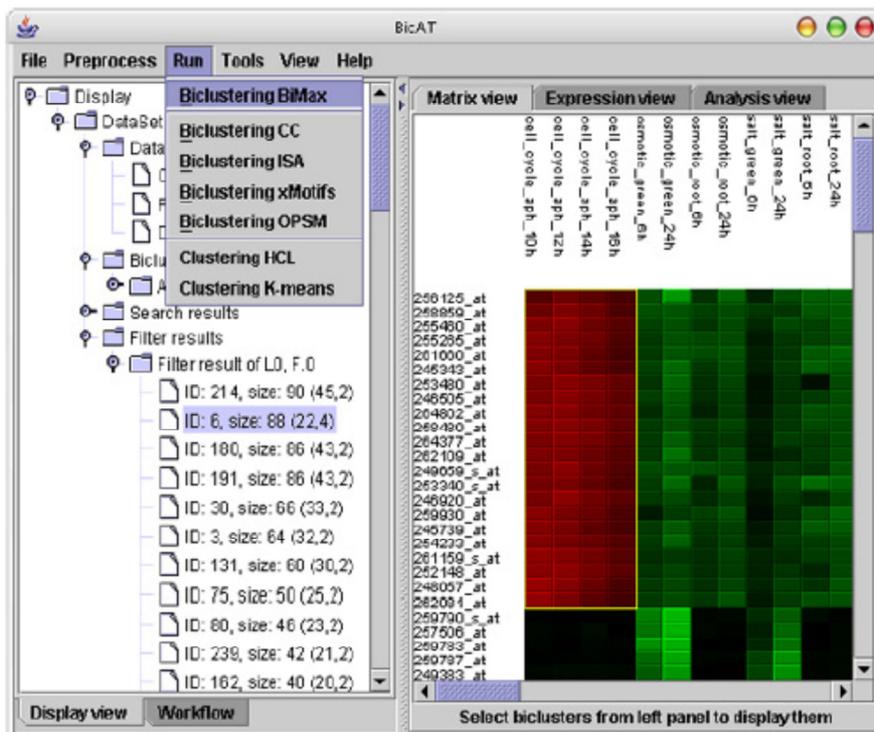


Fig. 5.3: Graphical user interface of the BicAT software.

BicAT implements the following biclustering methods: (i) Cheng and Church's algorithm (CC) which is based on a mean squared residue score; (ii) the Iterative Signature Algorithm (ISA) which searches for sub-matrices representing fix points; (iii) the Order-preserving Sub-matrix Algorithm (OPSM) which tries to identify large sub-matrices for which the induced linear order of the columns is identical for all rows; (iv) the xMotif algorithm, an iterative search method which seeks biclusters with quasi-constant expression values; (v) Bimax, an exact biclustering algorithm based on a divide-and-conquer strategy

that is capable of finding all maximal bicliques in a corresponding graph-based matrix representation. In addition, two standard clustering procedures, namely hierarchical clustering and K-means clustering, are included [141], [156].

B. EXPANDER Toolbox

EXPANDER (EXpression ANalyzer and DisplayER) is a bioinformatics software tool that was designed to help researchers in analyzing gene expression microarray (GEM) data, and allows viewing the raw data and analysis results via convenient graphical displays; see Fig. 5.4. The tool incorporates several conventional GEM analysis algorithms, and provides them with an easy-to-operate user interface. Among the tool's capabilities are clustering, biclustering, functional enrichment and promoter analysis, in addition to a variety of visualizations.

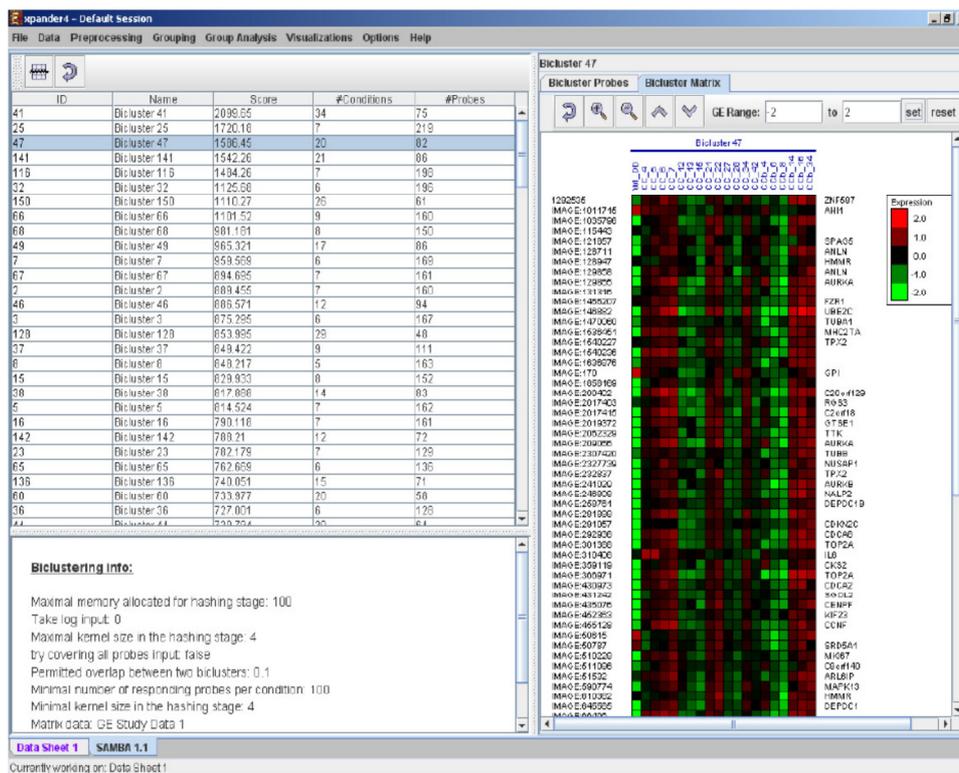


Fig. 5.4: Graphical user interface of the EXPANDER software.

EXPANDER was programmed using the Java programming language and it can be run on several platforms, including Windows and UNIX. It was written in an object oriented approach, suitable for such a large scale applications that requires many different modules that interact with one another. EXPANDER based analyses are demonstrated using three different biological datasets, and novel biological conclusions are drawn.

The EXPANDER tool is freely available for academic research, and is broadly used both for in-house research projects in biology and medicine. Over four hundred laboratories have downloaded the software over the last year. It is under ongoing development in order to keep it a state-of-the-art research tool with unique capabilities.

C. BicOverlapper Toolbox

BicOverlapper is a tool to visualize biclusters from gene-expression matrices in a way that helps to compare biclustering methods, to unravel trends and to highlight relevant genes and conditions. A visual approach can complement biological and statistical analysis and reduce the time spent by specialists interpreting the results of biclustering algorithms. The technique is based on a force-directed graph where biclusters are represented as flexible overlapped groups of genes and conditions [157]. BicOverlapper toolbox was developed in Java programming language see Fig. 5.5.

The overlap between biclusters is visualized by means of intersecting hulls, thus solving one of the most serious problems with bicluster visualization. The use of glyphs on gene and conditions nodes improves our understanding of instances of overlapping when the representation becomes complex. The effectiveness of BicOverlapper has been demonstrated using a lymphoma dataset, extracting actual biological features through the interaction with the tool without wasting time inspecting biclusters individually. Following these promising results, the tool is currently being upgraded with new linked visualizations within a visualization framework and by means of improvements in the graph layout algorithm.

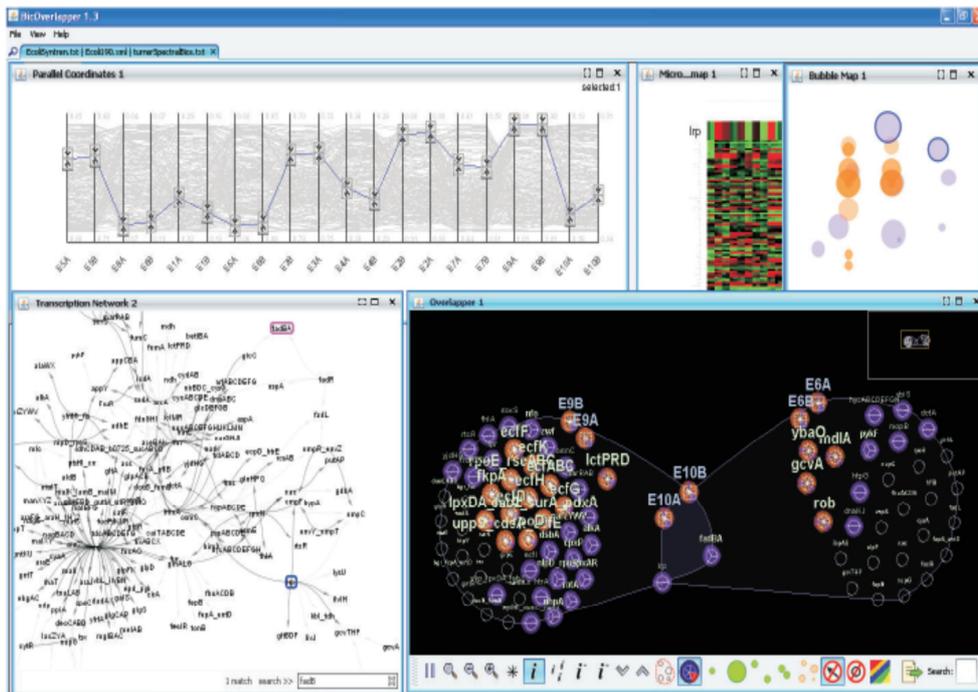


Fig. 5.5: Graphical user interface of the BicOverlapper software.

5.7 Results & Discussions

The experiments of the proposed Computer-Aided Diagnosis (CAD) system for digital mammograms using unsupervised classifiers (clustering) and biclustering techniques are conducted on the MIAS database. 224 features are extracted and 100 images for normal cases, 88 images for masses (circumscribed, spiculated, ill-defined, architectural distortion, and asymmetric) cases (51 benign images and 37 malignant images), and 25 images for microcalcification cases (13 benign images and 12 malignant images) are used.

The outcomes of the proposed system with unsupervised classifiers (clustering) and biclustering techniques are shown in following sections.

5.7.1 Unsupervised Clustering Results

Normal and abnormal (mass and microcalcification) heatmaps are illustrated in Fig. 5.6 and Fig. 5.7 respectively. From these figures the power of this visualization method becomes evident as the feature sets that show difference between the two categories can be easily identified visually.

Fig. 5.8 and Fig. 5.9 show the k-means and self-organizing map results for the normal vs. mass and normal vs. microcalcification images respectively based on the found discriminative set of features. The sensitivity and specificity for these results are shown in table 5.4 for k-means clustering, which indicates the significance performance of unsupervised clustering.

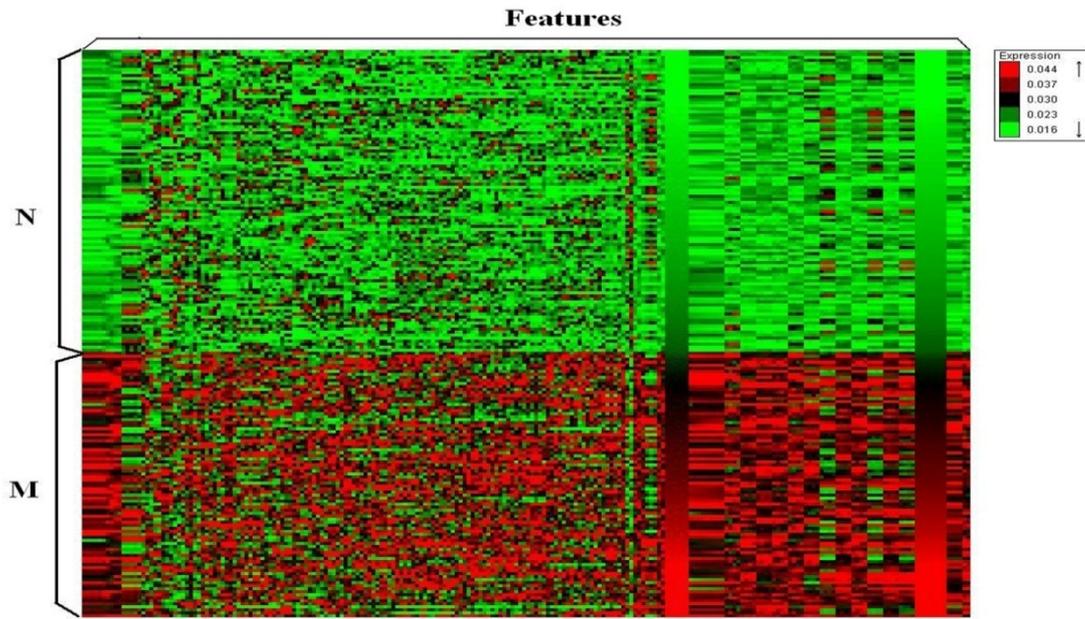


Fig. 5.6: Feature matrix heatmap for Normal(N) and mass(M) images.

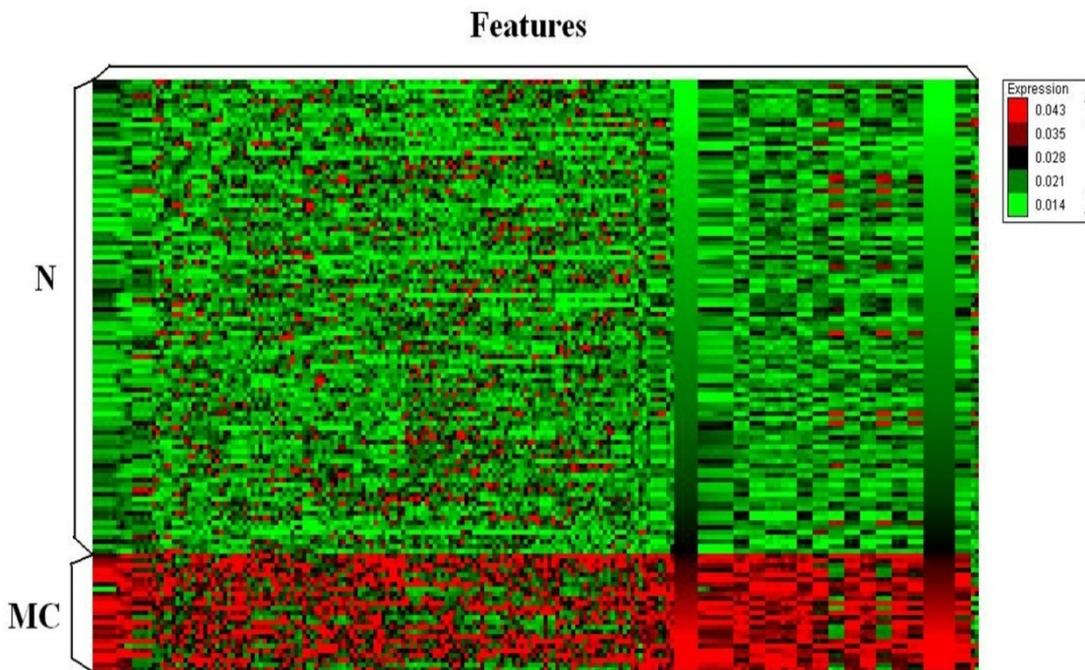


Fig. 5.7: Feature matrix heatmap for Normal(N) and Microcalcification images(MC).

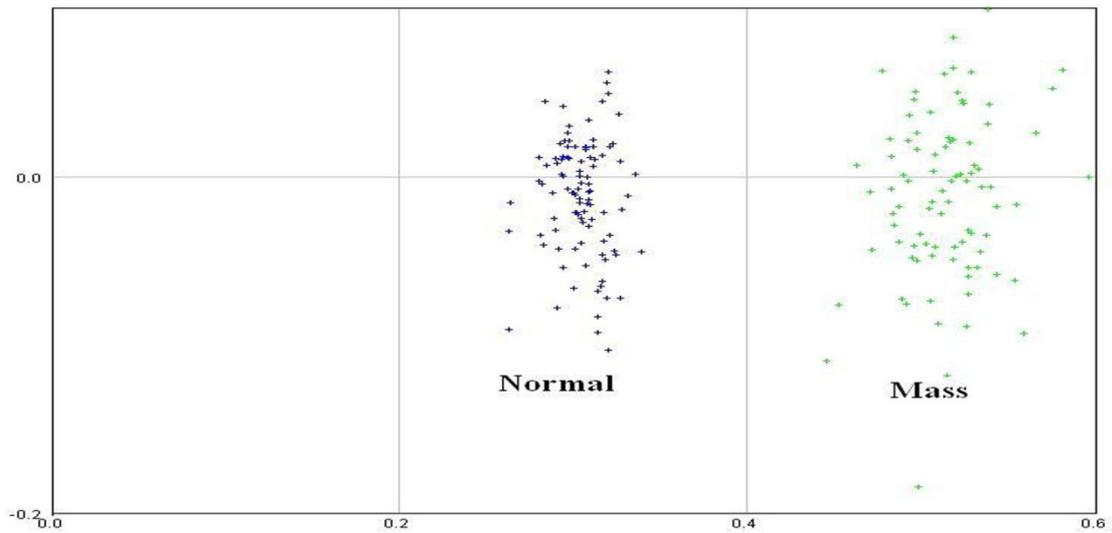


Fig. 5.8: k-means clustering results for normal and mass images.

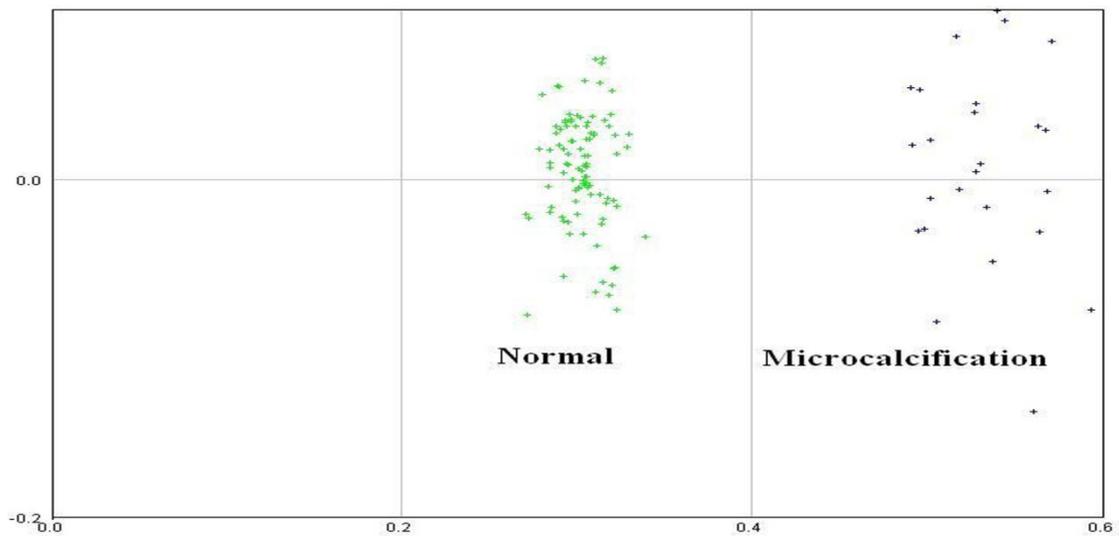


Fig. 5.9: k-means clustering results for normal and microcalcification.

Table 5.4: K-Means clustering results for normal, masses, and microcalcification

Types	K-Means Clustering	
	Sensitivity (Abnormal)	Specificity (Normal)
Normal & Masses	100%	100%
Normal & Microcalcification	100%	100%

The heatmap for feature matrix for mass (37 malignant, and 52 benign) and microcalcification (12 malignant, and 13 benign) images are shown in Fig. 5.10 and Fig. 5.11 respectively. Again, from these figures it was clearly visible that the fourteen features (6 from shape feature category and 8 from second order statistical feature category) represent the most significant features that could be used in differentiating between benign and malignant images.

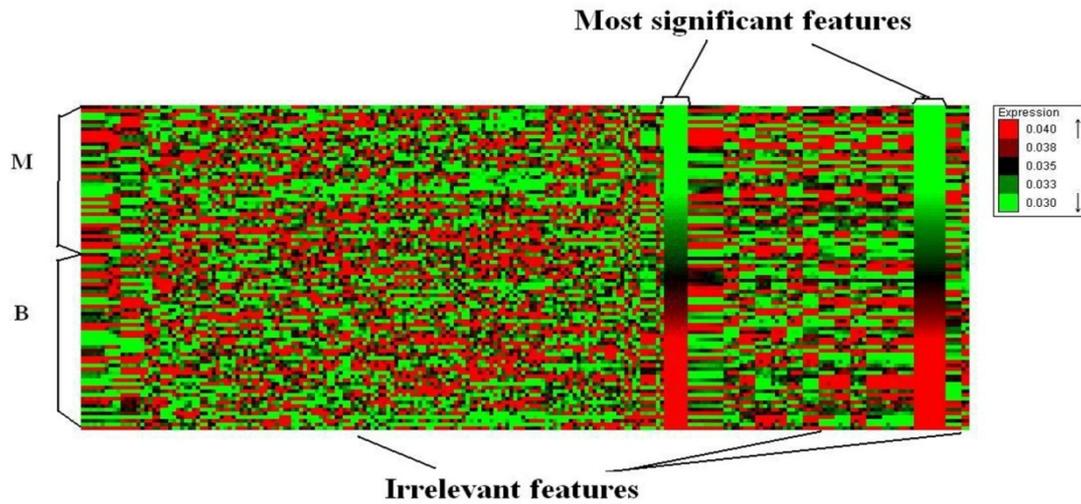


Fig. 5.10: Feature matrix heatmap for mass images (benign & malignant).

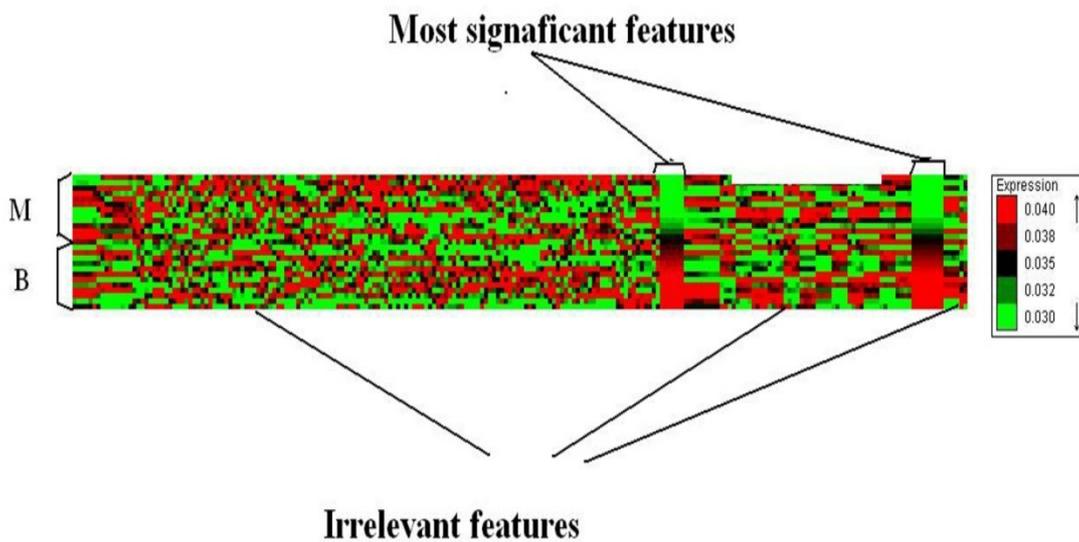


Fig. 5.11: Feature matrix heatmap for microcalcification images (benign & malignant).

Table 5.5 shows the sensitivity and specificity after applying k-means clustering using all the features compared with using the selected significant features, which indicate that using only these feature has an impact factor in classification between benign and malignant.

Table 5.5: K-Means clustering results for benign, and malignant

Types	K-Means Clustering			
	Clustering using all features		Clustering using significance features	
	Sensitivity (Malignant)	Specificity (Benign)	Sensitivity (Malignant)	Specificity (Benign)
Masses	48.6%	72.5%	100%	70.6%
Microcalcification	38.5%	76.9%	91.7%	100%

5.7.2 Biclustering techniques results

The biclustering results are obtained by applying the biclustering techniques (Bimax, CC, OPSM, and SAMBA) on the feature matrix which collected from the MIAS database to discover biclusters (pathologies) and the features are relevant to these biclusters. The feature matrix represent all features were extracted from each ROI of mammogram images; the rows of feature matrix represent images, and columns represent features.

In order to obtain the bicluster results in simplicity, the bicluster techniques were applied on four different feature matrices (mass feature matrix, microcalcification matrix, mass & microcalcification feature matrix, and normal & microcalcification feature matrix).

The results in table 5.6 represent the number of biclusters obtained by applying the biclustering techniques on the four different feature matrices.

Table 5.6: The number of biclusters was produced by biclusters techniques

Bicluster techniques	Number of biclusters			
	Mass	Micro-calcification	Mass & Micro-calcification	Normal & Micro-calcification
Bimax	5	7	10	8
CC	10	10	10	10
OPSM	17	9	15	11
SAMBA	9	1	15	16

Based on the information of the MIAS database, the biclusters which provided by SAMBA technique are the best results, i.e. We found the meaning for some of SAMBA biclusters by using the information of the MIAS database. For example; in Fig. 5.12, we draw the bicluster number 4 obtained from the SAMBA algorithm of masses images. All images belong to this bicluster are benign with overrepresentation of spiculated cancer type. Also the relevant features overrepresented with bicluster are the shape features (seven invariant moment) and second order statistics features (sum of square, difference entropy, and sum entropy).

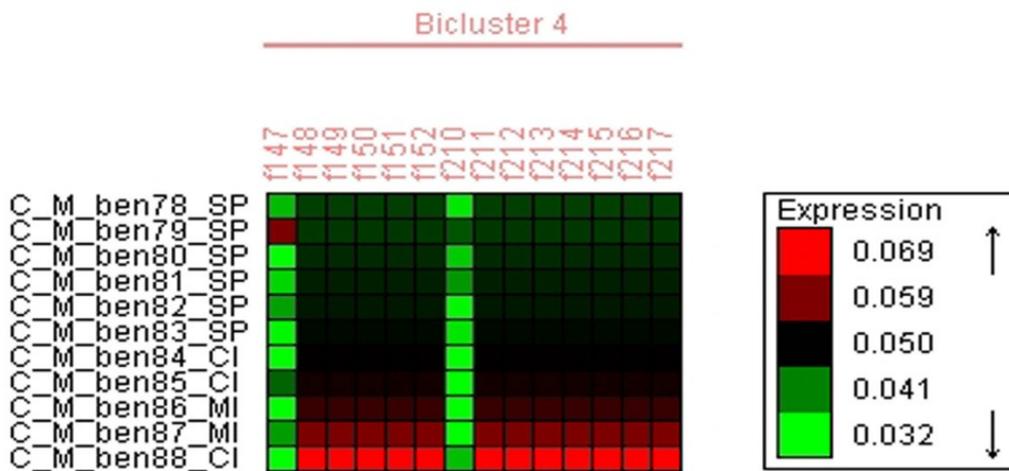


Fig. 5.12: Bicluster number 4 obtained from SAMBA algorithm of mass images where rows and columns represente images and features respectively.

In Fig. 5.13, 5.14, and 5.15 the biclusters number 2, 11, and 15 were produced by applying the SAMBA algorithm on the mass and microcalcification

images respectively. All images belong to the bicluster number 2 are microcalcifications. Also the relevant features overrepresented with bicluster are the wavelet features, first order statistics features (max. probability and entropy), and second order statistics features (contrast, correlation, energy, homogeneity, entropy of co-occurrence matrix, and second order inverse difference moment).

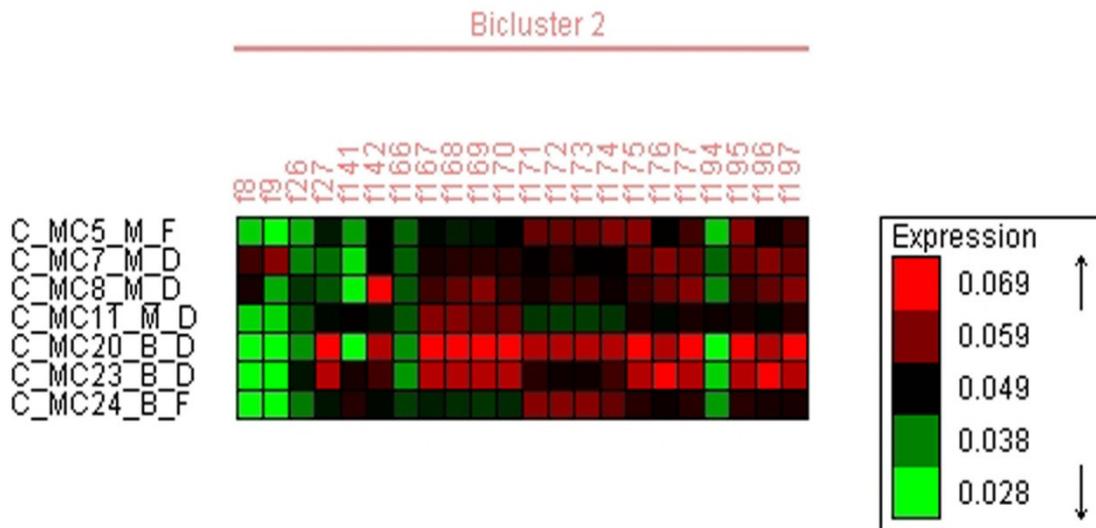


Fig. 5.13: Bicluster number 2 obtained from SAMBA algorithm of mass and microcalcification images where rows and columns represente images and features respectively.

All images belong to the bicluster number 11 are microcalcifications with overrepresentation of benign cancer type. Also the relevant features overrepresented with bicluster are the shape features (seven invariant moment), second order statistics features (sum of square, difference entropy, sum entropy, and sum average), and fractal dimension feature (piecewise modified box counting).

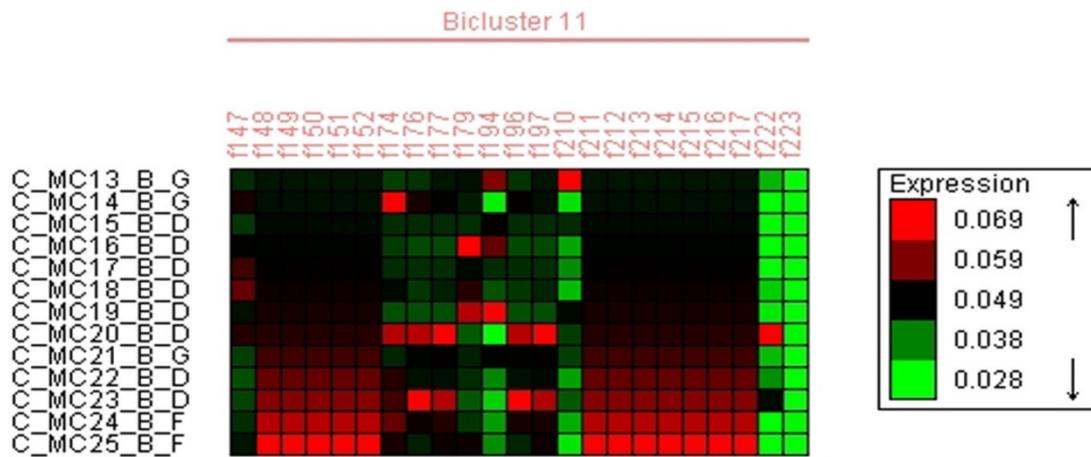


Fig. 5.14: Bicluster number 11 obtained from SAMBA algorithm of mass and microcalcification images where rows and columns represente images and features respectively.

All images belong to the bicluster number 15 are masses with overrepresentation of malignant cancer type. Also the relevant features overrepresented with bicluster are the shape features (seven invariant moment), and second order statistics features (sum of square, difference entropy, and sum entropy).

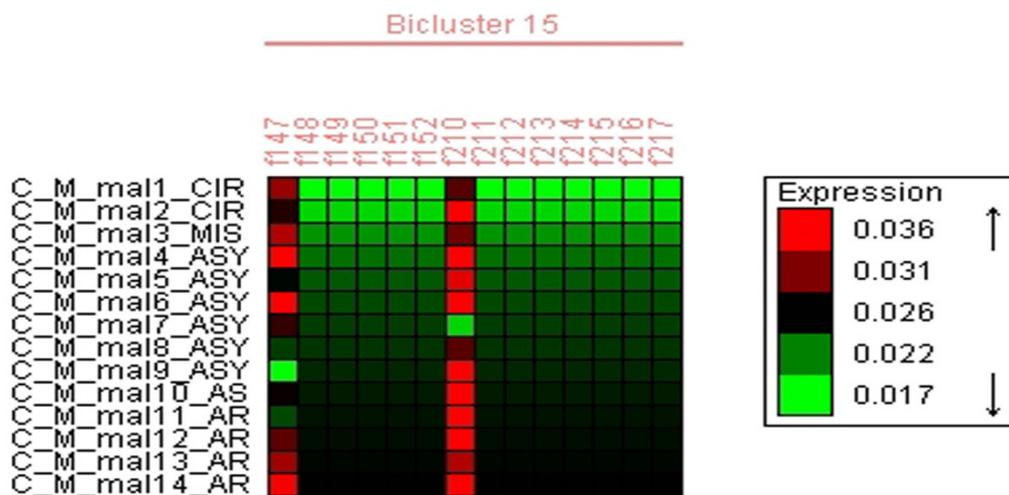


Fig. 5.15: Bicluster number 15 obtained from SAMBA algorithm of mass and microcalcification images where rows and columns represente images and features respectively.

Fig. 5.16, 5.17, and 5.18 are represented the biclusters number 1, 6, and 14 were produced by applying the SAMBA algorithm on the normal and microcalcification images respectively. All images belong to the bicluster number 1 are microcalcifications with overrepresentation of malignant cancer type. Also the relevant features overrepresented with bicluster are the wavelet features, first order statistics feature (standard deviation), shape feature (spreadness), and second order statistics features (contrast, homogeneity, first order moment, prominence, entropy of co-occurrence matrix, shade, information correlation2, and sum average).

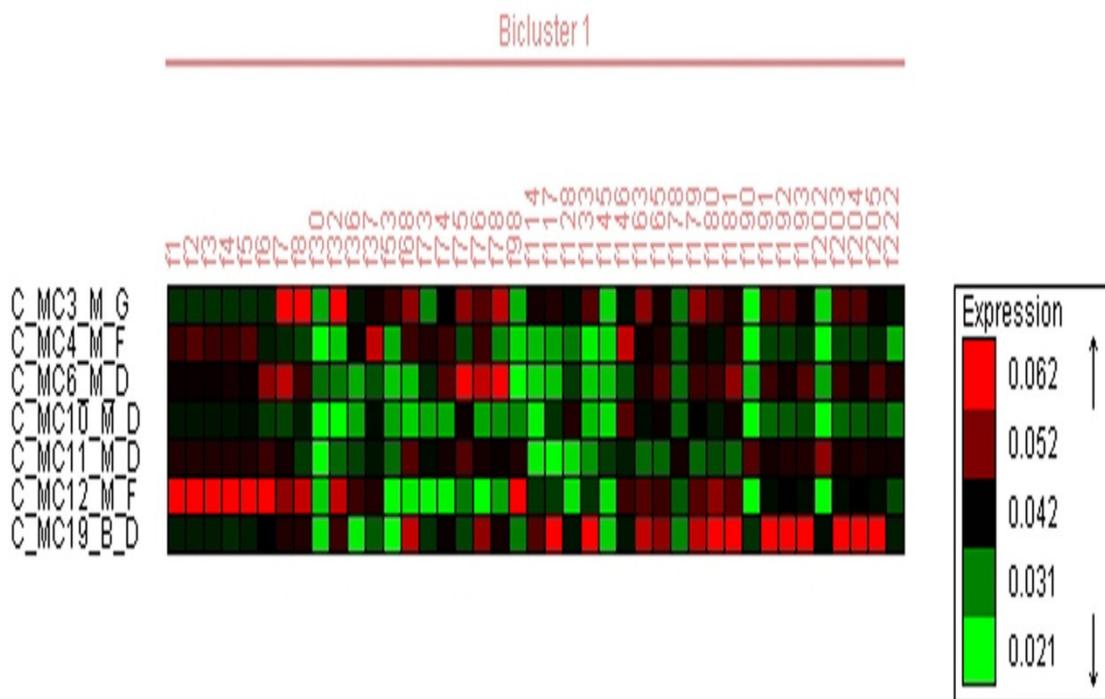


Fig. 5.16: Bicluster number 1 obtained from SAMBA algorithm of normal and microcalcification images where rows and columns represente images and features respectively.

All images belong to the bicluster number 6 are microcalcifications with overrepresentation of benign cancer type. Also the relevant features overrepresented with bicluster are the wavelet features, first order statistics features (kurtosis, standard deviation, and percentiles), shape features (seven

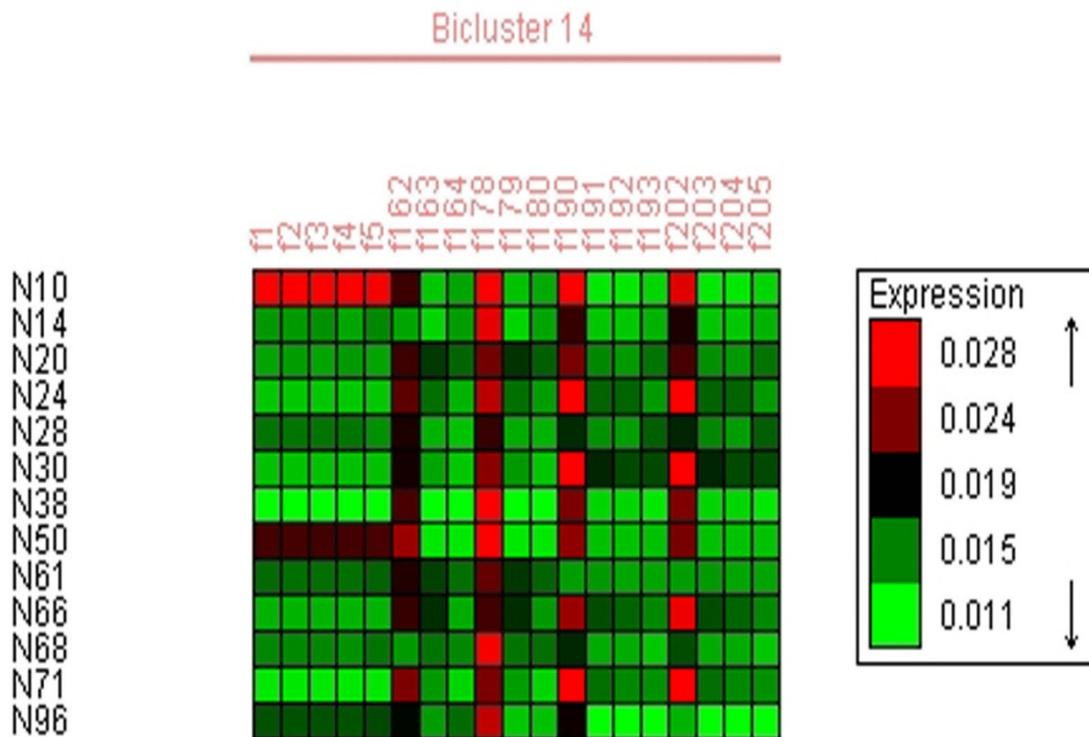


Fig. 5.18: Bicluster number 14 obtained from SAMBA algorithm of normal and microcalcification images where rows and columns represent images and features respectively.

Finally, Fig. 5.19, 5.20, 5.21, and 5.22 shows the bicluster results obtained from Bimax algorithm, CC algorithm, OPSM algorithm, and SAMBA algorithm using BicOverlapper toolbox which was applied on the mass feature matrix respectively.

In Fig. 5.19 an example of Overlapper for five biclusters from Bimax bi-clustering analysis of mass matrix; where images and features are represented by circle and square respectively. You can show or hide labels on the nodes, and search for node (image and feature) names.

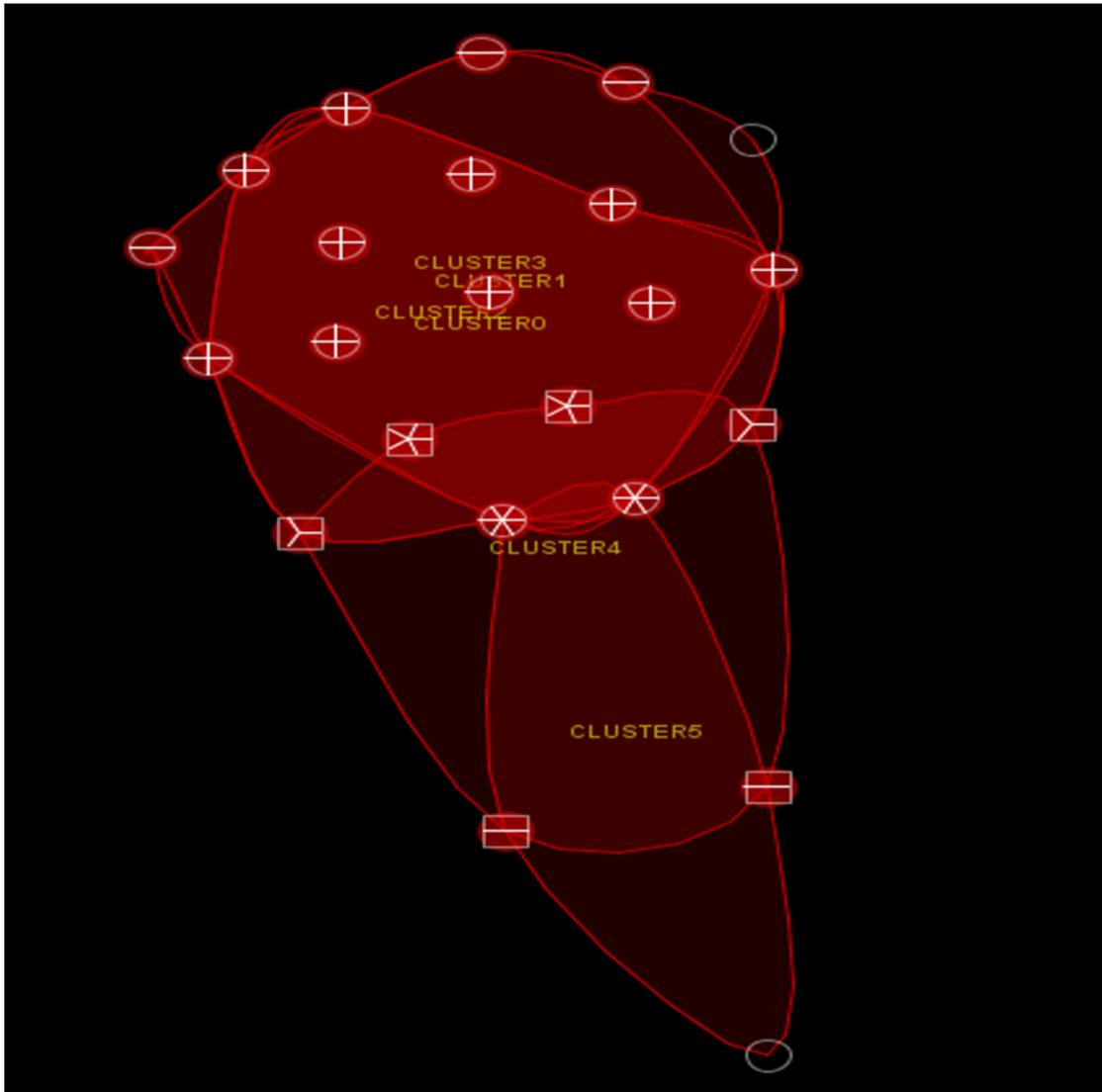


Fig. 5.19: An example of Overlapper for five biclusters from Bimax bi-clustering analysis of mass matrix; where images and features are represented by circle and square respectively.

In Fig. 5.20 an example of Overlapper for ten biclusters from CC bi-clustering analysis of mass matrix; where images and features are represented by circle and square respectively. A group of images in the center is present in all the biclusters, and a group of features without any line inside the square is present in one bicluster.

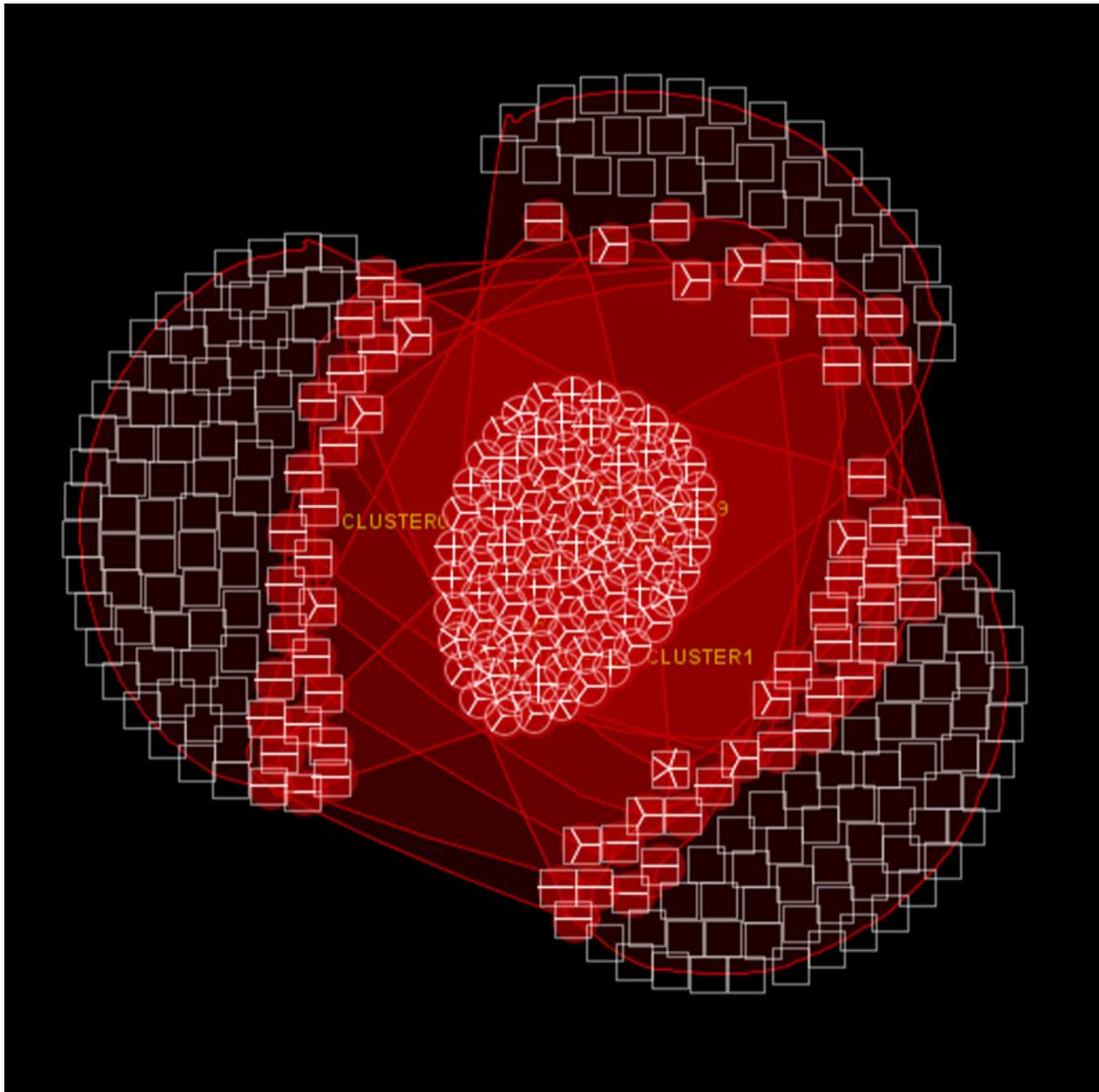


Fig. 5.20: An example of Overlapper for ten biclusters from CC bi-clustering analysis of mass matrix; where images and features are represented by circle and square respectively.

In Fig. 5.21 an example of Overlapper for seventeen biclusters from OPSM bi-clustering analysis of mass matrix; where images and features are represented by circle and square respectively. A group of images and features in the center is present in all the biclusters, and a group of features and images without any line inside the square and circle are present in one bicluster.

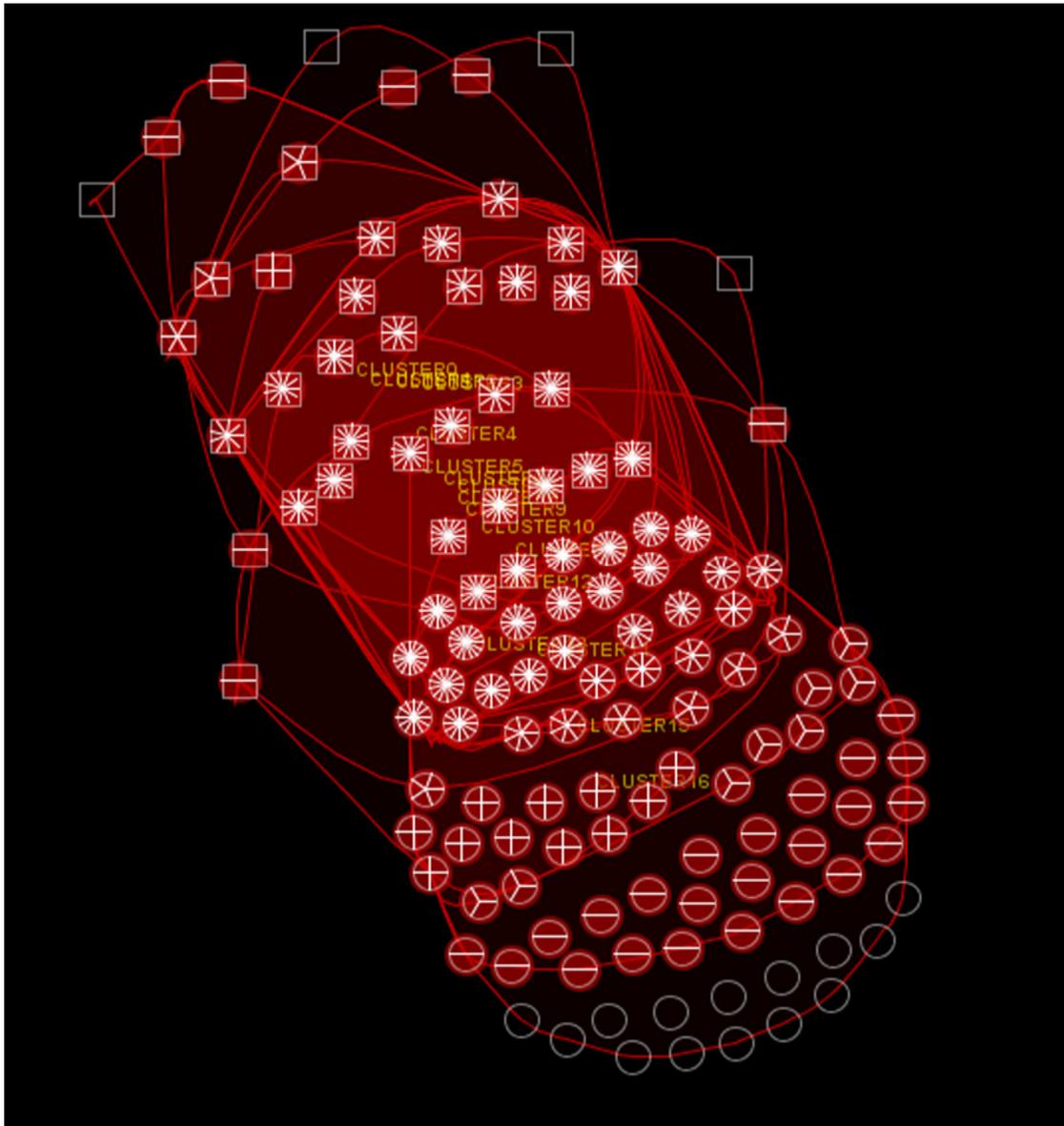


Fig. 5.21: An example of Overlapper for seventeen biclusters from OPSM bi-clustering analysis of mass matrix; where images and features are represented by circle and square respectively.

In Fig. 5.22 an example of Overlapper for nine biclusters from SAMBA bi-clustering analysis of mass matrix; where images and features are represented by circle and square respectively. A group of features and images without any line inside the square and circle are present in one bicluster.

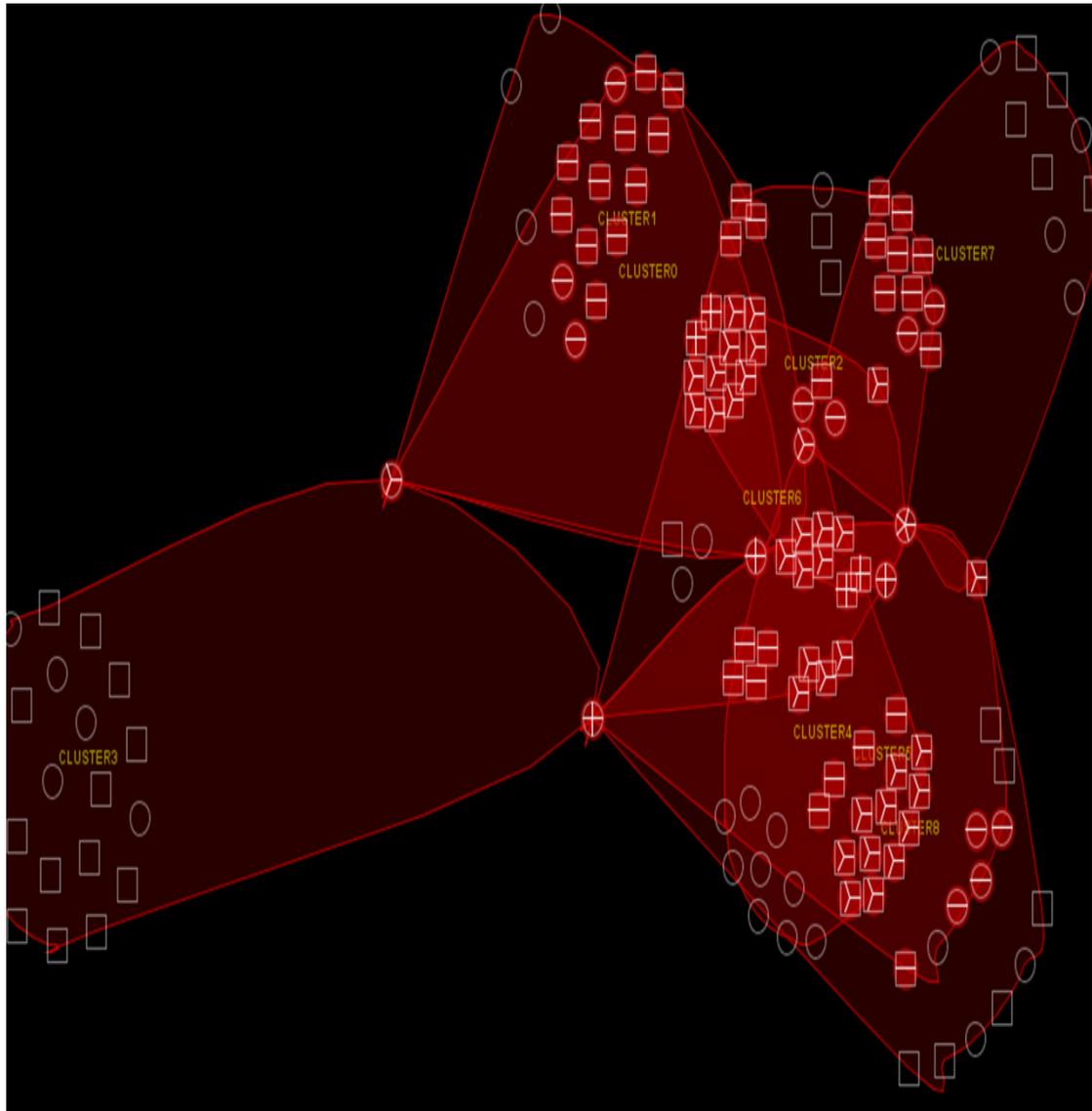


Fig. 5.22: An example of Overlapper for nine biclusters from SAMBA bi-clustering analysis of mass matrix; where images and features are represented by circle and square respectively.

The overlapping images and features corresponding to each its pathology are being under our investigation with pathologist in the future. Also for more interpretation of these results, we try to conduct the database curator to provide us with more pathology details on these images.

5.8 Summary

In this chapter; a new methodology for computer aided diagnosis in digital mammography using unsupervised classification and class-dependent feature selection is presented. This technique considers unlabeled data and provides unsupervised classes that give a better insight into classes and their interrelationships, thus improving the overall effectiveness of the diagnosis.

This technique is also extended to utilize biclustering methods, which allow for definition of unsupervised clusters of both pathologies and features. This has potential to provide more flexibility, and hence better diagnostic accuracy, than the commonly used feature selection strategies. The developed methods are applied to diagnose digital mammographic images from the MIAS database and the results confirm the potential for improving the current diagnostic rates.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

This thesis has focused on developing an approach for a Computer-Aided Diagnosis (CAD) system that can be very helpful for radiologist in detecting and diagnosing breast cancers' patterns (mass and microcalcification) in digitized mammograms earlier and faster than typical screening programs.

The new CAD system based on extracting the most effective sets of features from digital mammogram images and using them in detection and classification of breast cancer patterns by using supervised classifiers, unsupervised clustering, and biclustering methods.

A proposed system for fast fractal modeling of mammograms for microcalcifications detection is presented. The selected ROI is divided into non overlapping range blocks, these blocks are then classified into shade and non shade blocks according to their dynamic range. This system depends on mammographic microcalcification enhancement using the Collage Theorem for fractal modeling of only the non shade blocks.

All results obtained in this study are very encouraging, and indicate that the proposed fractal modeling method is an effective technique to extract mammographic patterns and to enhance microcalcifications embedded in inhomogeneous breast tissues, and this is done faster than the conventional method. Therefore, the proposed method may facilitate the radiologists' diagnosis of breast cancer at an early stage.

Another new methodology for CAD system in digital mammograms using unsupervised classification (K-means) and class-dependent feature selection is presented. This technique considers unlabeled data and provides unsupervised classes that give a better insight into classes and their interrelationships, thus improving the overall effectiveness of the diagnosis.

This technique is also extended to utilize biclustering methods (Bimax, CC, OPSM and SAMBA), which allows for definition of unsupervised clusters of

both pathologies and which features are relevant to these pathologies. The proposed system was shown to have the large potential for breast cancer diagnostic in digital mammograms and provide more flexibility, and hence better diagnostic accuracy, than the commonly used feature selection strategies.

Based on the results of this thesis, further research can be proposed in the following points:

1. Using a huge database with more representative cases. More number of images and training samples would help establish our results and observations.
2. Extracting more features from the images. In addition, non-visual features attached to the images such as age, with/without children and family history could be interesting and relevant as additional attributes for classification.
3. Developing better enhancement and segmentation algorithms.
4. Using the recent machine learning techniques such as Relevance Vector Machine (RVM).
5. Defining standard test set (database) are still very important. With some rigorous evaluations, objectives and fair comparison could determine the relative merit of competing algorithms and facilitate the development of better and robust system, this will lead to standardization of the available database.
6. Employing high resolution mammograms and wide dynamic range (gray-levels /pixel) required to adequately representing microcalcifications in digital mammograms.
7. Investigation 3D Mammograms. Mammograms are 2D of 3D structures. Then some information is inevitably distorted in the projected 2D Mammograms and will loss the depth and location of the imaged structures.
8. We would like to improve the performance of the CAD system using the information coming from the analysis of other mammographic views.

9. Using clustering and biclustering techniques with more explain and pathologies details database to investigate the overlapping images and features corresponding to each pathology, and also for more interpretation of the results of the biclustering algorithms.
10. It plain in the future to compare the algorithms implement with other difference neural network, genetic classifier as well as data mining systems.
11. Finally, although developed method is built as an offline diagnosing system, it can be rebuilt as an online diagnosing system in the future.

REFERENCES

- [1] H. D. Cheng, and M. Cui, "Mass Lesion Detection with a Fuzzy Neural Network," *Pattern Recognition*, vol. 37, no. 6, pp. 1189-1200, 2004.
- [2] H. D. Cheng, Y. M. Lui, and R. I. Freimanis, "A Novel Approach to Microcalcification Detection Using Fuzzy Logic Technique," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 442-450, June 1998.
- [3] R. Mousa, Q. Munib, and A. Moussa, "Breast Cancer Diagnosis System based on Wavelet Analysis and Fuzzy-Neural," *Expert Systems with Applications*, vol. 28, pp. 713-723, 2005.
- [4] P. M. Sampat, M. K. Markey, and A. C. Bovik, "Computer-aided detection and diagnosis in mammography," *Handbook of Image and Video Processing, 2nd ed.*, A. C. Bovik Ed. Academic Press, pp.1195-1217, 2005.
- [5] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms," *Computerized Medical Imaging and Graphics*, vol. 26, pp. 309-319, 2002.
- [6] F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," *Radiology*, vol. 89, pp. 211-5, 1967.
- [7] C. Marx, A. Malich, M. Facius, U. Grebenstein, D. Sauner, S. O. R. Pfleiderer, and W. A. Kaiser, "Are unnecessary follow-up procedures

- induced by computer-aided diagnosis (CAD) in mammography? Comparison of Mammographic diagnosis with and without use of CAD," *European Journal of Radiology*, vol. 51, pp. 66–72, 2004.
- [8] P. Sajda, C. Spence and J. Pearson, "Learning Contextual Relationships in Mammograms Using a Hierarchical Pyramid Neural Network," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 239-250, March 2002.
- [9] Songyang Yu and Ling Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Transactions Medical Imaging*, vol. 19, pp. 115-126, February 2000.
- [10] Huai Li, K. J. Ray Liu, and Shih-Chung B. Lo, "Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms," *IEEE Transactions Med. Imag.*, vol. 16, pp. 785-798, December 1999.
- [11] Kunio Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Science Direct, Computerized Medical Imaging and Graphics 31* (2007) 198–211.
- [12] O. M. Khatib and A. Modjtabai, "World Health Organization report: Guidelines for the early detection and screening of breast cancer," *EMRO Technical Publications Series 30*, 2006.
- [13] Mathers C., Loncar D., "Projections of global mortality and burden of disease from 2002 to 2030," *PLoS Medicine, Geneva*, vol. 3, pp. 442, 2006.

- [14] "Statistical Report for National Cancer Institute", *NCI 2004*, on Dec.2005
- [15] E. L. Thurfjell, K. A. Lernevall, and A. A. S. Taube., "Benefit of independent double reading in a population-based mammography screening program," *Radiology*, 191:241-244, 1994.
- [16] I. Anttinen, M. Pamilo, M. Soiva, and M. Roiha., "Double reading of mammography screening films: One radiologist or two?" *Clinical Radiology*, 48:414-421, 1993.
- [17] W. R. Hendee, C. Beam, and E. Hendrick., "Proposition: All mammograms should be double-read," *Medical Physics*, 26:115-118, 1999.
- [18] "Cancer facts and figures 2009, Technical report," *American Cancer Society*, 2009.
- [19] Shin-Yuan Hung, Chin-Yu Chen, "Mammographic case base applied for supporting image diagnosis of breast lesion," *Expert Systems with Applications* 30 (2006) 93–108.
- [20] C. Olsen, "Towards Automatic Image Analysis for Computerized Mammography," *Ph.D. thesis, Department of Computing Science, Umea University*, Sweden, 2008.
- [21] S. Feig, and M. Yaffe, "Digital Mammography, Computer-Aided Diagnosis and Telemammography," *The Radiologic Clinics of North America, Breast Imaging*, vol. 33, no. 6, pp. 1205-1230, January 1995.

- [22] B. Monsees, "Evaluation of Breast Microcalcifications," *The Radiologic Clinics of North America, Breast Imaging*, vol. 33, 6, pp.1109-1121, January 1995.
- [23] Interactive Mammography Analysis Web Tutorial, "<http://sprojects.mmi.mcgill.ca/mammography/calcifications.htm>." [Accessed: Dec. 5, 2009].
- [24] Interactive Mammography Analysis Web Tutorial, "<http://sprojects.mmi.mcgill.ca/mammography/masses.htm>." [Accessed: Dec. 5, 2009].
- [25] N. V. Candade, "Application of Support Vector Machines and Neural Networks in Digital Mammography: A Comparative Study," *M.S. thesis, University of South Florida*, 2004.
- [26] W. Phil, "Breast Masses Appropriate Evaluation," *The Radiologic Clinics of North America, Breast Imaging*, vol. 33, no. 6, pp. 1085-1108, January 1995.
- [27] T. W. Freer and M. J. Ulissey., "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology*, 220:781-786, 2001.
- [28] C. E. Metz and J. H. Shen., "Gains in accuracy from replicated readings of diagnostic images," *Medical Decision Making*, 12:60-75, 1992.

- [29] K. Doi, H. MacMahon, S. Katsuragawa, R. M. Nishikawa, and Y. Jiang. "Computer-aided diagnosis in radiology: Potential and pitfalls," *European Journal of Radiology*, 31(2):97-109, 1999.
- [30] R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles. "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology*, 219:192-202, 2001.
- [31] Linda J. Warren Burhenne, Susan A. Wood, Carl J. D'Orsi, Stephen A. Feig, Daniel B. Kopans, Kathryn F. O'Shaughnessy, Edward A. Sickles, Laszlo Tabar, Carl J. Vyborny, and Ronald A. Castellino. "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology*, 215:554-562, 2000.
- [32] "Summary of safety and effectiveness data: ISSI." *Technical Report POI0038, U.S. Food and Drug Administration*, January 15 2002.
- [33] "Summary of safety and effectiveness data: CADx Medical Systems." *Technical Report POI0034, U.S. Food and Drug Administration*, January 31 2002.
- [34] Ansgar Malich, Christiane Marx, Mirjam Facius, Thomas Boehm, Marlies Fleck, and Werner A. Kaiser. "Tumor detection rate of a new commercially available computer-aided detection system," *European Journal of Radiology*, 11:2454-2459, September 2001.
- [35] Mary Lechner, Michael Nelson, and Eugene Elvecrog. "Comparison of two commercially available computer-aided detection (CAD) systems," *Applied Radiology*, 31(4):31-35, April 2002.

- [36] B. Zheng, M. A. Ganott, C. A. Britton, C. M. Hakim, L. A. Hardesty, T. S. Chang, H. E. Rockette, and D. Gur. "Soft-copy mammographic readings with different computer-assisted detection cuing environments: Preliminary findings," *Radiology*, 221:633-640, 2001.
- [37] Rachel F. Brem and Joelle M. Schoonjans. "Radiologist detection of microcalcifications with and without computer-aided detection: A comparative study," *Clinical Radiology*, 56:150-154, 2001.
- [38] I. Leichter, S. Fields, R. Nirel, P. Bamberger, B. Novak, R. Lederman, and S. Buchbinder. "Improved mammographic interpretation of masses using computer-aided diagnosis," *European Radiology*, 10(2):377-383, 2000.
- [39] C. J. Vyborny. "Can computers help radiologists read mammograms?" *Radiology*, 191:315-317, 1994.
- [40] Digital Mammography: from theory to practice, "<http://www.moffitt.org/moffittapps/ccj/v5n1/department7.html>" [Accessed: Dec. 5, 2009].
- [41] "http://cancer.suite101.com/article.cfm/breast_cancer_imaging.htm" [Accessed: Dec. 5, 2009].
- [42] eMedicine - Breast Cancer Mammography. "<http://www.emedicine.com/radio/topic793.html>" [Accessed: Dec. 5, 2009].

- [43] eMedicine - Breast Cancer Ultrasonography. "<http://www.emedicine.com/radio/topic795.html>" [Accessed: Dec. 5, 2009].
- [44] eMedicine - Magnetic Resonance Mammography. "<http://www.emedicine.com/radio/topic792.html>" [Accessed: Dec. 5, 2009].
- [45] Hendrick RE. "Mammography legislation. In: Bassett LW, Jackson V, eds. *Diagnosis of diseases of the breast.*" Philadelphia, Pa: WB Saunders, 149–150, 1997.
- [46] US Food and Drug Administration (1998). Summary of Safety and Effectiveness Data: R2 Technologies. Technical report.
- [47] US Food and Drug Administration (2002a). Summary and Safety of Effectiveness Data: ISSI. Technical report.
- [48] US Food and Drug Administration (2002b). Summary of Safety and Effectiveness Data: CADx Medical Systems. Technical report.
- [49] Gromet M, "Comparison of Computer-Aided Detection to Double Reading of Screening Mammograms: Review of 231, 221 Mammograms" *AJR*: 190, April 2008 190:1–6.
- [50] American College of Radiology. "ACR BI-RADS - Mammography, Ultrasound & Magnetic Resonance Imaging." Reston, VA: American College of Radiology, Fourth Ed., 2003.

- [51] D. B. Kopans. "Breast Imaging." *Lippincott Williams*, January 1998.
- [52] S. Pohlman, K. Powell, N. Obuchowski, W. Chilcote, and S. Grundfest Broniatowski, "Quantitative classification of breast tumor in digitized mammograms," *Medical Physics*, vol. 23, no. 8, pp. 1337-1345, 1996. 122
- [53] D. Guliato, R Rangayyan, W. Carnelli, J. Zuffo, and J. Desautels, "Segmentation of breast tumors in mammograms by fuzzy region growing," *Proceedings of 20th Annual International Conference IEEE Engineering in Medicine and Biology*, pp. II: 1002-1004, Oct.29-Nov.1, 1998.
- [54] D. Guliato, R Rangayyan, J. Zuffo, and J. Desautels, "Detection of breast tumor boundaries using iso-intensity contours and dynamic thresholding," *Proceedings of the 4th International Workshop on Digital Mammography*, pp. 253260, Nijmegen, Netherlands, June 1996.
- [55] N. Petrick, H. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Transactions on Medical Imaging*, vol. 15, pp. 59-67, Jan. 1996.
- [56] N. Petrick, H. Chan, D. Wei, B. Sahiner, M. Helvie, and D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," *Medical Physics*, vol. 23, no. 10, pp. 16851696, 1996.
- [57] N. Mudigonda, R Rangayyan, and J. Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field

- analysis," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1215-1227, Dec. 2001.
- [58] J. Suckling, J. Parker, D. Dance, S. Astley, I. H. and C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, "The mammographic image analysis society digital mammogram database," *Proceedings of the 2nd International Workshop on Digital Mammography*, pp. 375-378, July 10-12 1994.
- [59] D. Catarious, A. Baydush, C. Abbey, and C. Floyd, "A mammographic mass cad system incorporating features from shape, fractal, and channelized hotelling observer measurements: preliminary results," *Proceedings of SPIE Medical Imaging 2003: Image Processing*, vol. 5032, pp. 111-119, 17-20 February 2003.
- [60] I. Christoyianni, E. Dermatas, and G. Kokkinakis, "Automatic detection of abnormal tissue in mammography," *Proceedings of the IEEE 2001 International Conference on Image Processing*, pp. 877 -880, October 7-10 2001.
- [61] S.-C. B. Lo, H. Li, Y. Wang, L. Kinnard, and M. Freedman, "A multiple circular path convolution neural network system for detection of mammographic masses," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 150-158, Feb. 2002.
- [62] Y. Xu, S. Neu, C. Ornes, J. Owens, J. Sklansky, and D. Valentino, "Optimization of active contour model parameters using genetic algorithms: segmentation of breast lesions in mammograms," *Proceedings of SPIE Medical Image 2002: Image Processing*, vol. 4684, pp. 1406-1414, 2002.

- [63] Y. Chu, L. Li, D. Goldgof, and R. Clark, "Classification of masses on mammograms using support vector machine," *Proceedings of SPIE Medical Imaging 2003: Image Processing*, vol. 5032, pp. 940-948, 17-20 February 2003.
- [64] X.-P. Zhang, "Multiscale tumor detection and segmentation in mammograms," *Proceedings of the IEEE 2002 International Symposium on Biomedical Imaging*, pp. 213-216, July 7-10 2002.
- [65] S. Kwok, R. Chandrasekhar, and Y. Attikiouzel, "Adaptation of the Daugman-Downing texture demodulation to highlight circumscribed mass lesions on mammograms," *Proceedings of the 14th International Conference on Digital Signal Processing*, vol. 1, pp. 449 -452, 2002.
- [66] Y. Chu, L. Li, and R. Clark, "Graph-based region growing for mass segmentation in digital mammography," *Proceedings of SPIE Medical Image 2002: Image Processing*, vol. 4684, pp. 1690-1697, 2002.
- [67] W. P. Kegelmeyer, Jr., J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper. "Computer-aided mammographic screening for speculated lesions," *Radiology*, 191:331-337, 1994.
- [68] K. Laws. "Textured image segmentation," *Ph.D. Thesis, University of Southern California*, January 1980.
- [69] K. Laws. "Rapid texture identification," *SPIE Image Processing for Missile Guidance*, 238:376-380, 1980.

- [70] N. Karssemeijer and G. M. te Brake. "Detection of stellate distortions in mammograms," *IEEE Transactions on Medical Imaging*, 15, 1996.
- [71] H. Li, Y. Wang, K. J. Liu, S. C. Lo, and M. T. Freedman. "Computerized radiographic mass detection- part i: Lesion site selection by morphological enhancement and contextual segmentation," *IEEE Transactions on Medical Imaging*, 20:289-301, 2001.
- [72] S. L. Liu, C. F. Babbs, and E. J. Delp. "Multiresolution detection of speculated lesions in digital mammograms," *IEEE Transactions on Image Processing*, 10, 2001.
- [73] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark. "Markov random field for tumor detection in digital mammography." *IEEE Transactions on Medical Imaging*, 14:565-576, 1995.
- [74] T. Matsubara, H. Fujita, T. Endo, K. Horita, M. Ikeda, C. Kido, and T. Ishigaki. "Development of mass detection algorithm based on adaptive thresholding technique in digital mammograms." *Presented at Digital Mammography*, June 1996.
- [75] N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie. "Combined adaptive enhancement and region growing segmentation of breast masses on digitized mammograms." *Medical Physics*, 26:1642-1654, 1999.
- [76] W. E. Polakowski, D. A. Cournoyer, S. K. Rogers, M. P. DeSimio, D. W. Ruck, J. W. Hoffmeister, and R. A. Raines. "Computer-aided breast cancer detection and diagnosis of masses using difference of gaussians and derivative-based feature saliency." *IEEE Transactions on Medical Imaging*, 16:811-819, 1997.

- [77] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano. "Computerized detection of malignant tumors on digital mammograms." *IEEE Transactions on Medical Imaging*, 18:369-378, 1999.
- [78] D. Brzakovic, X. M. Luo, and P. Brzakovic. "An approach to automated detection of tumors in mammograms." *IEEE Transactions on Medical Imaging*, 9:233-241, 1990.
- [79] W. Qian, L. Li, L. Clarke, R. A. Clark, and J. Thomas. "Comparison of adaptive and non adaptive cad methods for mass detection." *Academic Radiology*, 6:471-480, 1999.
- [80] W. Zhang, K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt. "An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms." *Medical Physics*, 23:595-601, 1996.
- [81] G. M. te Brake, N. Karssemeijer, and J. H. Hendriks. "An automatic method to discriminate malignant masses from normal tissue in digital mammograms." *Physics in Medicine & Biology*, 45:2843-2857, 2000.
- [82] C. Fortin et al., "Fractal Dimension in the Analysis of Medical Images," *IEEE Engineering in Medicine and Biology*, vol. 11, no. 2, pp. 65-71, June 1992.
- [83] B.B. Mandelbrot, "Fractal Geometry of Nature," *New York: W H Freeman & Co*, 1983.

- [84] C.C. Chen, J.S. Daponte, and M.D. Fox, "Fractal Feature Analysis and Classification in Medical Imaging," *IEEE Transactions on Medical Imaging*, vol. 8, no. 2, pp. 133-142, June 1989.
- [85] E.L. Chen et al., "An automatic diagnostic system for CT liver image classification," *IEEE Transactions Biomedical Engineering*, vol. 45, no. 6, pp. 783-794, June 1998.
- [86] A. I. Omara, A. S. Mohamed, A. M. Youssef, and Y. M. Kadah, "Computer Aided Diagnosis in Digital Mammography," *The Third Cairo International Biomedical Engineering Conference (CIBEC '06)*, 2006.
- [87] "<http://marathon.csee.usf.edu/Mammography/Database.html>"
[Accessed: Dec. 5, 2009].
- [88] William K. Pratt, "Digital Image Processing: PIKS Inside," *Third Edition. John Wiley & Sons, Inc.* 2001.
- [89] Rafael Gonzalez and Richard woods, "Digital image processing." Prentice Hall Inc., New Jersey, 2002.
- [90] CHAP T. LE, "Introductory Biostatistics." *John Wiley & Sons Publication. New York*, 2003.
- [91] Ruey-Feng Chang; Chii-Jen Chen; Ming-Feng Ho; Dar-Ren Chen; Woo Kyung Moon., "Breast Ultrasound Image Classification Using Fractal Analysis", *Fourth IEEE Symposium on Bioinformatics and Bioengineering, 2004*. Vol. 1, Issue, 19-21, May 2004 Page(s): 100 – 107.

- [92] C. T. Leondes, "Medical Imaging Systems Technology Analysis and Computational Methods," *World Scientific Inc., New Jersey*, 2005, pp. 63-85.
- [93] I. M. Ibrahim, A. A. Yassen, A. F. Qurany, G. E. Essam, M. A. Hefnawy, M. A. Yacoub, and Y. M. Kadah, "Computer-Aided Diagnostic System for Mass Detection in Digitized Mammograms," *Proceeding of Third Cairo International Biomedical Engineering Conference (CIBEC 06)*, pp. 1-4, 2006.
- [94] M. A. Yacoub, A. S. Mohamed, and Y. M. Kadah, "A CAD System for the Detection of Malignant Tumors in Digitized Mammogram Films," *Proceeding of Third Cairo International Biomedical Engineering Conference (CIBEC 06)*, 2006.
- [95] Wael A. Mohamed, and Y. M. Kadah, "Computer Aided Diagnosis of Digital Mammograms," *International Conference on Computer Engineering & Systems*, pp. 299-303, 27-29 Nov. 2007.
- [96] Y. M. Kadah, A. A. farag, A. M. badawy, and A. M. Youssef, "Classification algorithm for quantitative tissue characterization of diffuse liver disease from ultrasound," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, August 1996.
- [97] M. A. Alolfe, A. M. Youssef, A. S. Mohamed, and Y. M. Kadah, "Development of a computer-aided classification system for cancer detection from digital mammograms," *Proc. 25th National Radio Science Conference*, Tanta, Egypt, March 2008.
- [98] M. A. Alolfe, A. M. Youssef, A. S. Mohamed, and Y. M. Kadah, "Computer-aided diagnostic system using wavelet analysis for

- microcalcification detection in digital mammograms," *4th Cairo International Biomedical Engineering Conference (CIBEC'08)*, Cairo, 2008.
- [99] M. A. Alolfe, Wael A. Mohamed, A. M. Youssef, A. S. Mohamed, and Y. M. Kadah, "Feature Selection in Computer-aided diagnostic system for microcalcification detection in digital mammograms," *Proc. 26th National Radio Science Conference*, Future University, Egypt, March 2009.
- [100] Wael A. Mohamed, M. A. Alolfe, and Y. M. Kadah, "Microcalcification enhancement in digital mammograms using fractal modeling," *Proc. 4th Cairo International Biomedical Engineering Conf.*, Cairo, Egypt, 2008.
- [101] Wael A. Mohamed, M. A. Alolfe, Y. M. Kadah, "Fast Fractal Modeling of Mammograms for Microcalcifications Detection," *Proc. 26th National Radio Science Conference*, Cairo, Egypt, March 2009.
- [102] M. A. Alolfe, Wael A. Mohamed, A. M. Youssef, A. S. Mohamed, and Y. M. Kadah, "Computer-aided diagnosis in digital mammography using combined Support Vector Machine and Linear Discriminant Analysis," *IEEE International Conference on Image Processing*, Cairo, Egypt, 2009.
- [103] R. N. Strickland and H. I. Hahn. "Wavelet transforms for detecting microcalcifications in mammograms." *IEEE Transactions on Medical Imaging*, 15:218-229, 1996.
- [104] H. Yoshida, K. Doi, R. M. Nishikawa, M. L. Giger, and R. A. Schmidt. "An improved computer-assisted diagnostic scheme using

- wavelet transform for detecting clustered microcalcifications in digital mammograms." *Academic Radiology*, 3:621-627, 1996.
- [105] W. Zhang, H. Yoshida, R. M. Nishikawa, and K. Doi. "Optimally weighted wavelet transform based on supervised training for detection of microcalcifications in digital mammograms." *Medical Physics*, 25:949-956, 1998.
- [106] W. Qian, M. Kallergi, L. P. Clarke, H. D. Li, P. Venugopal, D. Song, and R. A. Clark. "Tree structured wavelet transform segmentation of microcalcifications in digital mammography." *Medical Physics*, 22:1247-1254, 1995.
- [107] M. N. Gurcan, Y. Yardimci, A. E. Cetin, and R. Ansari. "Detection of microcalcifications in mammograms using higher order statistics." *IEEE Signal Processing Letters*, 4:213-216, 1997.
- [108] E. J. Cands and D. L. Donoho. Ridgelets: "a key to higher-dimensional intermittency?" *Philosophical Transactions of The Royal Society, The Computer Journal*, pages 2495-2509, 1999.
- [109] T. Netsch and H. O. Peitgen. "Scale-space signatures for the detection of clustered microcalcifications in digital mammograms." *IEEE Transactions on Medical Imaging*, 18:774-786, 1999.
- [110] H. P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. MacMahon, and P. M. Jokich. "Image feature analysis and computer-aided diagnosis in digital radiography.i. automated detection of microcalcifications in mammography." *Medical Physics*, 14:538-548, 1987.

- [111] H. P. Chan, S. C. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie. "Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network." *Medical Physics*, 22:1555-1567, 1995.
- [112] D. H. Davies and D. R. Dance. "Automatic computer detection of clustered calcifications in digital mammograms." *Physics in Medicine & Biology*, 35(8):1111-1118, August 1990.
- [113] R. M. Nishikawa, M. L. Giger, R.A. Schmidt, C. J. Vyborny, W. Zhang, J. Papaioannou, U. Bick, R. Nagel, and K. Doi. "Performance of automated cad schemes for the detection and classification of clustered microcalcifications." *Presented at Digital Mammography*, Amsterdam, 1994.
- [114] B. Zheng, Y. H. Chang, M. Staiger, W. Good, and D. Gur. "Computer-aided detection of clustered microcalcifications in digitized mammograms." *Academic Radiology*, 2:655-662, 1995.
- [115] J. K. Kim and H. W. Park. "Statistical textural features for detection of microcalcifications in digitized mammograms." *IEEE Transactions on Medical Imaging*, 18:231-238, 1999.
- [116] R. H. Nagel, R. M. Nishikawa, J. Papaioannou, and K. Doi. "Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms." *Medical Physics*, 25:1502-1506, 1998.
- [117] M. F. Barnsley, "Fractals Everywhere", *Academic Press*, San Diego, CA, 1988,

- [118] A. E. Jacquin, "Image coding based on a fractal theory of Iterated Contractive Image Transformations," *IEEE Transactions Image Processing*, vol. 1, pp.18–30, Jan. 1992.
- [119] E. Jacquin, "Fractal Image Coding: A review," *IEEE Transactions on Imaging Processing*, vol. 81, pp. 1451–1465, Oct. 1993.
- [120] Y. Fisher, "Fractal Image Compression-Theory and Application", *New York: Springer-Verlag*, 1994.
- [121] B. Chaudhuri and Nintpam Sarkar, "Texture Segmentation Using Fractal Dimension" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, No.1, January 1995
- [122] Li, K. J. Ray Liu, and Shih-Chung B. Lo, "Fractal Model of Mammogram and Enhancement of Microcalcification", *IEEE Nuclear Science Symposium*, Vol. 3, pp-1850-1854, 1997
- [123] M. Haseyama, M.Takezawa, KKondo, H.Kitiajima, "An image restoration method using *IFS*", *IEEE Proc. International Conference on Image Processing* , Vol-3, pp-774-777, Sept 2000
- [124] H. Ebrahimpour-Komleh, V. Chandran, S. Sridharan, "Face recognition using fractal codes", *Proc. International Conference on Image processing*, Vol.3, pp.S8-6 1, Oct. 2001
- [125] C.M.Lai, KM.Lam, Y.H.Chan, W.C.Siu, "An Efficient Fractal Based Algorithm for Image Magnification", *Proc. of the International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp.S71-S74, Oct 2004.

- [126] L.Bocchi, G.Coppini, J.Nori,G.Valli, "Detection of Single and clustered Microcalcifications in Mammograms using Fractals models and Neural Networks," *Elsevier Medical Engineering & Physics*, pp303-312,2004
- [127] Yung-Gi, Wu, Ming Zhi, Huang, Yu-Ling, Wen, "Fractal Image compression with Variance and mean," *Proc. IEEE ICME*, vol-1, pp.353-356,2003
- [128] Sankar, T. Thomas, "Fractal Modeling of Mammograms based on Mean and Variance for the Detection of Microcalcifications," *International Conference on Computational Intelligence and Multimedia Applications*, Volume 2, Issue , 13-15 Dec. 2007 Page(s):334 - 348.
- [129] Sankar, T. Thomas, "A New Fast Fractal Modeling Approach for the Detection of Microcalcifications in Mammograms," *Journal of Digital Imaging*, Springer New York, Jul. 2009.
- [130] Sankar, T. Thomas, "Fast Fractal Coding Method for the Detection of Microcalcification in Mammograms", *International Conference on Signal Processing, Communications and Networking, 2008. ICSCN '08*.
- [131] I. El-Naqa, Y. Yang, and M Wernick, "A Support Vector Machine Approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, No. 12, December 2002.
- [132] V. N. Vapnik, "The Nature of Statistical Learning Theory," *Springer*, New York, 1995.

- [133] J. K. Kim, J. M. Park, S. S. Song, H. W. Park, "Detection of clustered microcalcifications on mammograms using surrounding region dependence method and artificial neural network," *J. VLSI Signal Processing*, pp. 251–262, vol.18, 1998.
- [134] N. Lee and H. Fujita, "K-means clustering for classifying unlabelled MRI data," *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications*, pp. 92-98, 2007.
- [135] D. Howard, S. C. Roberts, C. Ryan, and A. Brezulianu, "Textural Classification of Mammographic Parenchymal Patterns with the SONNET Self-organizing Neural Network," *Journal of Biomedicine and Biotechnology*, 2008.
- [136] D. Chen, R. Chang, C. Chen, M. Ho, S. Kuo, "Classification of breast ultrasound images using fractal feature," *Journal of Clinical Imaging*, pp. 235–245, vol. 29, 2005.
- [137] A. Meyer-Baese, O. Lange, T. Schlossbauer and A. Wismuller, "Computer-Aided Diagnosis and Visualization Based on Clustering and Independent Component Analysis for Breast MRI," *IEEE International Conference on Image Processing (ICIP2008)*, pp. 3000-3003, 2008.
- [138] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions System Man. Cybernetics*, vol. SMC-3, pp. 610–621, 1973.
- [139] R. F. Walker, P. Jackway, and I. D. Longstaff, "Improving co-occurrence matrix feature discrimination," *in Proc. 3rd Conference on*

- Digital Image Computing: Techniques and Applications (DICTA'95)*, pp. 643-648, Brisbane, Australia, 1995.
- [140] R. Shamir, A. Maron-Katz, A. Tanay, C. Linhart, I. Steinfeld, R. Sharan, Y. Shiloh, and R. Elkon, "EXPANDER - an integrative program suite for microarray data analysis," *BMC Bioinformatics*, vol. 6, p. 232, 2005.
- [141] S. Barkow, S. Bleuler, A. Prelic, P. Zimmermann, and E. Zitzler, "BicAT: a biclustering analysis toolbox," *Bioinformatics*, vol. 22, pp. 1282-1283, May 15, 2006.
- [142] A. Tanay, R. Sharan, and R. Shamir, "Discovering statistically significant biclusters in gene expression data," *Bioinformatics*, vol. 18, pp. S136-144, July 1, 2002.
- [143] "http://en.wikipedia.org/wiki/Cluster_analysis.html" [Accessed: Dec. 5, 2009].
- [144] Li, Wei, "Modified K-Means Clustering Algorithm", *Congress on Image and Signal Processing, 2008. CISP '08*, vol. 4, pp. 618-621
- [145] L. Wang, L. Bo, and L. Jiao, "A Modified K-Means Clustering with a Density-Sensitive Distance Metric," *Rough Sets and Knowledge Technology conferences (RSKT 2006)*, pp. 544-551, 2006.
- [146] Yi Lu, Shiyong Lu, Farshad Fotouhi, Youping Deng, and Susan Brown, "FGKA: A Fast Genetic K-means Algorithm," in *Proc. of the 19th ACM Symposium on Applied Computing*, pp. 162-163, Nicosia, Cyprus, March, 2004.

- [147] Yi Lu, Shiyong Lu, Farshad Fotouhi, Youping Deng, and Susan Brown, "Incremental Genetic K-means Algorithm and its Application in Gene Expression Data Analysis," *BMC Bioinformatics*, 5(172), 2004.
- [148] Binu Thomas, Raju G., and Sonam Wangmo, "A Modified Fuzzy C-Means Algorithm for Natural Data Exploration," *World Academy of Science, Engineering and Technology* 49, 2009.
- [149] Van Mechelen I, Bock HH, De Boeck P, "Two-mode clustering methods:a structured overview". *Statistical Methods in Medical Research* 13 (5): 363–94, 2004.
- [150] Kriegel, H.-P.; Kröger, P., Zimek, A., "Clustering High Dimensional Data: A Survey on Subspace Clustering, Pattern-based Clustering, and Correlation Clustering," *ACM Transactions on Knowledge Discovery from Data (TKDD)* 3 (1): 1-58, March 2009.
- [151] S. C. Madeira, and A. L. Oliveira, "Biclustering Algorithms for Biological Data Analysis: A Survey," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 1, no. 1, pp. 24-45, 2004.
- [152] Reiss DJ, Baliga NS, Bonneau R, "Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks," *BMC Bioinformatics* 2 (7): 280–302, 2006.
- [153] A. Prelic, S. Bleuler, P. Zimmermann, A. Wille, P. Hlmann, W. Gruissem, L. Hennig, L. Thiele, and E. Zitzler, "A systematic comparison and evaluation of biclustering methods for gene expression data," *Journal of Bioinformatics*, Vol. 22, no. 9, pp. 1122–1129, 2006.

- [154] Y Cheng, and G. M Church, "Biclustering of expression data," *In Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, vol. 1, pp. 93–103, 2000.
- [155] A. Ben-Dor, B. Chor, R. Karp, and Z. Yakhini, "Discovering local structure in gene expression data: The order-preserving submatrix problem," *Journal of Computational Biology*, vol. 10, pp. 373-384, 2003.
- [156] M. Al-Akwaa, Y. M. Kadah, M. H. Ali, "BicAT_Plus: An Automatic Comparative Java Tool For Bi/Clustering Algorithms Used In Analysis And Visualization of Gene Expression Data Obtained Using Microarrays," *Proc. 26th National Radio Science Conference, Future University, Egypt*, March 2009.
- [157] R. Santamaria, R. Theron, and L. Quintales, "BicOverlapper: A tool for bicluster visualization," *Journal of Bioinformatics*, vol. 24, pp. 1212 - 1213, 2008.