

# **DETECTION OF DIAGNOSTIC FEATURES FOR CLASSIFICATION OF MAMMOGRAMS**

**A DISSERTATION**

*Submitted in partial fulfillment of the  
requirements for the award of the degree*

*of*

**MASTER OF TECHNOLOGY**

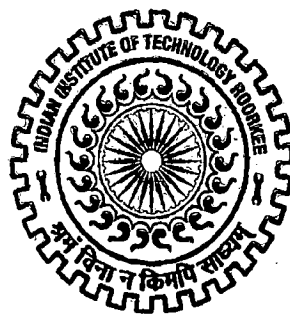
*in*

**ELECTRICAL ENGINEERING**

**(With Specialization in Measurement and Instrumentation)**

**By**

**HINGE SANDEEP JAGANNATH**



**DEPARTMENT OF ELECTRICAL ENGINEERING**

**INDIAN INSTITUTE OF TECHNOLOGY ROORKEE**

**ROORKEE - 247 667 (INDIA)**

**JUNE, 2007**

## CANDIDATE'S DECLARATION

---

I hereby declare that the work presented in this dissertation entitled "**DETECTION OF DIAGNOSTIC FEATURES FOR CLASSIFICATION OF MAMMOGRAMS**" submitted in partial fulfillment of the requirement for the award of the Degree of **Master of Technology in Electrical Engineering** with specialization in **Measurement and Instrumentation**, in the Department of Electrical Engineering, **Indian Institute of Technology, Roorkee** is an authentic record of my own work carried out from July 2006 to June 2007 under the guidance and supervision of Dr. Vinod Kumar (Professor, Electrical Engineering Department, Indian Institute of Technology, Roorkee)

I have not submitted the matter embodied in this report for the award of any other degree or diploma.

Date: June, 2007

Place: Roorkee

29-June-07

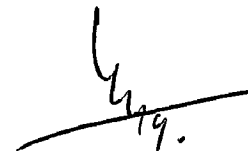


(HINGE SANDEEP JAGANNATH)

---

## CERTIFICATE

This is to certify that the above statement made by the candidate is true to the best of my knowledge and belief.



(Dr. Vinod Kumar)

Professor

Department of Electrical Engineering  
Indian Institute of Technology, Roorkee  
Roorkee-247667, Uttaranchal, INDIA .

## ACKNOWLEDGEMENTS

---

First and the foremost, I would like to thank my guide Prof. Vinod Kumar for his invaluable advices, guidance, encouragement and for sharing his broad knowledge. He has been very generous in providing the necessary resources to carry out my research. He is an inspiring teacher, a great advisor, and most importantly a nice person.

My special thanks to Lot.Col. Dr. M.R.Kural, Radiologist, M.D., CMO, IIT Roorkee Hospital & Dr. Krishna Prasad, M.D., Radiodiagnostist, PGI, Chandigrah, for their valuable help in evaluation of validity of mammogram images.

My special thanks to M & I group members for their patience during evaluations and discussions. My special thanks to Ranjit Patil, Ashwin Jain, Nageswara Reddy, Bhupendra Chandekar for their valuable suggestion and discussions(tech and non-tech). My special thanks to carrom group, Santosh Bhangare, Naresh for making me refresh for doing things in better way.

On a personal note, I owe everything to the Almighty and my parents. My father's hand written mails, long telephonic conversations with my mother and other family members provided me the mental support I needed.

(HINGE SANDEEP JAGANNATH)

## ABSTRACT

---

Mammography is currently the best technique for reliable early detection of breast cancer. Mammography is a procedure of obtaining mammogram i.e. X-Ray image of breast structures. Among all the mammogram abnormalities, microcalcifications are the most difficult type of tumour to detect. Area of thesis is mammogram image analysis for extraction of diagnostic information from it.

Thesis work proposed a scheme which focuses on enhancement of mammogram making use of morphological operators, segmentation of microcalcification using minimum cross entropy thresholding technique and extraction of number of diagnostic features for classification of mammogram into malignant or benign case.

Enhancement algorithms are developed using top-hat and h-dome filters. Both makes use of morphological operators. Top-hat filtering is dependent on selection of structuring element so is prone to be image feature dependent while h-dome extracts microcalcification irrespective of shape variation in microcalcifications present on mammogram. Traditional CLAHE enhancement algorithm is used for comparison. For the evaluation of the performance of enhancement algorithms; detail variance to background variance ratio and contrast improvement index these two performance indices are used.

After enhancement image background almost become uniform. Minimum Cross entropy based algorithm is developed for thresholding and extracting microcalcifications from the mammogram. This algorithm segments image for different thresholds and after each thresholding checks cross entropy between original image and segmented image. Threshold giving least cross entropy value is taken as final threshold. Qualitative analysis of enhancement as well as segmentation algorithms are done by presenting processed images to expert radiologist. It is found that h-dome enhancement is the best method among all stated methods for enhancement of microcalcifications without distorting the background parenchymal tissue.

Shape and texture of an image objects gives most of information about the nature of objects. Therefore we extracted 44 features providing diagnostic information based on shape and texture of image for classification.

SVM is used for classification as it is outperforming other conventional techniques as stated in literature. SVM with Radial Basis Function kernel with tuned parameters gave 82 % accuracy of classification. McGill University database images are used for quantitative and qualitative analysis with the assistance of radiologist.

# List of Figures

|      |   |    |
|------|---|----|
| 1.1  | Schematic Diagram of the Female Breast [1] . . . . .  | 11 |
| 1.2  | Mass shapes and margins [1] . . . . .   | 14 |
| 1.3  | Typically benign: a.Skin calcifications b.Vascular calcifications c.Coarse or popcorn calcifications d. Rod-shaped calcifications [1] . . . . . | 16 |
| 1.4  | Typically benign: a.Round calcifications b.Punctuate calcifications c.Spherical or lucent [1] . . . . .   | 17 |
| 1.5  | Typically benign: a. Rim or Egg calcification b.Suture calcifications c. Dystrophic calcifications [1] . . . . .                                | 17 |
| 1.6  | Intermediate concern calcifications [1] . . . . .   | 18 |
| 1.7  | Higher probability of malignancy calcifications [1] . . . . .   | 19 |
| 1.8  | Block diagram of a typical computer-aided diagnosis (CAD) system . .  | 22 |
| 2.1  | How to increase contrast of image . . . . .   | 25 |
| 2.2  | Case2.jpg from (McGill University database) a.Original image b.Cropped image . . . . .  | 26 |
| 2.3  | Tophat filtering [17] . . . . .   | 27 |
| 2.4  | Morphological reconstruction [20] . . . . .   | 29 |
| 2.5  | Block diagram of H-dome transformation mammogram enhancement .  | 29 |
| 2.6  | Synthetic images : a.Test1 b.Test2 c.Test3 d.Test4 d.Test5 . . .  | 31 |
| 2.7  | CLAHE : Test1 Original image & enhanced Image . . . . .   | 32 |
| 2.8  | CLAHE : Test2 Original image & enhanced Image . . . . .   | 32 |
| 2.9  | Tophat filtering with square SE with side=50,100 Test1 image . . . . .  | 33 |
| 2.10 | Tophat filtering with square SE with side=50,100 Test2 image . . . . .  | 33 |
| 2.11 | Tophat filtering with disk with radius 100 for Test1 image . . . . .  | 33 |
| 2.12 | Tophat filtering with disk with radius 100 for Test1 image . . . . .  | 34 |
| 2.13 | Tophat filtering with ball SE (non-flat) with height=50,radius=17, Test1 image . . . . .  | 34 |
| 2.14 | Tophat filtering with ball SE with height=50,radius=17, Test2 image .   | 35 |

|      |  |    |
|------|--|----|
| 2.15 | h-dome filtering Test1 image . . . . .   | 35 |
| 2.16 | h-dome filtering Test2 image . . . . .   | 35 |
| 2.17 | o.Cropped case25Lmlo.jpg (Mcgill database)a.CLAHE enhanced b.top-hat enhanced c.h-dome enhanced . . . . .  | 36 |
| 3.1  | Comparison of Contrast Improvement Index(CII) for different enhancement methods as applied to some ROIs of digital mammograms containing microcalcifications obtained from McGill database. . . . .                      | 41 |
| 3.2  | Comparison of Detail Variance(DV) for different enhancement methods as applied to some ROIs of digital mammograms containing microcalcifications obtained from McGill database. . . . .                                  | 43 |
| 3.3  | Comparison of Background Variance(BV) for different enhancement methods as applied to some ROIs of digital mammograms containing microcalcifications obtained from McGill database. . . . .                              | 45 |
| 3.4  | Comparison of Detail Variance(DV) to Background Variance(BV) ratio for different enhancement methods as applied to some ROIs of digital mammograms containing microcalcifications obtained from McGill database. . . . . | 47 |
| 3.5  | Mammogram ROI image:case5lcc.jpg [1] . . . . .   | 48 |
| 3.6  | Image profiles of case5lcc.jpg [1] . . . . .   | 48 |
| 4.1  | Original and segmented output image Test1 from synthetic images . . .  | 52 |
| 4.2  | Original and segmented output image Test2 from synthetic images . . .  | 53 |
| 4.3  | Original and segmented output image Test3 from synthetic images . . .  | 53 |
| 4.4  | Original Image case2.jpg(McGill Database) . . . . .  | 54 |
| 4.5  | Cropped ,segmented,superimposed output image(McGill Database) . .  | 54 |
| 4.6  | Original Image case19rcc.jpg(McGill Database) . . . . .  | 55 |
| 4.7  | Cropped, segmented output image and superimposed image case19rcc.jpg (Mcgill Database). . . . .  | 56 |
| 4.8  | Original Image case36lmlo.jpg(McGill Database) . . . . .   | 56 |
| 4.9  | Cropped, segmented output image and superimposed image case36lmlo.jpg (McGill Database). . . . .   | 57 |
| 5.1  | Co-occurrence matrix-angle $\theta$ and distance d . . . . .   | 62 |
| 5.2  | Original image: case2.jpg (McGill Database) . . . . .  | 63 |
| 5.3  | Cropped, segmented image of case2.jpg . . . . .  | 63 |
| 5.4  | Binary image matrix & labeled image matrix . . . . .   | 64 |

|     |  |    |
|-----|--|----|
| 5.5 | Labeled image matrix after object reduction . . . . .  | 64 |
| 5.6 | Schematic block diagram of feature extraction process . . . . .  | 65 |
| 6.1 | Linear separating hyperplanes for the separable case. (The support vectors are circled.) [27] . . . . .    | 75 |
| 6.2 | Linear separating hyperplanes for the non-separable case. (The support vectors are circled) [27] . . . . . | 78 |

# List of Tables

|     |  |    |
|-----|--|----|
| 3.1 | Comparison of Contrast Enhancement Index (CII) for different enhancement techniques applied on ROIs of mammograms containing microcalcifications . . . . .                     | 40 |
| 3.2 | Comparison of detail variance (DV) for different enhancement techniques applied on ROIs of mammograms containing microcalcifications. . . . .                                  | 42 |
| 3.3 | Comparison of Background variance (BV) for different enhancement techniques applied on ROIs of mammograms containing microcalcifications. . . . .                              | 44 |
| 3.4 | Comparison of Detail variance(DV) to Background variance (BV) ratio for different enhancement techniques applied on ROIs of mammograms containing microcalcifications. . . . . | 46 |
| 5.1 | Table 5.1: Sample feature calculation table - Angular Second moment based on Co-occurrence matrix with distance =1 . . . . .   | 68 |
| 5.2 | Table 5.2: Sample feature calculation table - Intensity Histogram Based features of image case17rmlo.jpg . . . . .   | 69 |
| 5.3 | Table 5.3: Sample feature calculation table - Shape based Features of image case17rmlo.jpg . . . . .   | 70 |
| 6.1 | List of extracted features . . . . .   | 81 |



# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>10</b> |
| 1.1      | Introduction . . . . .  | 10        |
| 1.2      | Basics of mammography . . . . .                                       | 10        |
| 1.2.1    | Anatomy of female breast . . . . .                                    | 10        |
| 1.2.2    | Symptoms of breast cancer . . . . .                                   | 11        |
| 1.2.3    | Mammography . . . . .   | 12        |
| 1.2.4    | Definitions & terms associated with mammography . . . . .             | 12        |
| 1.2.5    | Mammogram abnormalities . . . . .                                     | 13        |
| 1.2.6    | ACR BIRADS Classification of calcifications . . . . .                 | 15        |
| 1.2.7    | Problems in interpretation of microcalcifications on mammograms<br>18 |           |
| 1.3      | Motivation for the thesis . . . . .                                   | 19        |
| 1.4      | Literature survey . . . . .   | 20        |
| 1.5      | Area of thesis . . . . .  | 22        |
| 1.6      | Organization of thesis . . . . .                                      | 23        |
| <b>2</b> | <b>Mammogram enhancement</b>  | <b>24</b> |
| 2.1      | Introduction . . . . .  | 24        |
| 2.2      | How to increase contrast of image . . . . .                           | 24        |
| 2.3      | Preprocessing of mammogram . . . . .                                  | 25        |
| 2.4      | Mammogram enhancement techniques . . . . .                            | 25        |
| 2.5      | Results and discussions . . . . .                                     | 30        |
| 2.5.1    | Synthetic images . . . . .  | 30        |
| 2.5.2    | Image Database . . . . .  | 31        |
| 2.5.3    | Results . . . . .   | 32        |
| <b>3</b> | <b>Quantitative analysis of mammogram enhancement algorithms</b>      | <b>37</b> |
| 3.1      | Introduction . . . . .  | 37        |

|          |   |           |
|----------|---|-----------|
| 3.2      | Evaluation Indices . . . . .                    | 37        |
| 3.2.1    | Contrast Improvement Index (CII) . . . . .      | 38        |
| 3.2.2    | Background and detail region variance . . . . . | 38        |
| 3.2.3    | Results and discussions . . . . .               | 39        |
| 3.3      | Image Profiles . . . . .                        | 48        |
| <b>4</b> | <b>Detection of microcalcifications</b>         | <b>50</b> |
| 4.1      | Introduction . . . . .                          | 50        |
| 4.2      | Basics of entropy of an image . . . . .         | 50        |
| 4.3      | Algorithm . . . . .                             | 51        |
| 4.3.1    | Results and discussions . . . . .               | 52        |
| <b>5</b> | <b>Diagnostic features extraction</b>           | <b>58</b> |
| 5.1      | Introduction . . . . .                          | 58        |
| 5.2      | Labelling & object reduction . . . . .          | 58        |
| 5.3      | Feature Extraction . . . . .                    | 59        |
| 5.3.1    | Shape based features . . . . .                  | 59        |
| 5.3.2    | Texture based features . . . . .                | 60        |
| 5.3.3    | Results and discussions . . . . .               | 63        |
| <b>6</b> | <b>Classification of microcalcifications</b>    | <b>72</b> |
| 6.1      | Introduction . . . . .                          | 72        |
| 6.2      | Support Vector Machines (SVMs) . . . . .        | 72        |
| 6.2.1    | Linear SVM . . . . .                            | 74        |
| 6.2.2    | Lagrangian Formulation of the Problem . . . . . | 76        |
| 6.2.3    | Non-Separable Case . . . . .                    | 77        |
| 6.2.4    | Nonlinear SVM Classifier . . . . .              | 78        |
| 6.2.5    | SVM Kernel Functions . . . . .                  | 79        |
| 6.3      | Results and discussions . . . . .               | 80        |
| 6.3.1    | Input data . . . . .                            | 80        |
| 6.3.2    | Feature selection . . . . .                     | 80        |
| 6.3.3    | Parameter Selection and Training . . . . .      | 80        |
| <b>7</b> | <b>Conclusions and scope of future work</b>     | <b>82</b> |
| 7.1      | Conclusions . . . . .                           | 82        |
| 7.2      | Scope of future work . . . . .                  | 83        |
|          | Bibliography . . . . .                          | 84        |

# Chapter 1

## Introduction

### 1.1 Introduction

This chapter gives brief introduction of basics of mammography, it also highlights gaps in the current research and states aims of this thesis.

The brief overview of mammography which is best technique for early detection of breast cancer is cited. Important definitions and terminology used in mammography are described. Later an attempt is made to give an idea to reader about the intricacies involved in interpretation of mammograms. Mammogram abnormalities are discussed in detail and where this thesis work contributes in improvement of interpretation of mammogram is cited.

### 1.2 Basics of mammography

Mammography is acknowledged as a single most effective method for screening for breast cancer. Mammography is a technique that uses X-rays to provide an image of the breast structures. These images are known as mammograms which are used to find potential signs of breast cancer like tumours, small clusters of calcium deposits and abnormal changes in the skin.

#### 1.2.1 Anatomy of female breast

The anatomy of the adult female breast consists of 12 - 20 conical lobes. The base of a lobe lies on top of the pectoral muscles and ribs, and its apex is at the areola and nipple. Lobular (glandular) and ductal tissue lie within each lobe supported by intralobular connective tissue and adipose tissue. There is also extralobular connective

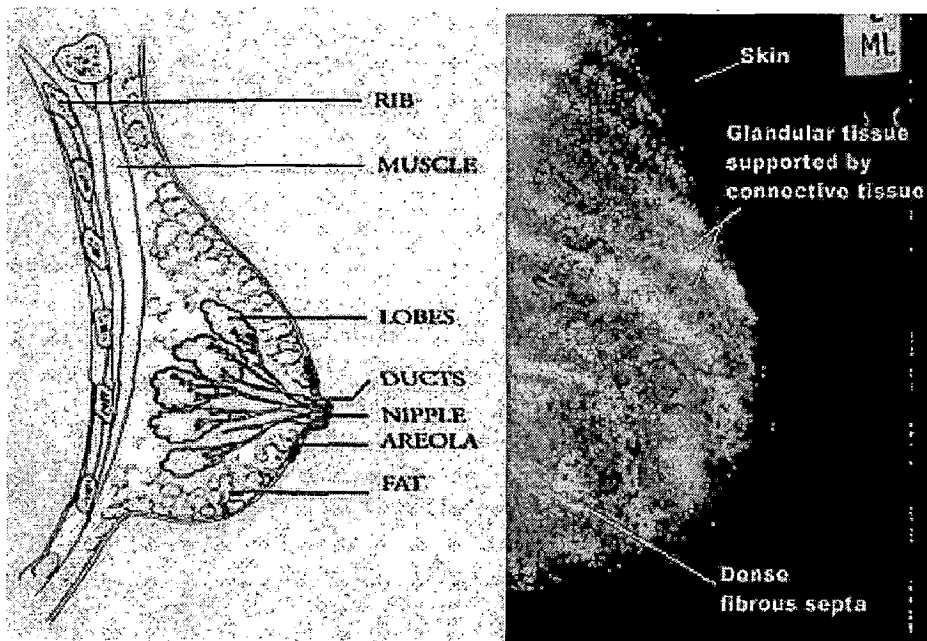


Figure 1.1: Schematic Diagram of the Female Breast [1]

tissue which binds the lobes together as well as extralobular adipose tissue. Adipose tissue is radiolucent, and the radiographically visible densities on a mammogram are the images of lobular elements, ducts, and fibrous connective tissue. Ducts may be seen as thin linear structures emanating from the nipple. Lobules and their ducts are often superimposed with connective tissue structures. Whether the mammographic appearance of a breast will appear more or less radiolucent will depend for the most part on the quantity of extralobular connective tissue [1].

### 1.2.2 Symptoms of breast cancer

Breast cancer in early stages usually, does not cause pain. In fact, when breast cancer first develops, there may be no symptoms at all. But as the cancer grows, the following symptoms generally occur [2]

- A lump or thickening in or near the breast or in the underarm area.
- A change in the size or shape of the breast.
- Nipple discharge or tenderness, or the nipple pulled back (inverted) into the breast. Ridges or pitting of the breast (the skin looks like the skin of an orange).

- A change in the way the skin of the breast, areola, or nipple looks or feels (for example, warm, swollen, red, or scaly).

### 1.2.3 Mammography

It is a technique, used to visualize normal and abnormal structures within the breasts. Mammography, therefore, can help in identifying cysts, calcifications, and tumors within the breast.

Mammography is currently the most effective way to detect early breast cancer. Breast self-examination (BSE) on a monthly basis and examination by a doctor are still important, but physical examinations typically find breast cancers when they are much larger than those detected by mammography. While mammography can find cancers at early stage, when they are small and most responsive to treatment. Breast cancer is detected on the basis of four types of signs observed on the mammograms:

1. The characteristic morphology of a tumor mass.
2. Certain presentations of mineral deposits as specks called microcalcifications.
3. Architectural distortion of normal tissue patterns caused by disease.
4. Asymmetry between corresponding regions of images of the left and the right breast.

### 1.2.4 Definitions & terms associated with mammography

**Mammogram**- X-ray image of breast structures.

**Calcifications**- These are the small tiny calcium deposits on breast tissue which looks as bright spots on mammogram image.

**Microcalcifications** - Calcifications of size 0.5 mm or less are known to be microcalcifications as an agreement by most of radiologist.

**Macrocalcifications** - Calcifications of size 2.0 mm or greater are known to be macrocalcifications as an agreement by most of radiologist.

**Suspicious or malignant** - A breast abnormality that may indicate breast cancer. On a mammogram, these abnormalities may be lesions such as spiculated masses or pleomorphic microcalcifications.

**Benign** - A breast abnormality that is not an indication of breast cancer.

### 1.2.5 Mammogram abnormalities

A radiologist looks for certain signs and characteristics indicative of cancer when evaluating a mammogram. Masses and calcifications are most common abnormalities on mammograms. A mass is a space-occupying lesion seen in at least two mammographic projections. A calcification is deposit of calcium salt on tissue. Both can be associated with either benign or malignant abnormalities, and can have a variety of appearances. The American College of Radiology (ACR) formulated Breast Imaging Reporting and Data System (BI-RADS), which contains a guide to standardized mammographic reporting, including a breast-imaging lexicon of terminology, a report organization and assessment structure and coding system, to help in standardized reporting. The lexicon used for describing mammographic abnormalities is organized by mass and calcifications. Masses are described by their geometry, border characteristics, and density. Calcifications are described by their size morphology and distribution. The ACR-BIRADS has classified findings of calcifications into three categories: (1) Typically benign; (2) Intermediate concern; and (3) Higher probability of malignancy.

#### Masses

Masses are three-dimensional lesions which may represent a localizing sign of breast cancer. They are described by their location, size, shape, margin characteristics, x-ray attenuation (radiodensity), effect on surrounding tissue, and any other associated findings (i.e. architectural distortion, associated calcifications, skin changes). Depending on the morphologic criteria of the mass, the likelihood of malignancy can be established.

*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 below. Architectural distortion is not technically a mass since there is no definite mass visible. It can be identified by dis-

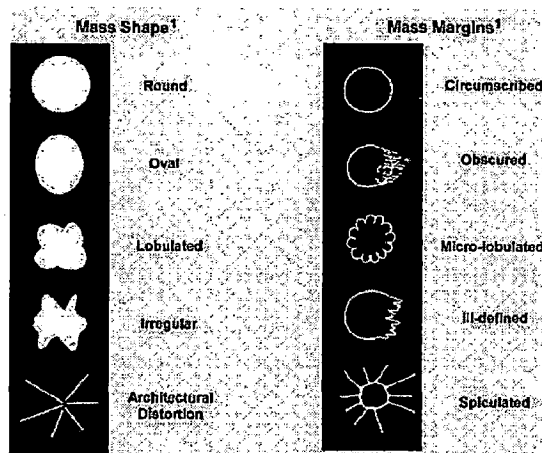


Figure 1.2: Mass shapes and margins [1]

tortion 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.

*Margins* - 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 type of margins as defined by BIRADS: Circumscribed, Obscured, Micro-lobulated, Ill-defined, and Spiculated. Circumscribed margins are well defined and sharply demarcated with an abrupt transition between the lesion and the surrounding tissue. Microlobulated 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.

## Calcifications

Calcifications are often important and common findings on a mammogram.. They can appear with or without an associated lesion, and their morphologies and distribution provide clues as to their etiology as well as whether they can be associated with a benign or malignant process.

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.

*Size* - Generally speaking, microcalcifications are associated with a malignant process and macrocalcifications are associated with a benign process. 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.

*Number* - The number of calcifications that make up a cluster has been used as an indicator of benign 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.

*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.

## 1.2.6 ACR BIRADS Classification of calcifications

The American College of Radiology (ACR) Breast Imaging Reporting and Data System (BIRADS) has classified findings of calcifications into three categories [1]:

Typically benign

Intermediate concern

Higher probability of malignancy

### Typically benign

*Skin Calcifications* - Skin calcifications have a typical lucent center and polygonal shape.

*Vascular Calcifications* - Vascular calcifications can be seen as parallel tracks or linear tubular calcifications that run along a blood vessel.

*Coarse or Popcorn-like Calcifications*

*Rod-Shaped Calcifications* - Large rod-like calcifications are typical of secretory disease but not of breast cancer. They are usually >1mm, are occasionally branching, and may have lucent centers.

*Round Calcifications* - Smooth round calcifications are associated with a benign process. They may vary in size in a cluster.



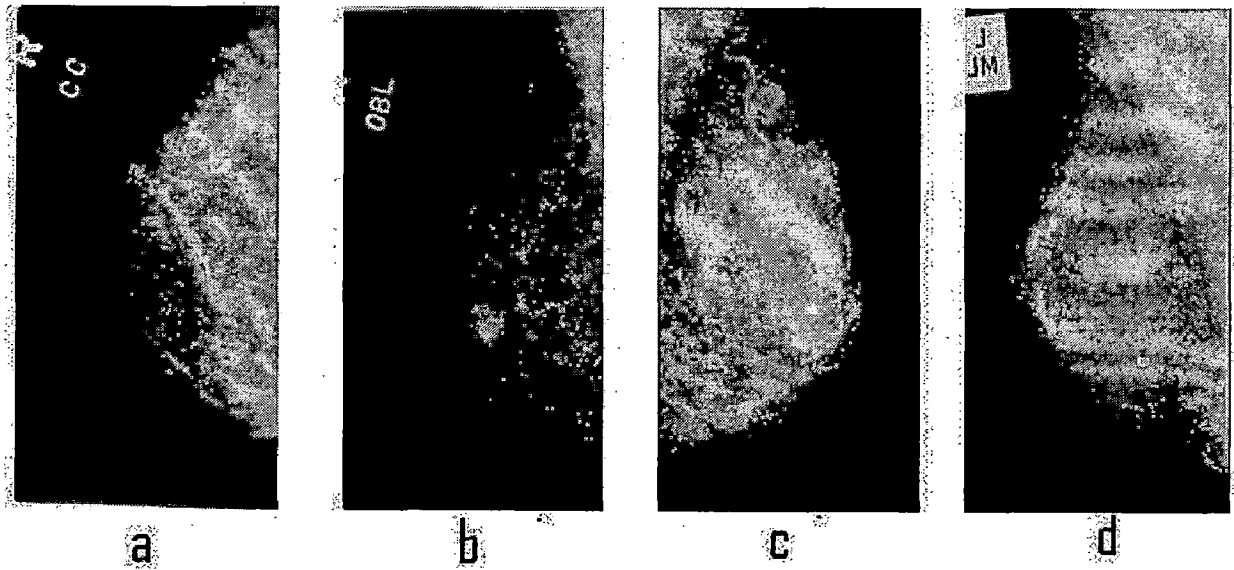


Figure 1.3: Typically benign: a.Skin calcifications b.Vascular calcifications c.Coarse or popcorn calcifications d. Rod-shaped calcifications [1]

*Punctuate Calcifications* - Round or oval calcifications <0.5mm are called punctuate. They appear as sharply defined, pinpoint deposits. Punctuate calcifications are rarely associated with cancer, but if they are found together with other heterogeneously shaped calcifications, greater suspicion is warranted.

*Spherical or Lucent - Centered Calcifications* Spherical or lucent-centered calcifications can range from <1mm to >1cm.

*Rim or Egg - shell Calcifications* These are thin calcifications that surround all or part of the margin of a mass.

*Suture Calcifications* - They typically appear as linear or tubular and knots may be visible.

*Dystrophic Calcifications* - These calcifications are irregular in shape but they are usually large, i.e. >0.5mm in size.

### **Intermediate concern calcification**

Calcifications that are of immediate concern are indistinct or amorphous microcalcifications. They may appear as round or flake-shaped calcifications that are sufficiently small or hazy such that their morphology cannot be reliably ascertained.

### **Higher probability of malignancy**

*Pleomorphic or Heterogeneous Calcifications* - Heterogeneous or pleomorphic calcifications in and of themselves are not associated with a benign or malignant process.

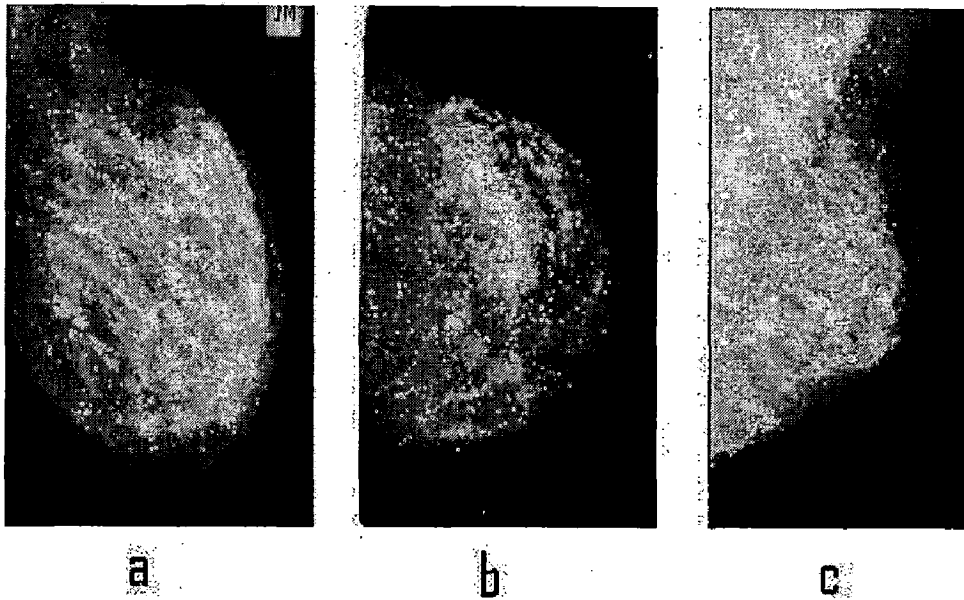


Figure 1.4: Typically benign: a.Round calcifications b.Punctuate calcifications c.Spherical or lucent [1]

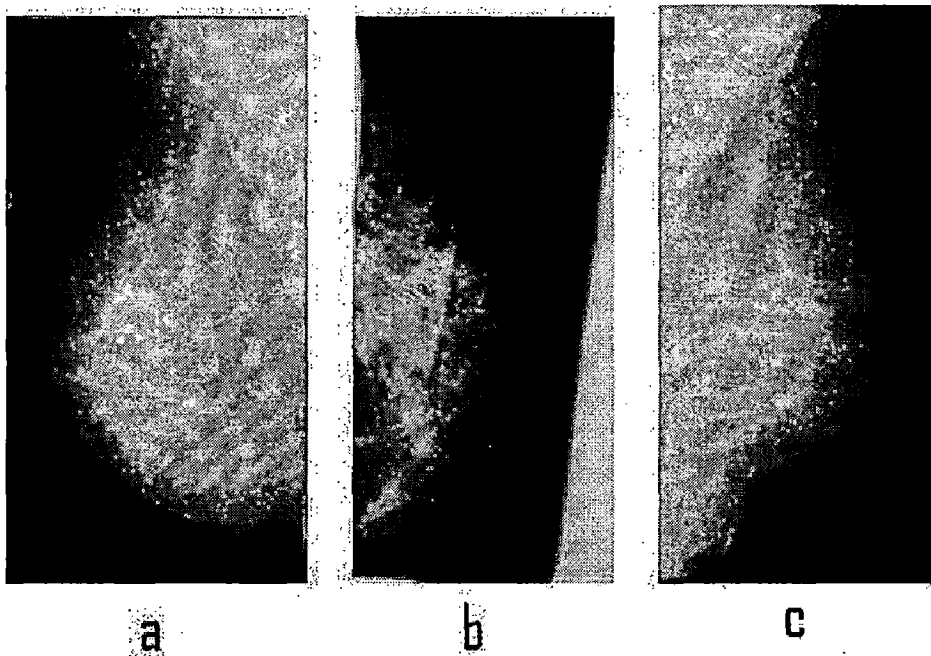


Figure 1.5: Typically benign: a. Rim or Egg calcification b.Suture calcifications c. Dystrophic calcifications [1]

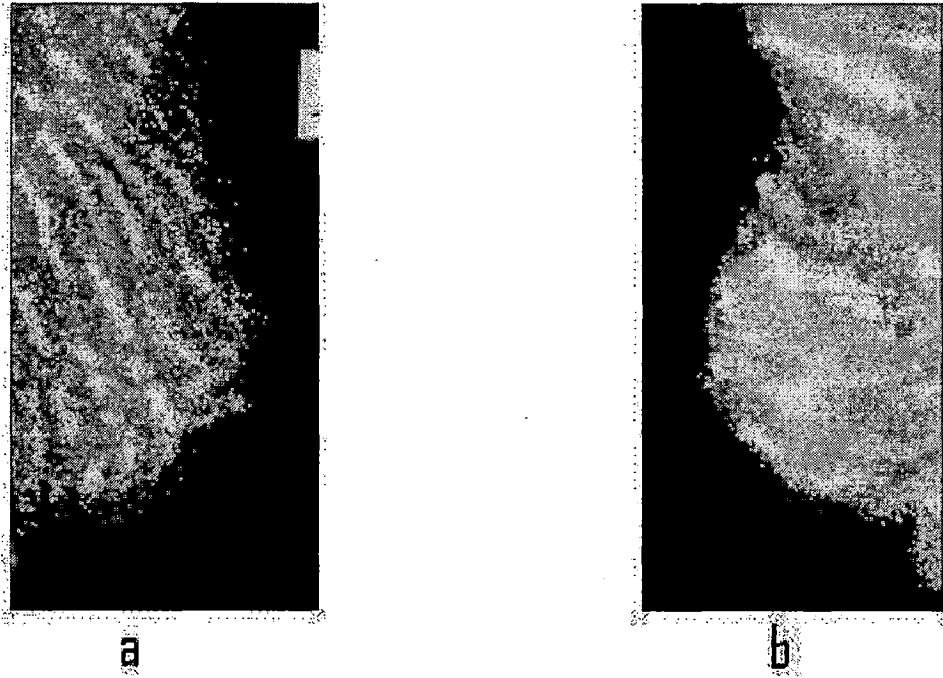


Figure 1.6: Intermediate concern calcifications [1]

However, a cluster of calcifications irregular in shape, size, and tend to be  $<0.5\text{mm}$  raises suspicion.

*Fine Linear or Branching Calcifications* - These are thin, irregular calcifications that appear linear from a distance. Closer examination reveals that they are distinct and  $<1\text{mm}$  in width.

### 1.2.7 Problems in interpretation of microcalcifications on mammograms

Microcalcifications are considered to be important signs of breast cancer. It has been reported that 30-50 % of breast cancer detected radiographically show microcalcifications on mammograms. The high correlation between the presence of microcalcifications and the presence of the breast cancer suggest that the accurate detection of microcalcification will improve the efficacy of mammography as a diagnostic feature.

1. Microcalcifications are very small. According to literature , the sizes of microcalcifications vary from  $0.1\text{mm}$  to  $1\text{mm}$  with an average diameter of about  $0.3\text{mm}$ . Small microcalcifications, ranging  $0.1\text{mm}$  to  $0.2\text{mm}$ , can hardly be seen on image due to their superimposition on the breast parenchymal textures and noise.
2. Microcalcifications often appear in an inhomogeneous background describing the

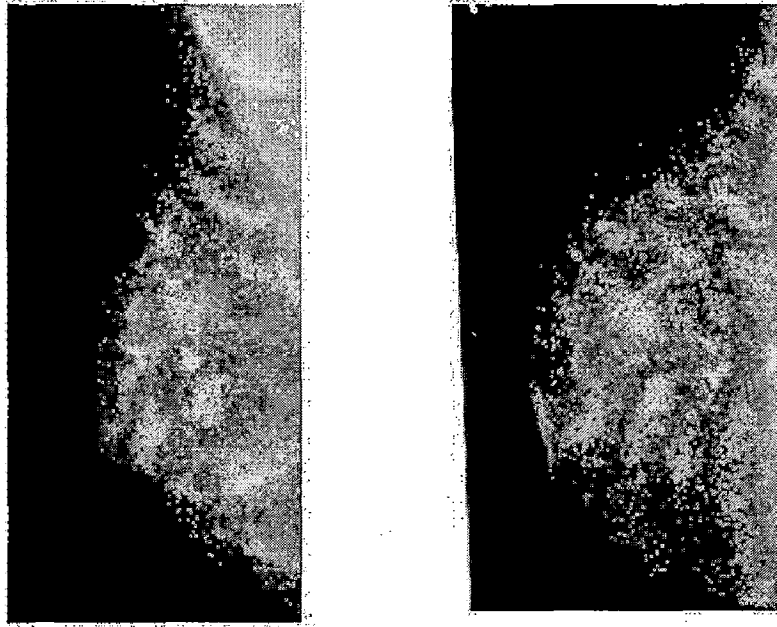


Figure 1.7: Higher probability of malignancy calcifications [1]

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

3. Some microcalcifications have low contrast to the background.

Due to above reasons it becomes too difficult to detect microcalcifications even for experienced radiologist. Consequently, computer assisted detection of microcalcification has aroused a great deal of interest.

### 1.3 Motivation for the thesis

The presence of microcalcification clusters (MCCs) is an important sign for the detection of early breast carcinoma. An early sign of breast cancer detected mammographically is the appearance of clusters of fine, granular microcalcification, and most of breast carcinomas reveal MCCs upon histological examinations. Despite of its proven effectiveness, screening still misses about 20% of cancers. Several studies have shown that double reading of mammograms (by second radiologist) improves the accuracy of mammogram interpretation. The desire to use computers in place of second radiologist, or as a prescreener to separate out clearly normal mammograms, is the motivations for computer aided detection system development

## 1.4 Literature survey

Diagnostic features in mammograms vary widely in size and shape. Classical image enhancement techniques cannot adapt to the varying characteristics of such features. An adaptive method for enhancing the contrast of mammographic features of varying size and shape is presented in this paper [4] [3]. Dongming Zhao et.al. proposed detection algorithm based on adaptive thresholding by making use of morphology [5]. Tomislav Stojic et.al. [6] proposed local contrast enhancement and background texture suppression algorithm based on morphology. This shows to what results on some images but still it fails for different databases. Adaptive histogram equalization has been shown to enhance contrast in radiological images, which in general have a large global dynamic range, but small local feature gray level variations [7].

Locally adaptive histogram equalization performs histogram equalization independently over different segments of the image. In these methods, determination of local neighborhood dimensions is a critical step. A given neighborhood size and shape may not be equally effective in enhancing all areas of an image. Enhancement by background removal [4] is a direct method of reducing the slowly varying portions of an image, to allow increased gray level variations in the image details. It is usually performed by subtracting the low pass filtered version of the image from itself. Unsharp masking is a simple version of this procedure. Spline filtering and gray scale morphological processing are two methods of estimating the image background. The background extraction technique should be adaptive to the local image characteristics to truly identify the image background.

Morphological operations can be employed for many image processing purposes, including edge detection, segmentation, and enhancement of images [5]. The simplicity of the mathematical morphology comes from the fact that a large class of filters can be represented as the combination of two simple operations on the image; the erosion and dilation, the top-hat transform is used for enhancement of microcalcifications. Many digital image enhancement algorithms have been developed and were reviewed in [4]. These techniques generally enhance image contrast, but they simultaneously amplify noise and artifacts. In interpreting mammograms, noise and artifacts are undesirable and can potentially lead to false diagnoses. Michael Wirth et.al. [8] presented an approach to enhancing the contrast of microcalcifications in mammograms using a contrast enhancement algorithm based on a combination of morphological enhancement and non-flat structuring elements. Given that microcalcifications appear as small domes on a 3D relief of a mammogram, enhancement is achieved by using structuring

elements which have a 3D form.

Entropy thresholding forms another class of algorithms widely used in thresholding. It has also been used in detection of MCCs in mammograms. Moti Melloul and Melloul et al. [9] have for the first time used a three dimensional co-occurrence matrix for threshold calculation.

The threshold selection problem is solved by minimizing the cross entropy [10] between the image and its segmented version. The cross entropy is formulated in a pixel-to-pixel basis between the two images and a computationally attractive algorithm employing the histogram is developed. Without making a priori assumptions about the population distribution, this method provides an unbiased estimate of a binarized version of the image in an information theoretic sense.

Most systems extract features from the texture, spatial and spectral domains. Textural features represent properties of the object's surface. Haralick et al. [11] developed a graytone spatial-dependence matrix method (known as the co-occurrence matrix method) to describe patterns of graylevel repetition. Statistics derived from the co-occurrence matrix are used to characterize the textural pattern. The co-occurrence matrix method is widely used in biomedical image processing [12]. Spatial domain features encode shape, pixel intensity, and other object characteristics. Zheng et al. [13] calculate the pixel intensity variance and energy variance of each image block. Zheng et al. (1996) calculated the discrete cosine transform (DCT) of X-ray images and derived the block activity and spectral entropy from the DCT coefficients [12].

Back-propagation neural networks (BPNN) were once commonly used for microcalcification classification [14]. However, a major disadvantage of BPNNs is that they lack a scheme for choosing training parameters such as the transfer function, number hidden layer nodes, and convergence algorithm. General regression neural networks (GRNNs) were later introduced to replace the BPNN [12]. Unlike BPNNs, the GRNN is based on non-linear regression theory for function estimation. One major advantage of a GRNN is the simplicity of the network training, which does not require time-consuming procedures based on trial-and error. The drawback of GRNN's is that large data sets increase network size and computational load. In addition to the problem of dimensionality, the GRNN does not generate the optimal boundary for problems such as microcalcification classification, which have two classes (microcalcification present and microcalcification absent). For two-class problems, an optimal classifier should in principle seek the hyperplane that produces the largest margin of separation (structural risk minimization) instead of minimizing learning error alone [16]. Support Vector Machines (SVMs) minimize the structural risk by using non-linear mapping to

transform the input space to a high dimensional feature space, where a hyperplane is constructed to maximize the separation margin. SVM has recently been employed in mammographic microcalcification classification. Though experimental results show that SVM outperforms other types of neural networks, the impact of relevant feature selection on classification performance has not been addressed.

## 1.5 Area of thesis

The basic objective of computer-aided diagnosis (CAD) is to provide a second opinion on radiologist's image readings, with a view to improve the quality and productivity by improving the accuracy of radiological diagnosis and reducing the image reading time. As enhancement and detection of microcalcification are most important and crucial from development of CAD system for early detection of breast cancer. The main goal of this work is to make contributions to a computer aided diagnosis system, which can provide a second opinion to radiologists on a routine clinical basis. Proposed CAD system is as shown in figure 1.8

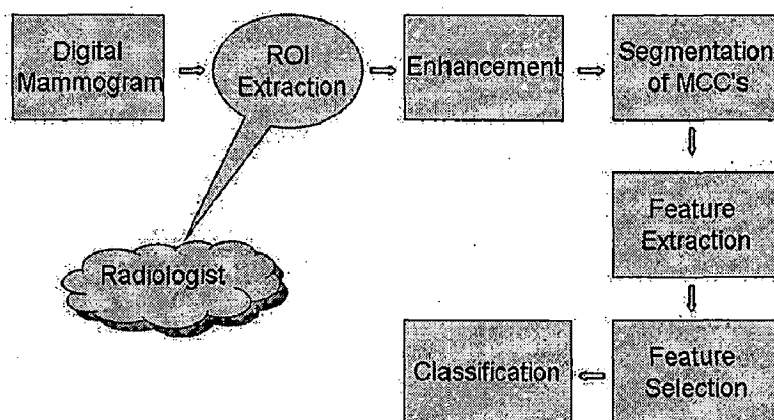


Figure 1.8: Block diagram of a typical computer-aided diagnosis (CAD) system

In the proposed method, first step is selection or marking of region of interest (ROI) on mammogram. Then, the ROI selected from the digitized mammogram is de-noised and enhanced. The next stage is designed to find suspicious areas containing MCCs, and to separate the MCCs from the background that will be used for extracting features of MCCs. Next the features of MCCs are extracted and selected, and finally MCCs are classified into benign and malignant.

## 1.6 Organization of thesis

### Chapter 1.

gives an introduction to basics of mammography, mammogram abnormalities on the mammogram, Problems in interpretation of the mammogram. It also gives a brief literature review and also list area of thesis work done.

### Chapter 2

describes preprocessing of mammograms, morphological enhancement algorithms studied and implemented and finally formation of synthetic images for evaluation and results.

### Chapter 3

describes different evaluation indices such as Contrast Improvement Index (CII), and detail variance to background variance (DV/BV) ratio, used for evaluating different enhancement techniques along with this image profiles is also described.

### Chapter 4

describes minimum cross entropy thresholding applied to synthetic mammogram images for segmentation of the microcalcifications.

### Chapter 5

describes features extracted from microcalcifications and

### chapter 6

describes labelling and object reduction techniques and finally basics of Support vector machine (SVM) used for classification along with results.

### Chapter 7

summarizes the results, conclusions and describes directions of future work.



# Chapter 2

## Mammogram enhancement

### 2.1 Introduction

Image enhancement refers to attenuation or sharpening of image features of interest such as edges, boundaries or contrast to make the enhanced image more useful for analysis. In our case as we are interested in microcalcifications i.e bright spots of higher intensities on the mammogram image. There are two approaches by which we can enhance contrast of mammogram image. First one is by increasing contrast of suspicious areas of mammogram. Second one involves the removal of background noise from the mammogram image [15].

This chapter gives brief overview of traditional contrast enhancement techniques, basics of morphology. Finally describes two methods used in our thesis work by removal of background for enhancement of mammogram.

### 2.2 How to increase contrast of image

Contrast of an gray scale image is difference in gray level values of foreground and background of an image in a vague language.

Low contrast image means image pixel intensity values concentrated near a narrow range (mostly dark, or mostly bright, or mostly medium values)

to enhance the contrast of an image means change the image value distribution to cover a wide range.

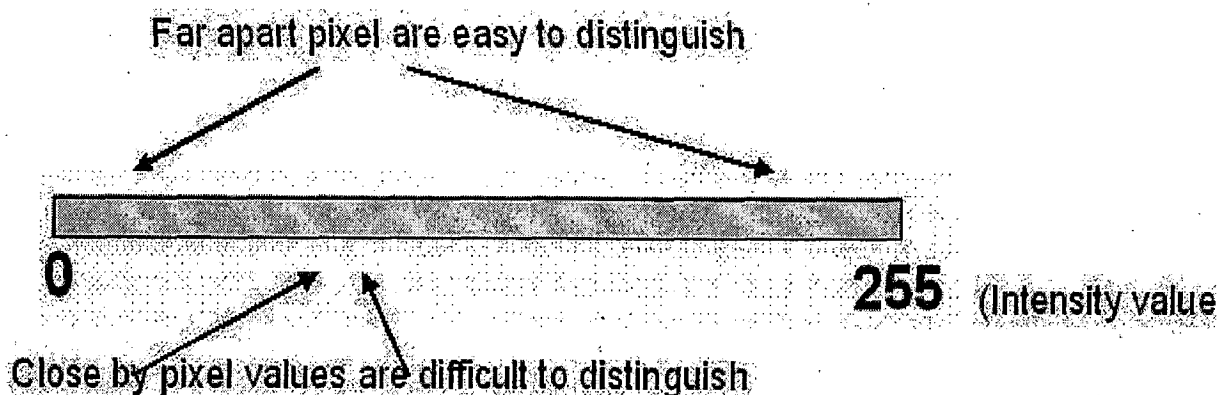


Figure 2.1: How to increase contrast of image

## 2.3 Preprocessing of mammogram

Images from McGill University database where region of interest are marked by experienced radiologist used for analysis. We cropped region of interest from original mammogram.

Cropping cuts of the unwanted portions of the image, thus all the unnecessary background information and most of noise are eliminated. An example of the cropping that eliminates the unwanted black background is given in figure. By clipping large number of background pixels from the images, storage requirements, I/O time, and image processing time is significantly reduced. Cropping operation is as shown in figure 2.2.

## 2.4 Mammogram enhancement techniques

In thesis work we have implemented CLAHE traditional technique for comparison and two morphological enhancement methods which make use of background removal for enhancement.

In this section we will study these techniques in detail.

### Contrast Limited Adaptive Histogram Equalization (CLAHE)

Histogram equalization (HISTEQ) method is a well-known gray scale manipulation technique. In histogram equalization, the goal is to map the input image to the output image so that gray values in the output image are uniformly distributed. For most practical images, gray values need to be redistributed. In histogram equalization we try to spread gray values uniformly over the full gray-scale range. It increases the contrast

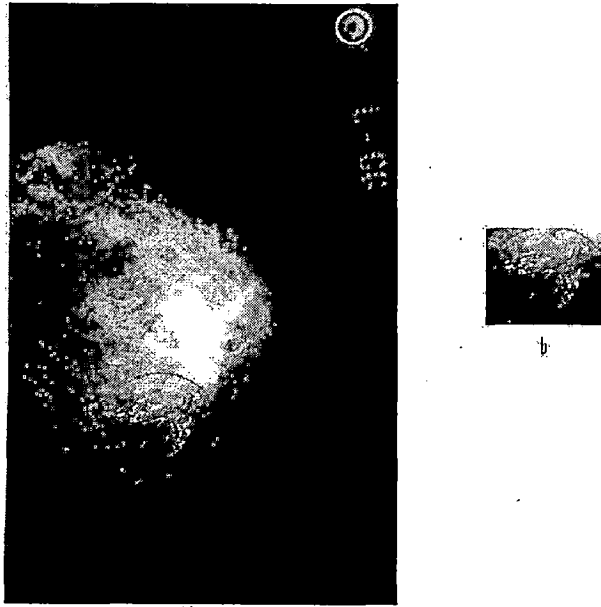


Figure 2.2: Case2.jpg from (McGill University database) a. Original image b. Cropped image

range in an image by increasing the dynamic range of gray levels. Another variation of conventional histogram equalization is Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithm .

Contrast-limited adaptive histogram equalization operates on small data regions (tiles) rather than the entire image. Each tile's contrast is enhanced so that the histogram of each output region approximately matches the specified histogram (uniform distribution by default). The contrast enhancement can be limited in order to avoid amplifying the noise, which might be present in the image. The histogram filter performs a so-called contrast limited adaptive histogram equalization (CLAHE) on the data set. The CLAHE algorithm partitions the images into contextual regions and applies the histogram equalization to each one. This evens out the distribution of used gray values and thus makes hidden features of the image more visible.

### Top-hat filtering

Morphological contrast enhancement is based on the notion of morphological top-hats which were first proposed by Meyer [5]. A top-hat is a residual filter which preserves those features in an image that can fit inside the structuring element (SE) and removes those that cannot.

The top-hat transform is used to segment objects that differ in brightness from the

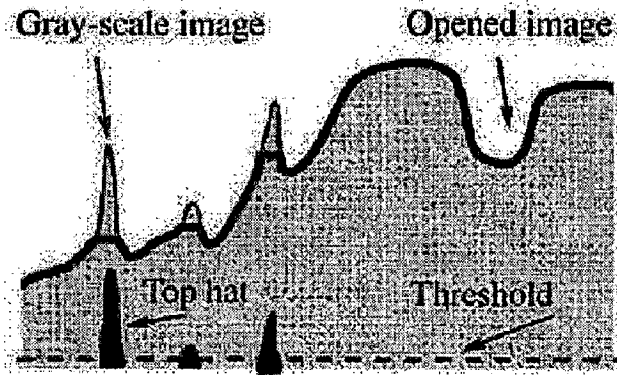


Figure 2.3: Tophat filtering [17]

surrounding background in the images with uneven background intensity.

The high-intensity regions, i.e. features that cannot accommodate the structuring element are removed from an image by performing a structural opening. The top-hat by opening,  $\hat{A}$ , is defined as the difference between the original image,  $A$ , and its grayscale opening using structuring element  $SE$ .

TopHat transform = Original image - Opened Image

$$B = A - [(A \ominus SE) \oplus SE] \quad (2.1)$$

In top-hat filtering for microcalcifications, the selection of structuring element is crucial and plays a central role in enhancement of microcalcifications in the mammogram. But as microcalcifications vary in shape and size, it becomes too difficult to decide structuring element for morphological operators. In literature previous approaches used small square or disk shaped flat structuring element [18], which may be inappropriate for natural shape of microcalcifications. Some has also used non-flat structuring element for enhancement by taking care of intensity along with shape of microcalcifications [8]

### H-dome transformation [19]

Considering each mammogram as a topographic representation, each microcalcification appears as an elevation or dome constituting a regional maximum. So problem of enhancement reduces to enhancement of domes from the mammogram. It makes use of morphological reconstruction.

Morphological reconstruction is part of a set of image operators often referred to as

geodesic. In the binary case, reconstruction simply extracts the connected components of a binary image  $I$  (the mask) which are "marked" by a (binary) image  $J$  contained in  $I$ . This transformation can be extended to the grayscale case, where it turns out to be extremely useful for several image analysis tasks.

### Geodesic distance

Given a set  $X$  (the mask), the geodesic distance between two pixels  $p$  and  $q$  is the length of the shortest paths joining  $p$  and  $q$  which are included in  $X$ . Note that the geodesic distance between two pixels within a mask is highly dependent on the type of connectivity which is used.

### Geodesic dilation

Geodesic dilation is defined as:

Let  $X \in Z^2$  be a discrete set of  $Z^2$  and  $Y \subseteq X$ . The geodesic dilation of size  $n \geq 0$  of  $Y$  is subset of  $X$  is the set of the pixels of  $X$  whose geodesic distance to  $Y$  is smaller or equal to  $n$ :

$$\partial_X^{(n)}(Y) = \{p \in X | d_{X(p,Y)} \leq n\} \quad (2.2)$$

From this definition, it is obvious that geodesic dilations are extensive transformations. In addition, geodesic dilation of a given size  $n$  can be obtained by iterating  $n$  elementary geodesic dilations.

The elementary geodesic dilation can itself be obtained via a standard dilation of size one followed by an intersection

$$\partial_X^{(1)} = (Y \oplus B) \cap X \quad (2.3)$$

The reconstruction of  $X$  from  $Y \subseteq X$  is obtained by iterating elementary geodesic dilations of  $Y$  inside  $X$  until stability.

In other words

$$p_X(Y) = \bigcup_{n \geq 1} \partial_X^n(Y) \quad (2.4)$$

Just like binary reconstruction extracts those connected components of the mask which are marked, grayscale reconstruction extracts the peaks of the mask which are marked by the marker-image. As stated in earlier chapters that microcalcifications

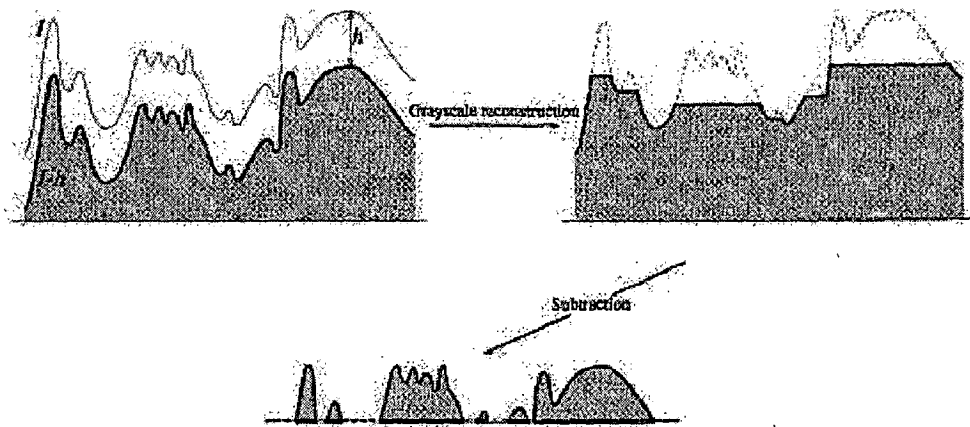


Figure 2.4: Morphological reconstruction [20]

are regional maximum in mammogram if we consider mammogram as a topographic representation.

### The grayscale reconstruction

Grayscale reconstruction  $\rho_I(J)$  of  $I$  from  $J$  is obtained by iterating grayscale dilations of  $J$  under  $I$  until stability is reached.

$$\rho_I(J) = \bigvee_{n \geq 1} \partial_I^{(n)}(J) \quad (2.5)$$

Reconstruction provides a very efficient method to extract regional maximum from grayscale images. This technique extends to the determination of “maximal structures”, which we call h-domes.

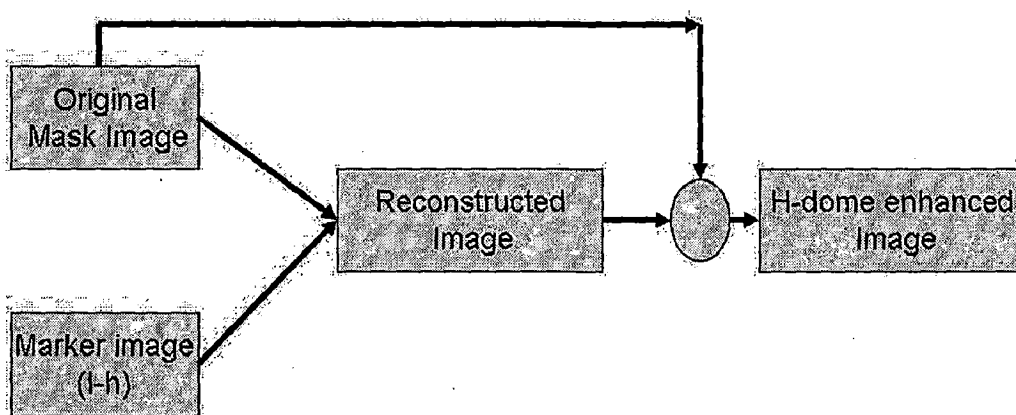


Figure 2.5: Block diagram of H-dome transformation mammogram enhancement

Regional maximum  $M$  of grayscale image  $I$  is a connected components of pixels with a given value  $h$ , such that pixel in the neighborhood of  $M$  has strictly lower value.

## **h-dome Transformation**

h-dome transformation (h-dome) of the image is given by

$$h - dome(I) = I - \rho(I - h) \quad (2.6)$$

where

$\rho$  = reconstruction of image which removes peaks from the image

Thus after subtracting morphologically reconstructed image from the original image we get peaks of the image. h-dome is independent of size and shape of domes in the images.

## **2.5 Results and discussions**

### **2.5.1 Synthetic images**

One of the fundamental problems in evaluating the performance of algorithms for detection of microcalcifications is that one is never certain if a false positive is not an actual calcification. This is because the radiologist is prone to human error. While most of the research has concentrated on detection methods little attention has been devoted to defining a gold standard for an objective evaluation of computerized detection accuracy. Some authors compare their results to the locations indicated by an experienced radiologist while others confirm the presence of microcalcification by biopsy or magnification. Biopsy and magnification are able to confirm the presence of microcalcifications and can accurately evaluate the rate of false positives when the aim is to detect clusters of microcalcifications on the whole mammogram. However these tools are not able to determine the number of microcalcifications and to localize them individually.

Therefore for evaluating and analysing different enhancement algorithms applied on mammogram, synthetic images are created with known microcalcifications on them. While preparing different synthetic images we have taken care of shapes and sizes of microcalcifications. Some of the synthetic images are shown in figure 2.6. In first image microcalcifications are shown by arrows.

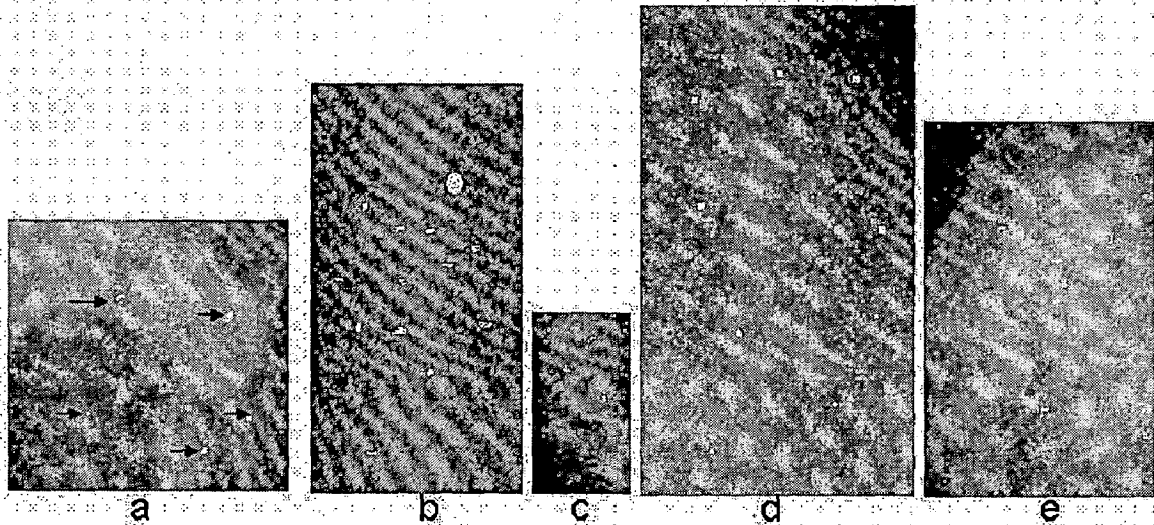


Figure 2.6: Synthetic images : a.Test1 b.Test2 c.Test3 d.Test4 d.Test5

In test1 image microcalcifications are indicated by arrows. Test1 image contains 5 microcalcifications as shown. Test2,Test3,Test4,Test5 images contains 13,5,10,10 microcalcifications respectively. In all images we induced simulated microcalcifications with different sizes and varying shapes as shown in above figure 2.6.

## 2.5.2 Image Database

After getting confidence of performance of different algorithms on test synthetic images, these are algorithms applied on standard mammogram images from the McGill University database with known region of interest ROI .

Because of the minute sizes of the microcalcification, the image resolution needs to be very high. The image size in database has been clipped and padded to become 1,024\*1,024 pixels. The dynamic range of the pixel is eight bit- that is, a gray scale of 0 to 255. Background tissues included in this database include fatty, fatty glandular, and dense glandular. The database yields abnormalities such as calcification, well defined circumscribed masses, masses with speckles, ill defined masses, architectural distortion, asymmetry and normal. The ground truth of each mammogram is also included in the database, which provides the location where the abnormality is present by marking a circle marked around it.



### 2.5.3 Results

Algorithms stated applied on test synthetic and after getting confidence applied on standard database, results are shown as below.

#### Contrast Limited Adaptive Histogram Equalization (CLAHE)

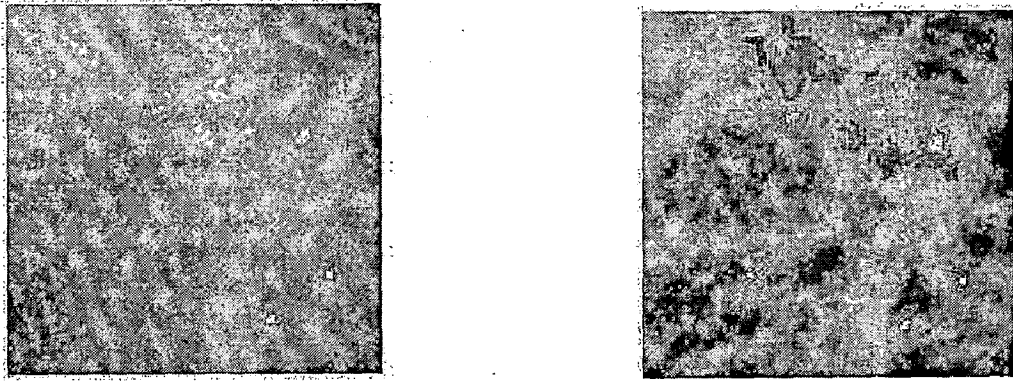


Figure 2.7: CLAHE : Test1 Original image & enhanced Image

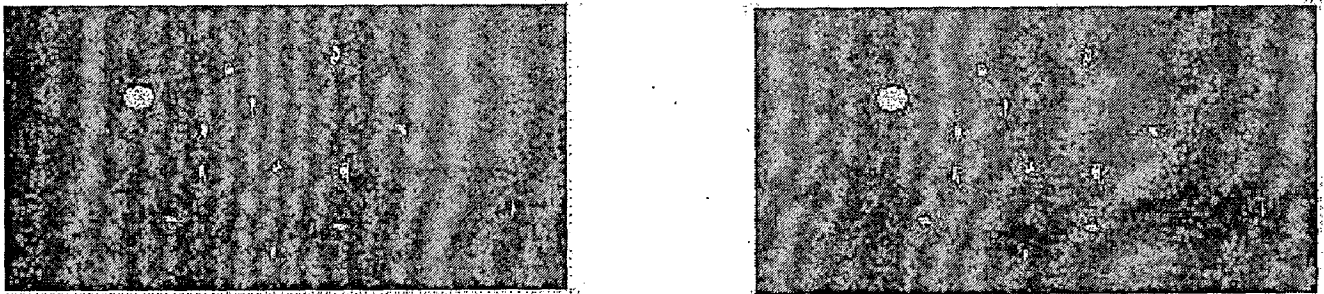


Figure 2.8: CLAHE : Test2 Original image & enhanced Image

As shown in above figures 2.7 & 2.8 ,CLAHE technique enhances simulated microcalcifications along with background noise which is not desirable. Clearly, in Test1 image it can be seen that background noise is also enhanced along with simulated microcalcifications, which makes it difficult to detect microcalcifications in dense structures. Even for the Test2 image where background is smooth it shows noise around microcalcifications. This is not suitable for microcalcification enhancement. All synthetic images are tested for this and showing similar enhancement.

#### Tophat filtering

As stated in literature most of the researcher tried and used disk and square shaped structuring element for tophat filtering. But as microcalcifications are of different sizes

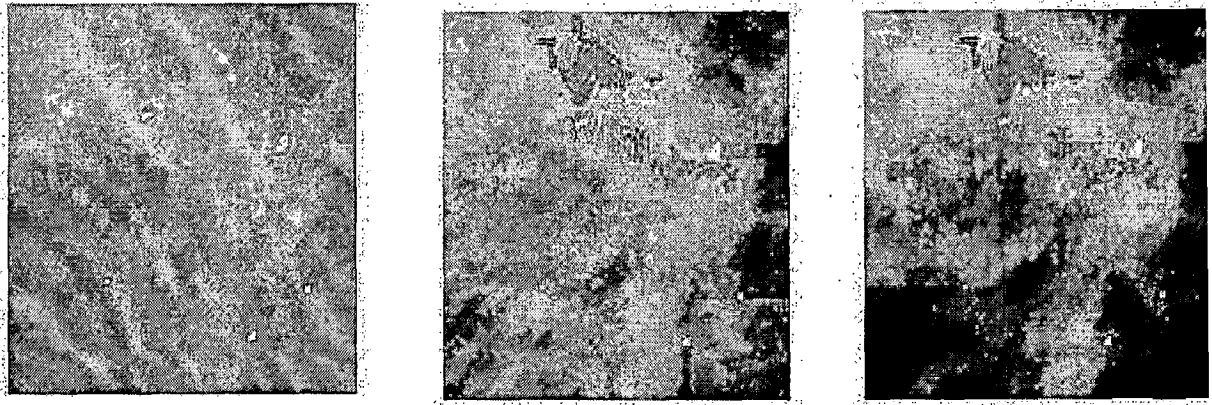


Figure 2.9: Tophat filtering with square SE with side=50,100 Test1 image

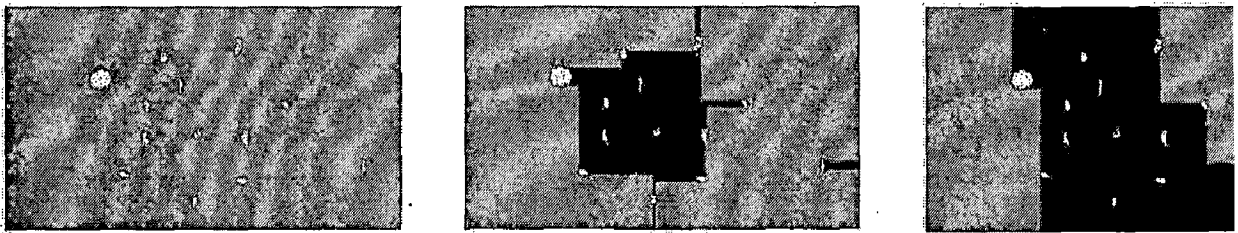


Figure 2.10: Tophat filtering with square SE with side=50,100 Test2 image

and shapes it becomes too difficult to choose structuring element.

We have tested tophat with different structuring element enlisted in literature, proposed to be better than other structuring elements.

From the above two figures 2.9 & 2.10 we can see the effect of the structuring element in tophat filtering. Even though if some structuring element works for one image it may happen that it will not give good result on other images. This is drawback of tophat filtering in enhancement of mammograms.

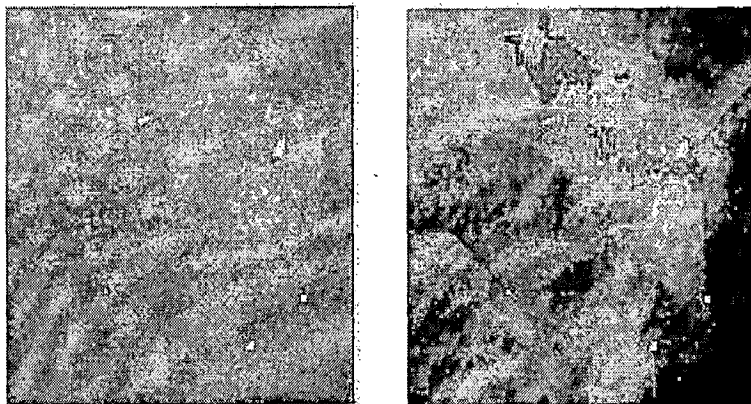


Figure 2.11: Tophat filtering with disk with radius 100 for Test1 image

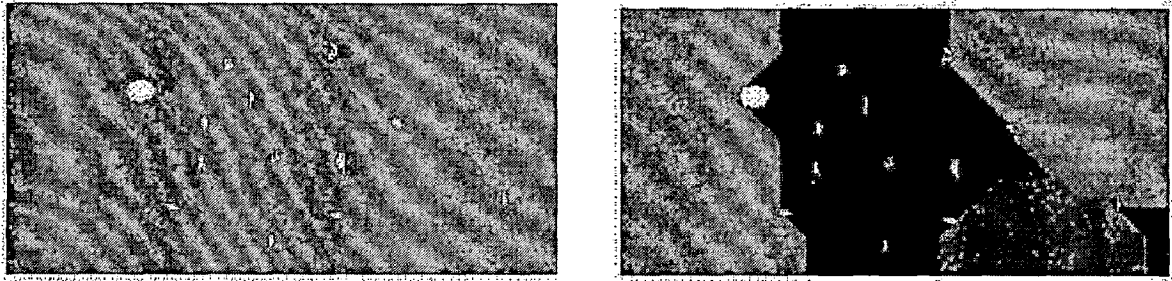


Figure 2.12: Tophat filtering with disk with radius 100 for Test1 image

In literature some researchers also used non-flat structuring element for enhancement [8], we have also tested non-flat structuring element on our images.

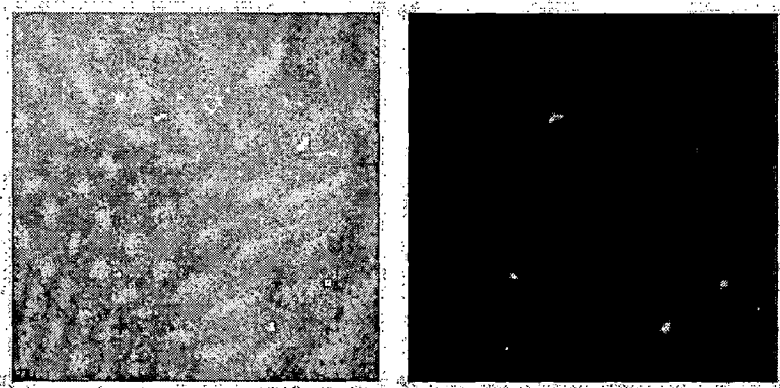


Figure 2.13: Tophat filtering with ball SE (non-flat) with height=50, radius=17, Test1 image

It seems that this structuring element works for some images but when we tested on other images it shows similar effect of dependency on structuring element.

From above we can conclude that tophat filtering technique can be used for particular sets with specific structuring element. Clearly it is dependent on structuring element, choice of which is most difficult.

Now we will move to h-dome transform and will see how it overcomes this drawback.

### **h-dome filtering**

h-dome transformation is applied to test images with different values of  $h$ . As stated in [20] The choice of  $h$  turns out not to be a critical operation, since a range of values yield correct results.

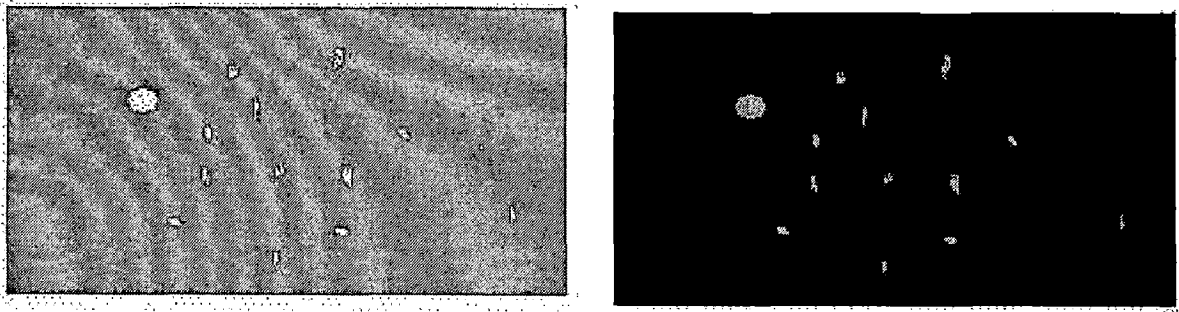


Figure 2.14: Tophat filtering with ball SE with height=50,radius=17, Test2 image

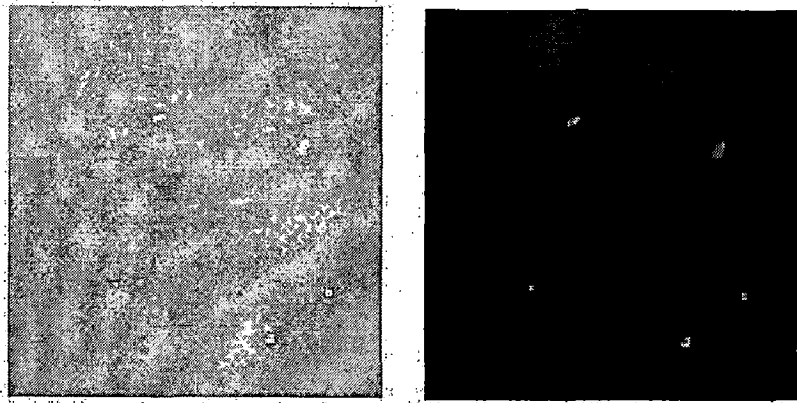


Figure 2.15: h-dome filtering Test1 image

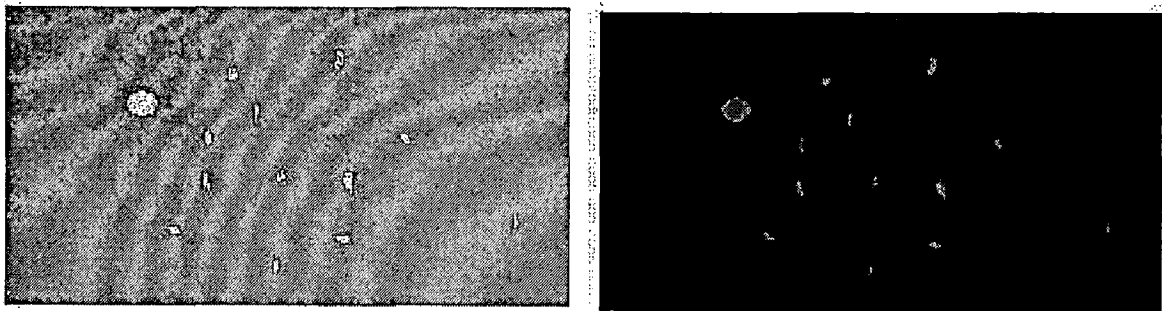


Figure 2.16: h-dome filtering Test2 image

In h-dome it removes completely background portion in image. This can be very clear if we give close look on details in Test2 image, original image is having uniform background, which is completely removed by h-dome in enhanced image. As seen from both the images shapes and sizes of microcalcifications are preserved after enhancement.

Results of applying proposed algorithms on standard database images are shown in figure 2.17.

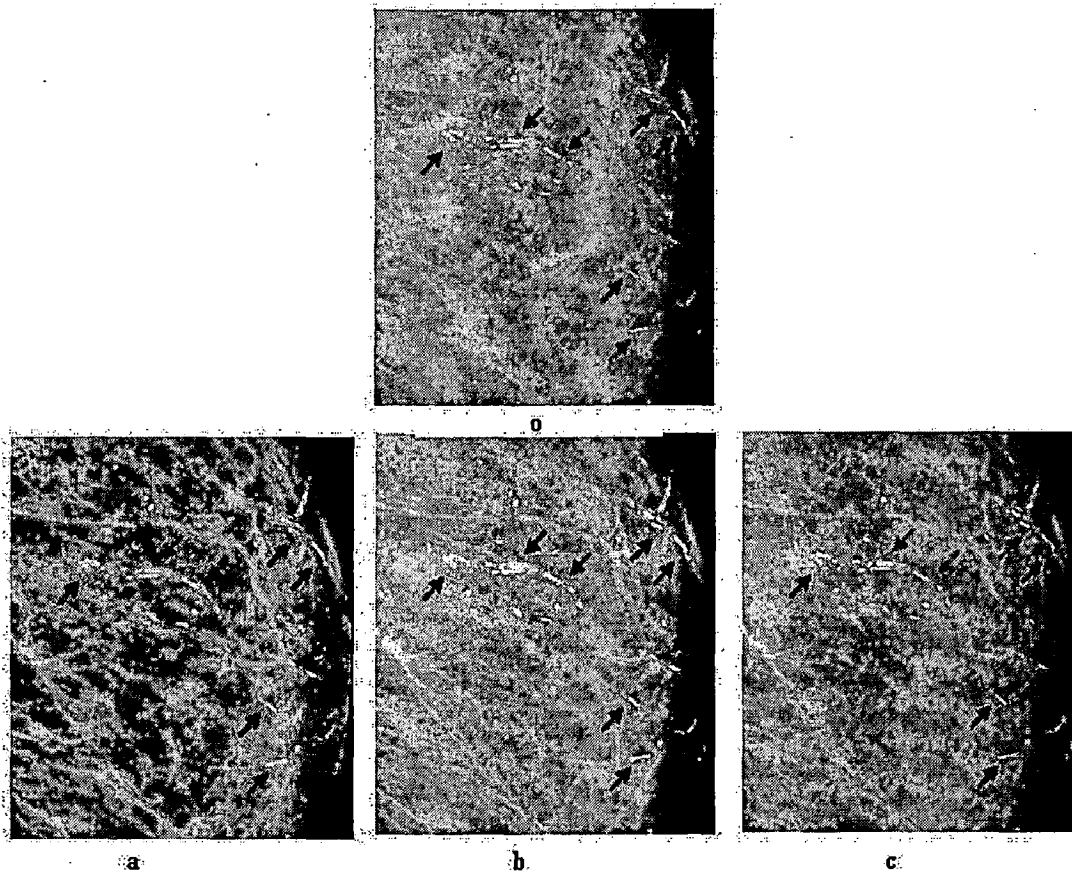


Figure 2.17: o.Cropped case25Lmlo.jpg (Mcgill database)a.CLAHE enhanced b.tophat enhanced c.h-dome enhanced

From the figure 2.17 (a) , if we closely observe, it can be noted that noise enhancement is more in CLAHE as compared to other techniques. So though it seems that it is increasing contrast of image. But it from microcalcifications point of view, it is making image difficult for detecting microcalcifications.

In figure 2.17(b), tophat filtered image is shown , which clearly indicates good enhancement, but if we take a close look it can be found that tophat has changed shapes of some of microcalcifications.

In figure 2.17(c), h-dome enhanced image shows clearly removal of background without affecting shapes and sizes of microcalcifications.

*It can be observed that the H-dome transformation gives better results as compared to tophat and CLAHE irrespective of structuring element used. Occasionally tophat may give better results for particular image but overall h-dome outperforms tophat filtering method.*

# Chapter 3

## Quantitative analysis of mammogram enhancement algorithms

### 3.1 Introduction

This chapter addresses the various quantitative measures for evaluation of enhancement algorithms used to improve quality the image.

In this chapter, comparison of our proposed algorithms using two of the evaluation indices listed in literature i.e. CII(Contrast Improvement Index) and DV/BV ratio is done. Finally we used image profiles of a mammogram for visual inspection of performance of enhancement.

### 3.2 Evaluation Indices

The improvement in images after enhancement is often very difficult to measure. A processed image can be said to be an enhanced over the original image if it allows the observer to better perceive the desirable information in the image. In mammograms, the improved perception is difficult to quantify. Use of statistical measures of gray level distribution as measures of local contrast enhancement (for example, variance or entropy) have not been particularly meaningful for mammogram images. A number of images which clearly showed improved contrast showed no consistency, as a class, using these statistical methods.

### 3.2.1 Contrast Improvement Index (CII)

Laine et al. [21] proposed the contrast improvement index CII as the measure of the enhancement performance:

$$CII = C_{processed}/C_{Original} \quad (3.1)$$

Where  $C_{processed}$  and  $C_{Original}$  are the contrasts for a ROI in the processed and the original images, respectively. The contrast  $C$  of a region is defined by

$$C = \frac{f - b}{f + b} \quad (3.2)$$

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 the background. This definition of contrast has the advantage of being independent of the actual range of gray levels in the image. The bigger the value of CII, the better the performance.

### 3.2.2 Background and detail region variance

Even if direct visual inspection remains the most effective way for evaluating the qualities of an enhancement scheme, it can be useful to assign them some objective measure; a simple but fitting method has been devised in order to make the comparison among the processed images easier. It is based on the separate estimate of the local variance on the image details (detail variance) and in uniform areas (background variance).

First, we define such two regions in the ideal image; for each pixel we evaluate the variance in a  $n \times n$  window: if the variance is larger than a fixed threshold, the pixel belongs to the detail region; otherwise, it belongs to the background region. In this way, a binary map image is generated where, say, white pixels indicate details and black ones indicate background. Then, we evaluate the variance in a  $n \times n$  window centered on each pixel of the processed image: if the corresponding pixel in the binary map image is white, this variance is accumulated in a detail region variance register, otherwise in a background region register. We end up with two numbers: the average background variance (BV), obtained dividing the detail region variance register by the number of pixels in the detail region, and the average detail variance (DV), obtained in an analogous way [22].

Reasonably enhancement techniques should yield for the processed image a DV value larger than the one of the original image, while the BV value should remain

unchanged or, if possible, slightly decrease. In order to apply this method, one has to choose two parameters, i.e. the size  $n$  of the window and the detail/background threshold. Of course, the absolute values of DV and BV depend on such parameters; nevertheless, their relative values remain very similar in a wide range of choice, indicating that they represent two reliable quality factors when it is needed to compare the performance of different operators.

### 3.2.3 Results and discussions

First Index used for evaluation of enhancement algorithms is Contrast Improvement Index(CII).As we have seen higher the value of CII, better is the enhancement of mammogram.CII is calculated for images in the database explained in section 2.5.2. A graph of McGill Images vs CII is plotted as shown in figure 3.1. It can be observed that CII profile for H-dome algorithm has maximum area under the curve as compared to two other techniques used for comparison i.e.tophat filtering and CLAHE. From graph it can be concluded that H-dome seems to better than other two methods.

Second evaluation index used is DV/BV ratio.Graph of Detail variance(DV) vs McGill Images ,Background variance(BV) vs McGill Images and finally combining DV/BV ratio vs McGill Images are plotted. By looking at figure 3.2 , it can be observed that DV profile for CLAHE shows maximum area under curve.That means CLAHE enhances detail portion of image in better way as compared to other two techniques. But as we have seen it also enhances noise so it is not suitable. H dome shows least area under DV curve, means it enhances detail portion comparatively less effectively than other methods.

By looking at figure 3.3, it can be observed that BV profile for CLAHE shows maximum area under curve, as we have seen it also enhances noise so it is not suitable. H dome shows least area under BV curve that indicates it rejects slow varying background better than other two methods.

Finally by looking at figure 3.4 , it can be observed that DV/BV profile for h-dome shows maximum area under curve , from which we can conclude that h-dome is best amongst these three methods.



**Table 3.1 Comparison of Contrast Enhancement Index (CII) for different enhancement techniques applied on ROIs of mammograms containing microcalcifications .**

| name of image              | CLAHE    | tophat  | hdome    |
|----------------------------|----------|---------|----------|
| case11Lcc small ans.jpg    | 0.493    | 1.2987  | 5.6131   |
| case11Lmlo small ans.jpg   | 2.7793   | 0.5437  | 1.9579   |
| case11Rcc small ans.jpg    | 0.1079   | 2.2725  | 8.7316   |
| case11Rmlo small ans.jpg   | 27.1092  | 13.0023 | 58.1538  |
| case11 lLcc small ans.jpg  | 184.6198 | 61.48   | 6.8459   |
| case11 lRmlo small ans.jpg | 0.1241   | 3.2574  | 8.7838   |
| case15Lcc small ans.jpg    | 0.6527   | 1.2991  | 2.2072   |
| case15Lmlo small ans.jpg   | 1.3765   | 1.8636  | 5.6711   |
| case15Rcc small ans.jpg    | 0.481    | 4.5599  | 5.2196   |
| case15Rmlo small ans.jpg   | 0.6418   | 1.833   | 2.392    |
| case15 lLcc small ans.jpg  | 0.2499   | 5.0706  | 4.918    |
| case15 lLmlo small ans.jpg | 0.6932   | 1.498   | 1.8621   |
| case15 lRmlo small ans.jpg | 0.0547   | 0.1008  | 6.2143   |
| case22Lcc small.jpg        | 0.8216   | 0.9868  | 0.596    |
| case25Lcc small ans.jpg    | 1.2564   | 1.1051  | 5.9704   |
| case25Lmlo small ans.jpg   | 2.2343   | 1.559   | 12.5288  |
| case25Rcc small ans.jpg    | 3.1077   | 0.8041  | 7.2974   |
| case25Rmlo small ans.jpg   | 13.0431  | 5.0641  | 116.1576 |
| case37Lcc ans small.jpg    | 2.5223   | 1.5354  | 2.7563   |
| case37Lmlo ans small.jpg   | 0.9028   | 1.1311  | 2.4467   |
| case37Rcc ans small.jpg    | 1.9805   | 2.3513  | 15.2404  |
| case37Rmlo ans small.jpg   | 1.3164   | 1.3675  | 3.8337   |
| case46Lcc small.jpg        | 0.4639   | 0.8848  | 2.8152   |
| case46Lmlo small.jpg       | 0.7666   | 0.7883  | 3.5919   |
| case46Rcc small.jpg        | 0.8822   | 1.0089  | 1.3031   |
| case46Rmlo small.jpg       | 0.4022   | 0.6451  | 3.2164   |
| case49Lcc small.jpg        | 0.9036   | 1.7152  | 6.4065   |
| case49Lmlo small.jpg       | 7.0585   | 5.499   | 60.241   |
| case49Rcc small.jpg        | 1.2976   | 0.8067  | 12.9275  |
| case49Rmlo small.jpg       | 1.1905   | 2.3978  | 5.208    |
| case6Lcc ans small.jpg     | 0.2116   | 1.1966  | 3.3881   |
| case6Lmlo ans small.jpg    | 0.4271   | 1.1629  | 1.7586   |
| case7Lcc small ans.jpg     | 0.8138   | 1.1458  | 1.9992   |
| case7Lmlo small ans.jpg    | 0.9706   | 1.1872  | 2.3027   |
| case7Rcc small ans.jpg     | 0.858    | 1.1433  | 0.4832   |
| case7Rmlo small ans.jpg    | 0.8328   | 1.1995  | 0.2002   |
| case17rmlo.jpg             | 0.5289   | 1.2171  | 0.5787   |
| case19Rcc.jpg              | 5.9842   | 1.0575  | 2.2162   |
| case2.jpg                  | 0.7273   | 1.1058  | 0.3297   |
| case21Lcc.jpg              | 3.3547   | 0.4764  | 55.5751  |
| case21Lmlo.jpg             | 2.1772   | 1.5631  | 13.3021  |
| case24Lcc.jpg              | 1.1596   | 0.8069  | 8.3485   |
| case24Lmlo.jpg             | 0.8293   | 0.6032  | 5.3621   |

|                |        |        |         |
|----------------|--------|--------|---------|
| case26Lcc.jpg  | 1.1631 | 0.7045 | 9.2249  |
| case26Lmlo.jpg | 1.7403 | 3.4294 | 70.565  |
| case27Rcc.jpg  | 0.6366 | 0.9595 | 0.975   |
| case27Rmlo.jpg | 0.2922 | 0.8349 | 3.3956  |
| case2lmlo.jpg  | 0.7183 | 1.0391 | 0.0819  |
| case36Lcc.jpg  | 5.7666 | 0.7903 | 94.6143 |
| case36Lmlo.jpg | 0.7396 | 1.2465 | 13.0567 |
| case39Lcc.jpg  | 0.8749 | 1.07   | 1.0026  |
| case39Lmlo.jpg | 0.7471 | 1.314  | 0.9289  |
| case44Rcc.jpg  | 0.1877 | 0.9878 | 3.7911  |
| case44Rmlo.jpg | 0.4768 | 1.1194 | 5.0349  |
| case5lcc.jpg   | 0.6757 | 0.8759 | 8.4431  |
| case5lmlo.jpg  | 0.8292 | 1.0102 | 1.0968  |

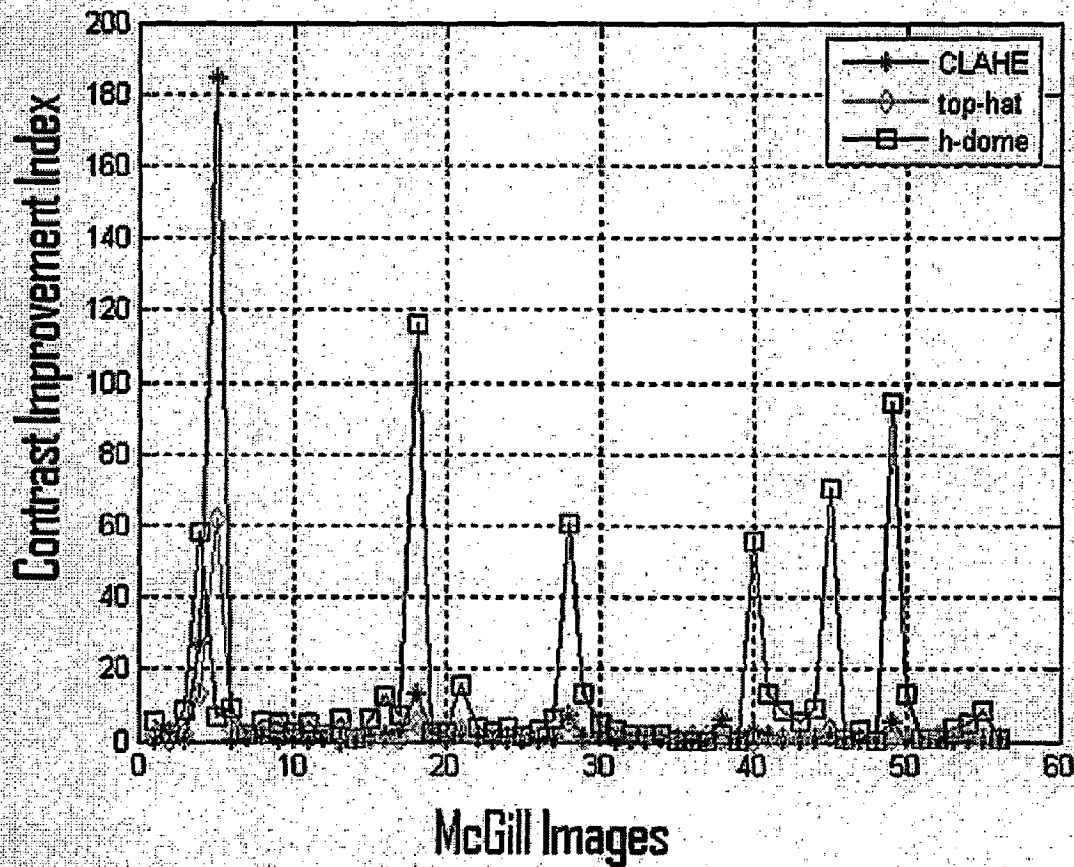


Figure 3.1: Comparison of Contrast Improvement Index(CII) for different enhancement methods as applied to some ROIs of digital mam-mograms containing microcalcifications obtained from McGill database.

**Table 3.2 Comparison of detail variance (DV) for different enhancement techniques applied on ROIs of mammograms containing microcalcifications.**

| Name of file               | CLAHE   | tophat  | hdome   |
|----------------------------|---------|---------|---------|
| case11Lcc small ans.jpg    | 119.776 | 86.3368 | 38.7288 |
| case11Lmlo small ans.jpg   | 126.182 | 89.4503 | 44.1579 |
| case11Rcc small ans.jpg    | 112.814 | 89.9484 | 72.9425 |
| case11Rmlo small ans.jpg   | 132.411 | 99.9194 | 42.0985 |
| case11 1Lcc small ans.jpg  | 129.036 | 95.5496 | 36.4984 |
| case11 1Rmlo small ans.jpg | 135.842 | 93.8154 | 57.5082 |
| case15Lcc small ans.jpg    | 124.678 | 111.813 | 82.6724 |
| case15Lmlo small ans.jpg   | 117.135 | 80.4082 | 43.4477 |
| case15Rcc small ans.jpg    | 128.954 | 92.0032 | 47.0946 |
| case15Rmlo small ans.jpg   | 112.421 | 86.1503 | 36.1243 |
| case15 1Lcc small ans.jpg  | 126.169 | 94.3014 | 59.7354 |
| case15 1Lmlo small ans.jpg | 127.263 | 114.6   | 78.4887 |
| case15 1Rmlo small ans.jpg | 125.435 | 72.7343 | 43.543  |
| case22Lcc small.jpg        | 117.872 | 112.971 | 78.3867 |
| case25Lcc small ans.jpg    | 117.816 | 86.4123 | 26.976  |
| case25Lmlo small ans.jpg   | 115.205 | 91.9685 | 24.772  |
| case25Rcc small ans.jpg    | 122.352 | 84.3921 | 14.6501 |
| case25Rmlo small ans.jpg   | 124.89  | 90.0429 | 23.3756 |
| case37Lcc ans small.jpg    | 119.728 | 79.9616 | 36.5203 |
| case37Lmlo ans small.jpg   | 106.264 | 45.7874 | 28.2992 |
| case37Rcc ans small.jpg    | 99.1138 | 57.2457 | 18.5983 |
| case37Rmlo ans small.jpg   | 100.519 | 56.3781 | 21.6777 |
| case46Lcc small.jpg        | 88.2482 | 46.48   | 19.4532 |
| case46Lmlo small.jpg       | 96.8439 | 62.4293 | 31.5731 |
| case46Rcc small.jpg        | 110.611 | 104.838 | 55.2147 |
| case46Rmlo small.jpg       | 100.819 | 65.7743 | 38.9152 |
| case49Lcc small.jpg        | 83.0532 | 50.2534 | 43.7349 |
| case49Lmlo small.jpg       | 116.52  | 96.2095 | 86.1048 |
| case49Rcc small.jpg        | 95.248  | 86.3938 | 64.2145 |
| case49Rmlo small.jpg       | 117.333 | 90.1161 | 83.643  |
| case6Lcc ans small.jpg     | 112.273 | 57.627  | 28.0965 |
| case6Lmlo ans small.jpg    | 116.902 | 44.7081 | 39.3306 |
| case7Lcc small ans.jpg     | 115.199 | 92.8786 | 22.7526 |
| case7Lmlo small ans.jpg    | 111.276 | 86.8622 | 30.4011 |
| case7Rcc small ans.jpg     | 126.971 | 109.583 | 38.5841 |
| case7Rmlo small ans.jpg    | 117.142 | 99.9647 | 22.9932 |
| case17rmlo.jpg             | 109.631 | 67.2741 | 26.879  |
| case19Rcc.jpg              | 119.856 | 77.3274 | 28.8286 |
| case2.jpg                  | 115.571 | 92.2304 | 52.313  |
| case21Lcc.jpg              | 104.615 | 62.8795 | 28.7914 |
| case21Lmlo.jpg             | 116.117 | 71.629  | 42.3082 |
| case24Lcc.jpg              | 109.219 | 52.1466 | 47.7569 |

|                |         |         |         |
|----------------|---------|---------|---------|
| case24Lmlo.jpg | 134.275 | 111.611 | 47.2136 |
| case26Lcc.jpg  | 110.472 | 43.2295 | 14.108  |
| case26Lmlo.jpg | 98.2392 | 56.3716 | 16.8233 |
| case27Rcc.jpg  | 94.6747 | 69.4211 | 36.7046 |
| case27Rmlo.jpg | 98.7278 | 72.7195 | 36.5211 |
| case21mlo.jpg  | 114.111 | 96.4431 | 59.1301 |
| case36Lcc.jpg  | 118.198 | 62.3338 | 17.6813 |
| case36Lmlo.jpg | 103.522 | 39.0239 | 7.4051  |
| case39Lcc.jpg  | 94.3056 | 61.3534 | 19.0714 |
| case39Lmlo.jpg | 105.372 | 69.9173 | 18.2997 |
| case44Rcc.jpg  | 108.137 | 64.0485 | 29.3574 |
| case44Rmlo.jpg | 122.717 | 83.7531 | 44.946  |
| case5lcc.jpg   | 109.869 | 62.2445 | 29.2852 |
| case5lmlo.jpg  | 114.791 | 108.552 | 65.5999 |

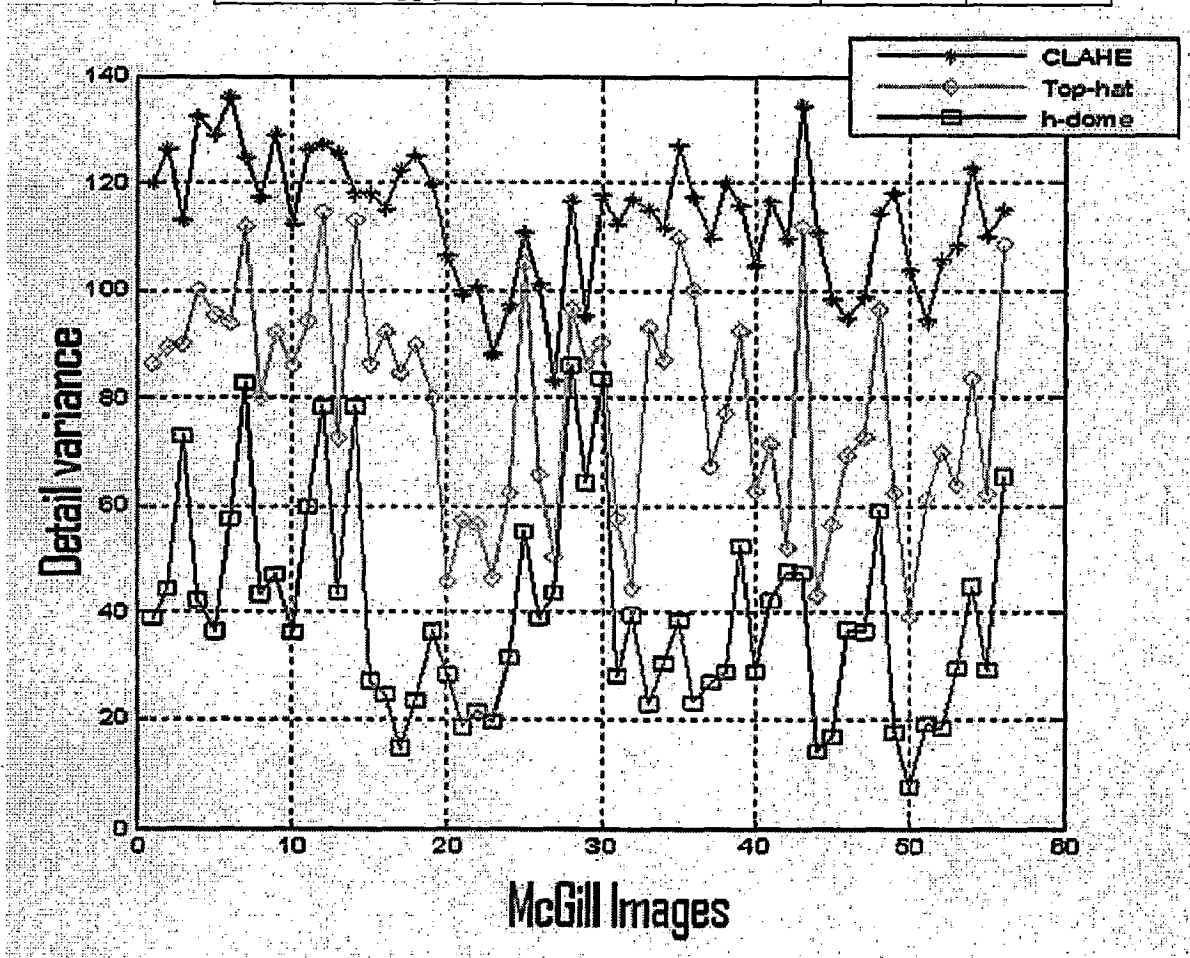


Figure 3.2: Comparison of Detail Variance(DV) for different enhancement methods as applied to some ROIs of digital mam-mograms containing microcalcifications obtained from McGill database.

Table 3.3 Comparison of Background variance (BV) for different enhancement techniques applied on ROIs of mammograms containing microcalcifications.

| Name of file               | CLAHE   | tophat  | hdome   |
|----------------------------|---------|---------|---------|
| case11Lcc small ans.jpg    | 65.7407 | 15.8149 | 5.2828  |
| case11Lmlo small ans.jpg   | 91.2847 | 25.8385 | 12.9519 |
| case11Rcc small ans.jpg    | 109.925 | 16.4882 | 5.3785  |
| case11Rmlo small ans.jpg   | 114.537 | 39.9862 | 13.9527 |
| case11 1Lcc small ans.jpg  | 101.205 | 23.8829 | 12.1611 |
| case11 1Rmlo small ans.jpg | 115.94  | 39.8565 | 12.4973 |
| case15Lcc small ans.jpg    | 87.8566 | 26.4083 | 8.8012  |
| case15Lmlo small ans.jpg   | 85.3867 | 23.3267 | 7.1336  |
| case15Rcc small ans.jpg    | 114.358 | 18.3642 | 5.5138  |
| case15Rmlo small ans.jpg   | 99.7888 | 30.4593 | 9.5311  |
| case15 1Lcc small ans.jpg  | 115.791 | 22.3212 | 8.5519  |
| case15 1Lmlo small ans.jpg | 102.121 | 40.2307 | 12.9579 |
| case15 1Rmlo small ans.jpg | 115.033 | 18.5551 | 7.7475  |
| case22Lcc small.jpg        | 23.1572 | 7.1073  | 2.9518  |
| case25Lcc small ans.jpg    | 52.3913 | 12.2884 | 4.6665  |
| case25Lmlo small ans.jpg   | 61.3048 | 14.6129 | 4.8919  |
| case25Rcc small ans.jpg    | 54.9611 | 12.3586 | 3.9521  |
| case25Rmlo small ans.jpg   | 64.2992 | 14.315  | 4.8156  |
| case37Lcc ans small.jpg    | 71.9347 | 15.7604 | 7.4904  |
| case37Lmlo ans small.jpg   | 63.1592 | 7.3134  | 3.0754  |
| case37Rcc ans small.jpg    | 66.488  | 14.7954 | 4.9113  |
| case37Rmlo ans small.jpg   | 44.5239 | 9.1136  | 2.629   |
| case46Lcc small.jpg        | 48.3099 | 8.8642  | 3.9508  |
| case46Lmlo small.jpg       | 56.6193 | 11.4309 | 3.6581  |
| case46Rcc small.jpg        | 45.3557 | 9.857   | 2.247   |
| case46Rmlo small.jpg       | 64.1948 | 15.0134 | 5.7276  |
| case49Lcc small.jpg        | 48.8262 | 8.2752  | 5.1027  |
| case49Lmlo small.jpg       | 54.0595 | 11.916  | 5.8115  |
| case49Rcc small.jpg        | 45.2983 | 6.6021  | 1.3718  |
| case49Rmlo small.jpg       | 53.707  | 10.1534 | 6.0692  |
| case6Lcc ans small.jpg     | 51.9116 | 9.534   | 4.4319  |
| case6Lmlo ans small.jpg    | 71.6962 | 7.5438  | 5.209   |
| case7Lcc small ans.jpg     | 47.1428 | 15.318  | 4.9046  |
| case7Lmlo small ans.jpg    | 54.3043 | 18.2928 | 6.1598  |
| case7Rcc small ans.jpg     | 61.8618 | 24.6397 | 8.4048  |
| case7Rmlo small ans.jpg    | 64.2621 | 25.1122 | 8.5309  |
| case17rmlo.jpg             | 83.3709 | 26.0836 | 9.1561  |
| case19Rcc.jpg              | 77.103  | 15.7121 | 10.1174 |
| case2.jpg                  | 59.7236 | 21.5586 | 7.1512  |
| case21Lcc.jpg              | 86.6151 | 22.2302 | 5.8418  |

|                |         |         |         |
|----------------|---------|---------|---------|
| case21Lmlo.jpg | 84.494  | 29.7057 | 9.6538  |
| case24Lcc.jpg  | 49.463  | 4.2826  | 2.2734  |
| case24Lmlo.jpg | 98.2403 | 30.589  | 16.1846 |
| case26Lcc.jpg  | 59.6111 | 8.4871  | 3.5104  |
| case26Lmlo.jpg | 51.1466 | 8.8898  | 3.3563  |
| case27Rcc.jpg  | 55.5713 | 18.5925 | 9.3226  |
| case27Rmlo.jpg | 61.8729 | 23.6434 | 12.4143 |
| case2lmlo.jpg  | 59.7656 | 25.8374 | 8.3435  |
| case36Lcc.jpg  | 48.4189 | 10.5125 | 4.0835  |
| case36Lmlo.jpg | 56.1102 | 9.602   | 3.628   |
| case39Lcc.jpg  | 73.4592 | 17.5956 | 7.6439  |
| case39Lmlo.jpg | 73.4274 | 23.692  | 6.8882  |
| case44Rcc.jpg  | 56.5989 | 16.0424 | 7.76    |
| case44Rmlo.jpg | 75.1761 | 23.5834 | 10.5149 |
| case5lcc.jpg   | 65.1534 | 14.314  | 4.5944  |
| case5lmlo.jpg  | 28.2253 | 4.6292  | 1.4759  |

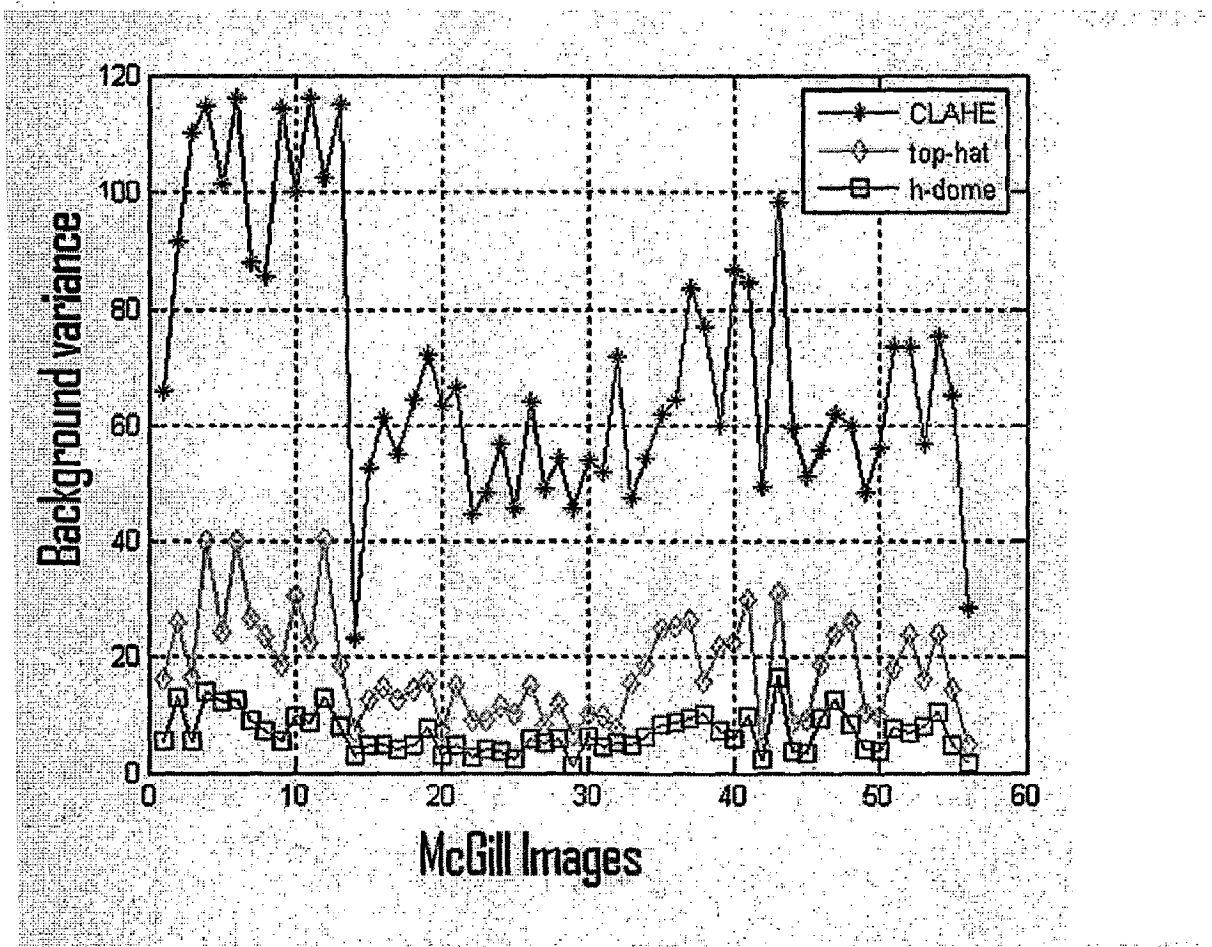


Figure 3.3: Comparison of Background Variance(BV) for different enhancement methods as applied to some ROIs of digital mam mograms containing microcalcifications obtained from McGill database.

Table 3.4 Comparison of Detail variance(DV) / Background variance (BV) for different enhancement techniques applied on ROIs of mammograms containing microcalcifications.

| Name of file               | CLAHE  | tophat  | hdome   |
|----------------------------|--------|---------|---------|
| case11Lcc small ans.jpg    | 1.8219 | 5.4592  | 7.3312  |
| case11Lmlo small ans.jpg   | 1.3823 | 3.4619  | 3.4094  |
| case11Rcc small ans.jpg    | 1.0263 | 5.4553  | 13.5619 |
| case11Rmlo small ans.jpg   | 1.1561 | 2.4988  | 3.0172  |
| case11 1Lcc small ans.jpg  | 1.275  | 4.0008  | 3.0012  |
| case11 1Rmlo small ans.jpg | 1.1717 | 2.3538  | 4.6016  |
| case15Lcc small ans.jpg    | 1.4191 | 4.234   | 9.3933  |
| case15Lmlo small ans.jpg   | 1.3718 | 3.447   | 6.0906  |
| case15Rcc small ans.jpg    | 1.1276 | 5.0099  | 8.5412  |
| case15Rmlo small ans.jpg   | 1.1266 | 2.8284  | 3.7901  |
| case15 1Lcc small ans.jpg  | 1.0896 | 4.2247  | 6.9851  |
| case15 1Lmlo small ans.jpg | 1.2462 | 2.8486  | 6.0572  |
| case15 1Rmlo small ans.jpg | 1.0904 | 3.9199  | 5.6203  |
| case22Lcc small.jpg        | 5.0901 | 15.8951 | 26.5558 |
| case25Lcc small ans.jpg    | 2.2488 | 7.032   | 5.7807  |
| case25Lmlo small ans.jpg   | 1.8792 | 6.2937  | 5.0638  |
| case25Rcc small ans.jpg    | 2.2262 | 6.8286  | 3.7069  |
| case25Rmlo small ans.jpg   | 1.9423 | 6.2901  | 4.8542  |
| case37Lcc ans small.jpg    | 1.6644 | 5.0736  | 4.8756  |
| case37Lmlo ans small.jpg   | 1.6825 | 6.2607  | 9.2017  |
| case37Rcc ans small.jpg    | 1.4907 | 3.8692  | 3.7868  |
| case37Rmlo ans small.jpg   | 2.2576 | 6.1862  | 8.2456  |
| case46Lcc small.jpg        | 1.8267 | 5.2436  | 4.9238  |
| case46Lmlo small.jpg       | 1.7104 | 5.4614  | 8.6311  |
| case46Rcc small.jpg        | 2.4387 | 10.6359 | 24.573  |
| case46Rmlo small.jpg       | 1.5705 | 4.381   | 6.7943  |
| case49Lcc small.jpg        | 1.701  | 6.0728  | 8.571   |
| case49Lmlo small.jpg       | 2.1554 | 8.074   | 14.8162 |
| case49Rcc small.jpg        | 2.1027 | 13.0858 | 46.8117 |
| case49Rmlo small.jpg       | 2.1847 | 8.8755  | 13.7816 |
| case6Lcc ans small.jpg     | 2.1628 | 6.0444  | 6.3396  |
| case6Lmlo ans small.jpg    | 1.6305 | 5.9265  | 7.5506  |
| case7Lcc small ans.jpg     | 2.4436 | 6.0634  | 4.639   |
| case7Lmlo small ans.jpg    | 2.0491 | 4.7484  | 4.9354  |
| case7Rcc small ans.jpg     | 2.0525 | 4.4474  | 4.5907  |
| case7Rmlo small ans.jpg    | 1.8229 | 3.9807  | 2.6953  |
| case17rmlo.jpg             | 1.315  | 2.5792  | 2.9356  |
| case19Rcc.jpg              | 1.5545 | 4.9215  | 2.8494  |
| case2.jpg                  | 1.9351 | 4.2781  | 7.3153  |
| case21Lcc.jpg              | 1.2078 | 2.8286  | 4.9285  |
| case21Lmlo.jpg             | 1.3743 | 2.4113  | 4.3826  |

|                |        |         |         |
|----------------|--------|---------|---------|
| case24Lcc.jpg  | 2.2081 | 12.1763 | 21.0072 |
| case24Lmlo.jpg | 1.3668 | 3.6487  | 2.9172  |
| case26Lcc.jpg  | 1.8532 | 5.0935  | 4.0189  |
| case26Lmlo.jpg | 1.9207 | 6.3411  | 5.0125  |
| case27Rcc.jpg  | 1.7037 | 3.7338  | 3.9372  |
| case27Rmlo.jpg | 1.5957 | 3.0757  | 2.9419  |
| case2lmlo.jpg  | 1.9093 | 3.7327  | 7.087   |
| case36Lcc.jpg  | 2.4412 | 5.9295  | 4.3299  |
| case36Lmlo.jpg | 1.845  | 4.0641  | 2.0411  |
| case39Lcc.jpg  | 1.2838 | 3.4869  | 2.495   |
| case39Lmlo.jpg | 1.4351 | 2.9511  | 2.6567  |
| case44Rcc.jpg  | 1.9106 | 3.9924  | 3.7832  |
| case44Rmlo.jpg | 1.6324 | 3.5514  | 4.2745  |
| case5lcc.jpg   | 1.6863 | 4.3485  | 6.3741  |
| case5lmlo.jpg  | 4.067  | 23.4492 | 44.4479 |

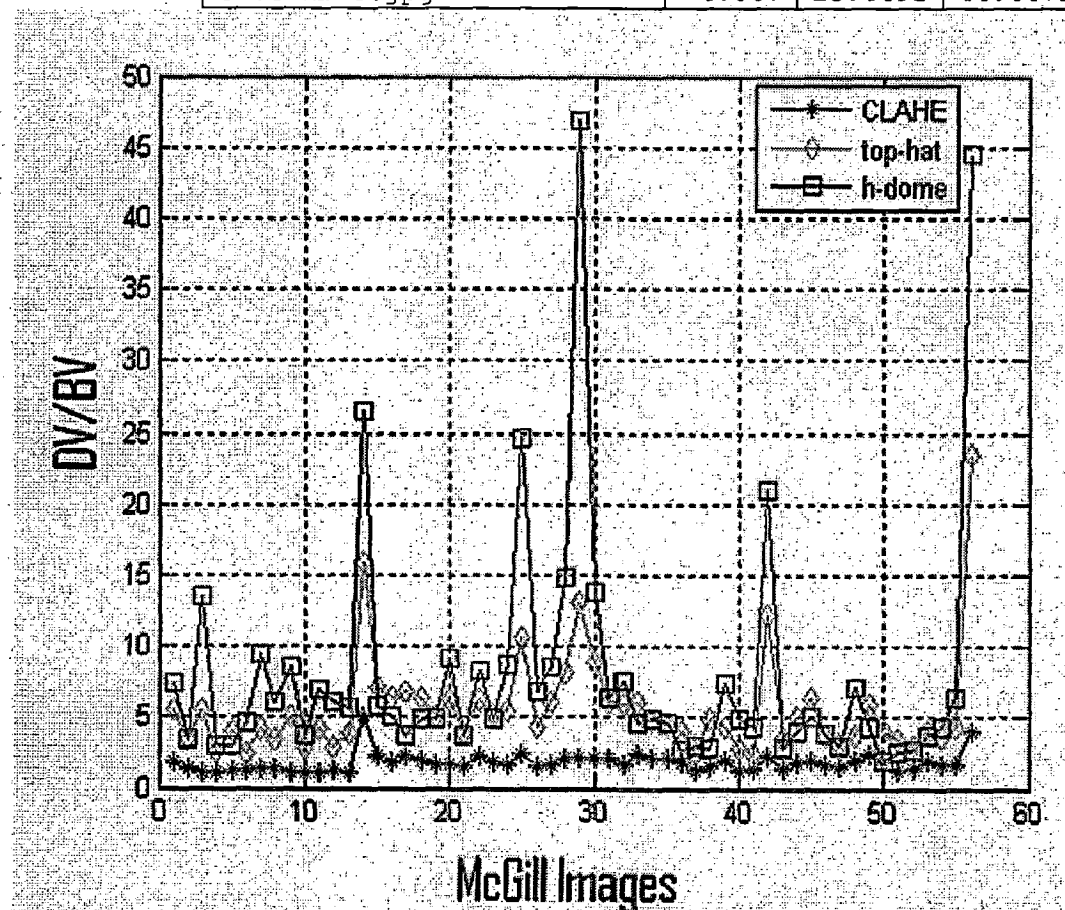


Figure 3.4: Comparison of Detail Variance(DV) / Background Variance(BV) for different enhancement methods as applied to some ROIs of digital mam-mograms containing microcalcifications obtained from McGill database.



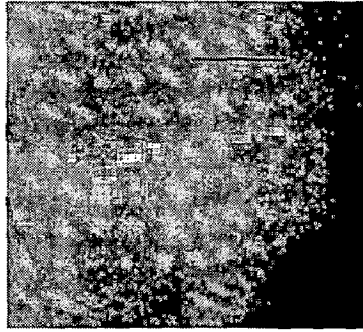


Figure 3.5: Mammogram ROI image:case5lcc.jpg [1]

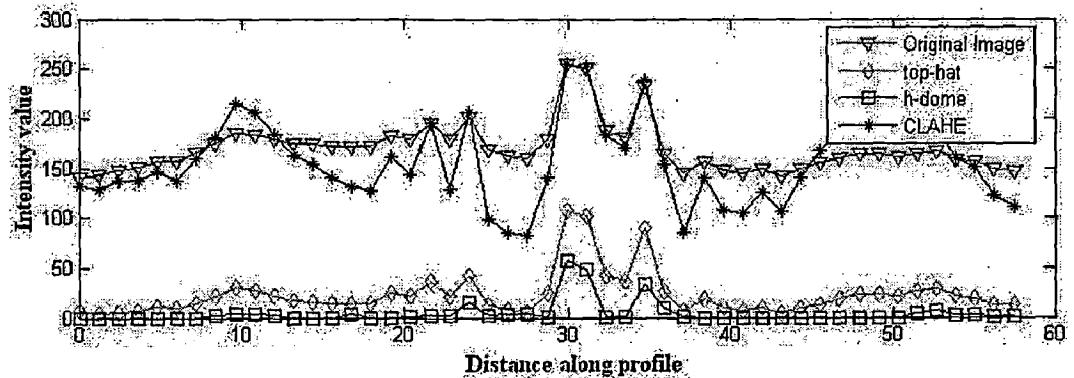


Figure 3.6: Image profiles of case5lcc.jpg [1]

### 3.3 Image Profiles

The effect of enhancement can be visualized using a series of profiles taken across a series of microcalcifications from the ROI in Figure 3.5 . The resulting profiles for the original, CLAHE, Tophat, h-dome are shown in Fig. As clear from figure 3.6 , it has been noticed that images enhanced with h-dome exhibit deep valleys, separating the peaks of the microcalcifications more effectively than the two other enhancement methods. Fig. shows that the h-dome method is only enhancing the microcalcifications but other methods enhances the background also, which is undesirable.

Figure 3.5 below shows a ROI of the mammogram indicating the line along which image profile is taken.

*This study measures the CII & DV/BV of processed results to elucidate the enhancement of an image. In this simulation, it is found that the h-dome transformation can be used to enhance clinically acquired digital mammograms without distortion and can therefore provide more detailed*

*image information. The promising results of this work suggest that the proposed technique can be routinely used and the proposed method extended to computer-assisted diagnosis of breast diseases in a clinical environment.*



# Chapter 4

## Detection of microcalcifications

### 4.1 Introduction

Mammogram image, enhanced by h-dome transformed image, is used as input image for segmentation algorithm. This chapter describes the way to extract microcalcifications from the mammograms.

Entropy thresholding forms another class of algorithms widely used in thresholding. It has also been used in detection of MCC's in mammograms. Minimum cross entropy method applied for segmentation of mammograms for extracting microcalcifications.

### 4.2 Basics of entropy of an image

Shannon defined the entropy of a system as a function of the probability of occurrence of different states of the system. If a system has  $n$  different states with probability of occurrence  $p_i, i=1,2,\dots,n. \sum_{i=1}^n p_i = 1$ . then the gain in information from the occurrence of the event  $i$  is defined as  $\Delta I = -\log_2 p_i$ . The expected value of such a gain in information is defined as the entropy of the system.

Thus the entropy  $H$  of the system is:

$$H = - \sum_{i=1}^n p_i \quad (4.1)$$

Let  $F = [f(x, y)]_{P \times Q}$  where  $f(x, y)$  is gray value at  $(x, y)$ ;  $f(x, y) \in G_L = \{1, 2, 3, \dots, L-1\}$ , the set of gray levels.

$N_i$  be the frequency of gray level  $i$ .  $\sum_{i=1}^{L-1} N_i = N$

Following Shannon's definition of entropy, entropy of the image histogram can be defined as:

$$H = - \sum_{i=0}^{L-1} p_i \log_2 p_i; p_i = N_i/N \quad (4.2)$$

for the image segmentation problem.

Cross Entropy of the system

The cross entropy was proposed by Kullback under the name of directed divergence. The cross entropy measures the information theoretic distance between two distributions.  $p = \{p_1, p_2, \dots, p_N\}$  and  $Q = \{q_1, q_2, \dots, q_N\}$  by

$$D(P, Q) = \sum_1^N q_k \log_2 (q_k/p_k) \quad (4.3)$$

formula can be interpreted as the expectation of the change in the information content when we are using Q instead of P. The minimum cross entropy method can be seen as an extension of the maximum entropy method by setting equal initial estimates for all  $p_i$  when no prior information is available. Cross entropy gives a measure of how close two distributions are.

### 4.3 Algorithm

For a histogram  $h$  defined on the gray level range  $[1, L]$ , the zeroth and the first moments of the foreground and background portions of the thresholded histogram are respectively, [10]

$$m_{0a}(t) = \sum_{i=1}^{t-1} h(i) \quad (4.4)$$

$$m_{0b}(t) = \sum_{i=t}^L h(i) \quad (4.5)$$

$$m_{1a}(t) = \sum_{i=1}^{t-1} i * h(i) \quad (4.6)$$

$$m_{1b}(t) = \sum_{i=t}^L i * h(i) \quad (4.7)$$

The portion's means are defined as

$$\mu_a(t) = \frac{m_{1a}(t)}{m_{0a}(t)} \quad (4.8)$$

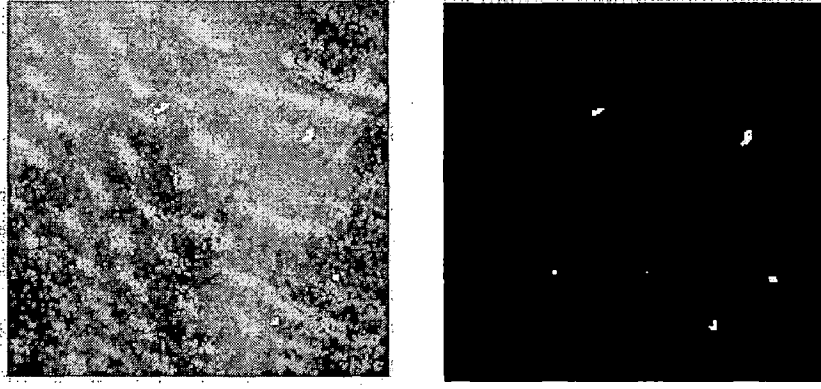


Figure 4.1: Original and segmented output image Test1 from synthetic images

$$\mu_b(t) = \frac{m_{1b}(t)}{m_{0b}(t)} \quad (4.9)$$

The minimum cross entropy [23] method selects the threshold which minimizes the cross entropy of the image and its segmented version. The criterion function is found to be

$$\eta(t) = -m_{1a}(t) * \log(\mu_a(t)) - m_{1b}(t) * \log(\mu_b(t)) \quad (4.10)$$

The optimal threshold  $t$  is given by the minimizer of Eq.4.10.,

$$t_{opt} = argmin_t \eta(t) \quad (4.11)$$

The calculation of the optimal threshold involves the evaluation of  $\eta(t)$  for all possible threshold values. Minimum cross entropy means distributions of original image and segmented image is maintained as close as possible after thresholding.

### 4.3.1 Results and discussions

Segmentation algorithm is applied on test synthetic images as described in section 2.5.1.

As shown in figure 4.1, original image Test1 with 5 simulated microcalcifications, after applying proposed segmentation algorithm proposed we get a segmented image with 5 microcalcifications as in original without distorting shapes and sizes of microcalcifications.

As shown in figure 4.2, original image Test2 with 13 microcalcifications, after applying proposed segmentation algorithm we get segmented image with 13 microcalcifications with some small dark details added in microcalcifications. Which is taken

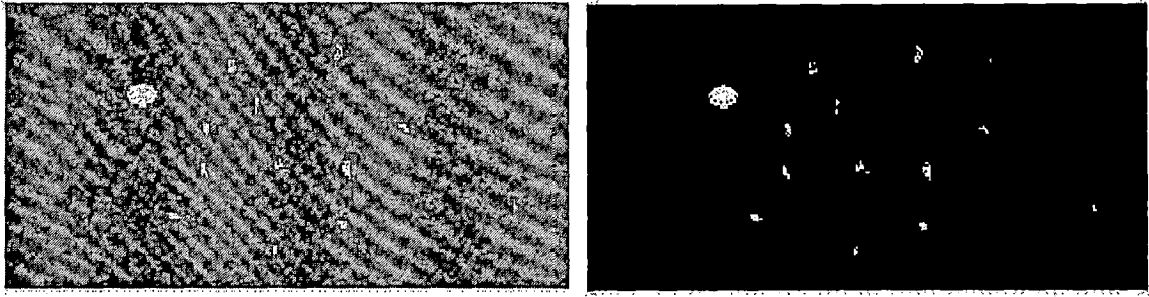


Figure 4.2: Original and segmented output image Test2 from synthetic images

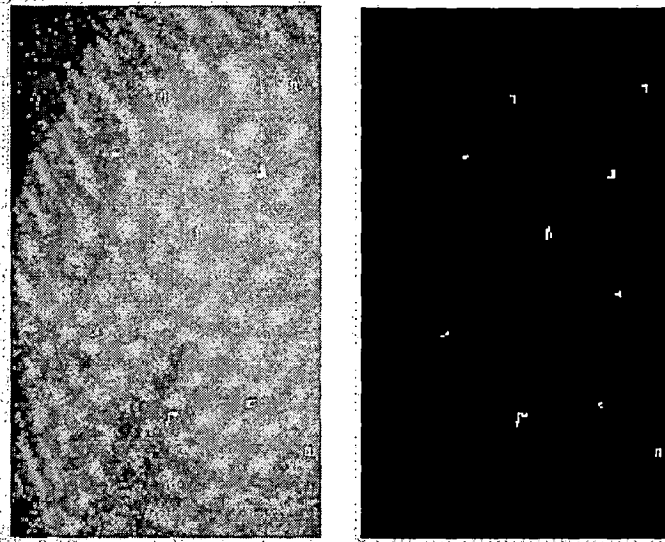


Figure 4.3: Original and segmented output image Test3 from synthetic images

care by object reduction techniques stated in next chapter in section 5.2.

As shown in figure 4.3, original image with 10 simulated microcalcifications on it, after applying proposed segmentation algorithm, we get segmented image with all 10 microcalcifications without noticeable distortion in sizes and shapes of microcalcifications.

After testing this algorithms on all of our generated synthetic images, we found that most of the cases are segmenting all microcalcifications in segmented images as in original image without any noticeable loss.

After getting confidence on synthetic images, we applied our proposed segmentation algorithm on some typical cases from McGill Database which is described in detail in section 2.5.2.

Figure 4.4, containing calcifications of having Pleomorphic/Heterogeneous characteristics (malignant case). Which needs further evaluation as per [1]. Figure clearly

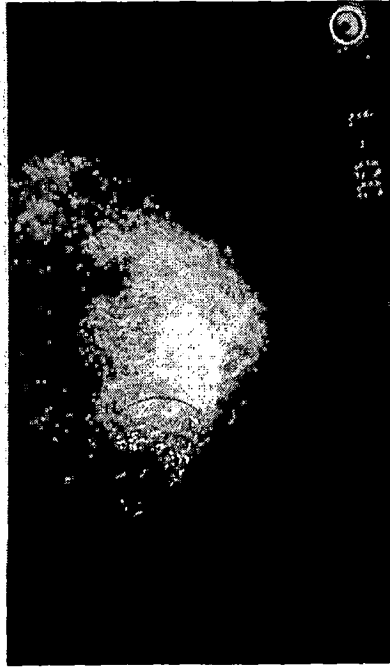


Figure 4.4: Original Image case2.jpg(McGill Database)



Figure 4.5: Cropped ,segmented,superimposed output image(McGill Database)

indicates that it extracts microcalcification which are small bright spots present in the image. But it is noticeable that along with microcalcifications it is abstracting some of the regions with similar intensity values in background. In object reduction step we have taken care of this unwanted parts. We may call this as false positives, but one thing is noticeable and most important is that it is not missing any true positive microcalcifications. This is validated by expert radiologist by taking feedback from him.

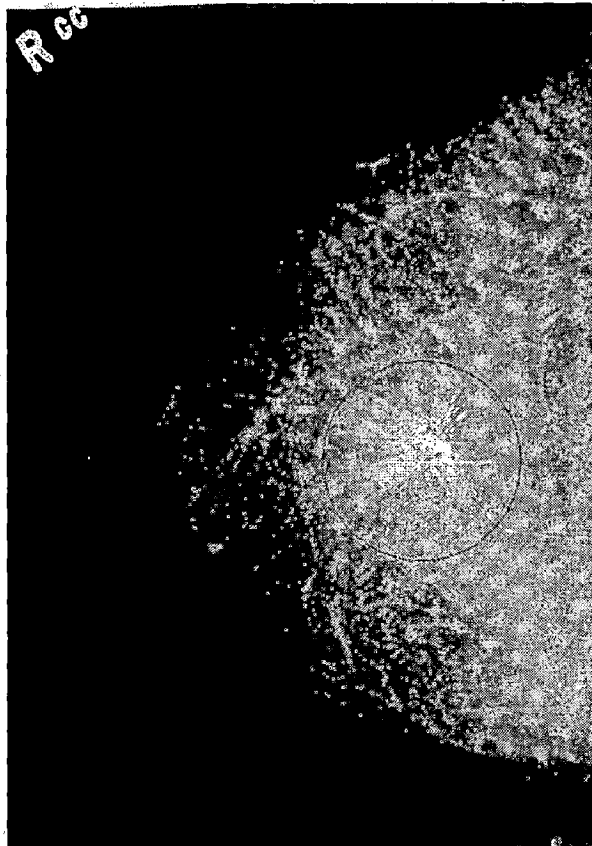


Figure 4.6: Original Image case19rcc.jpg(McGill Database)





Figure 4.7: Cropped, segmented output image and superimposed image case19rcc.jpg (Mcgill Database).

As shown in figure 4.6, it contains pleomorphic microcalcifications as per [1]. In this case microcalcifications are spreading through ducts, after segmenting we observe similar response as in earlier case.

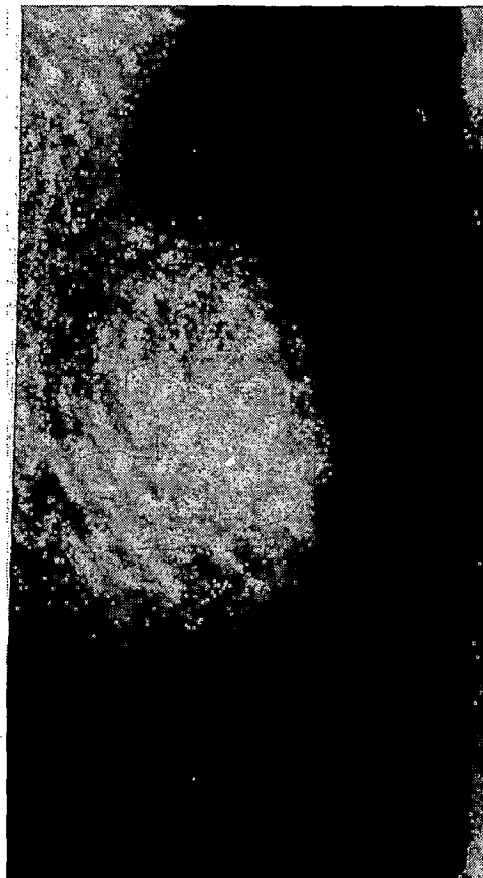


Figure 4.8: Original Image case36lmlo.jpg (McGill Database)

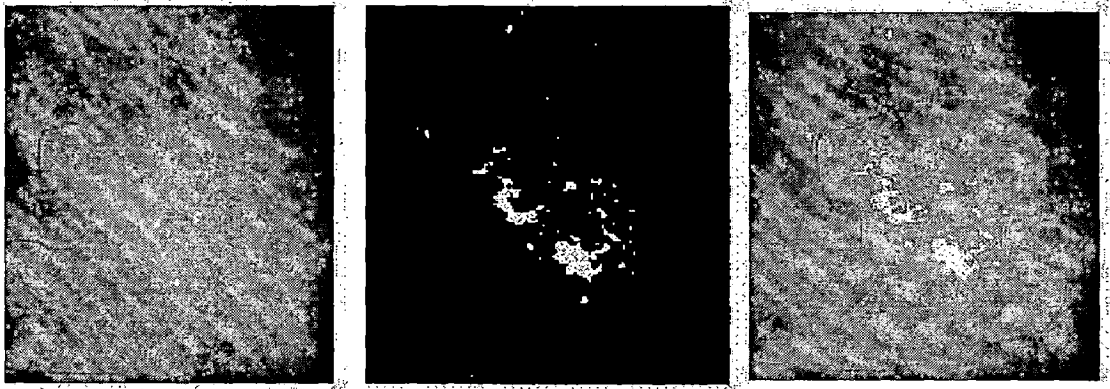


Figure 4.9: Cropped, segmented output image and superimposed image case36lmlo.jpg (McGill Database).

As shown in figure 4.8, it contains Punctuate , Indistinct or Amorphous calcifications which is most difficult to detect for doctors. But as seen from the segmented image, we can clearly see that microcalcifications are present and extracted from the image.

*From the above results and discussion on test and database images, it can be concluded that minimum cross entropy thresholding method gave good results as verified and validated by radiologist response.*

# Chapter 5

## Diagnostic features extraction

### 5.1 Introduction

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 abnormal and normal regions of mammograms, there is a need to present all the available information that exists in mammograms to the diagnostic system so that it can easily discriminate between the abnormal and the normal 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. Therefore, reliable feature vectors should be considered that reduce the amount of irrelevant information thus producing robust mammographic descriptors of compact size.

In this chapter different diagnostic information based features of mammogram are extracted such as shape, texture.

### 5.2 Labelling & object reduction

As in last chapter, we obtained segmented image from original image. Next part is giving labels to different regions to analyse them.

This is done for gaining the ability to handle every single microcalcification. The labeling can be done using an efficient method that labels the microcalcifications in the image in only two passes. In the first pass, starting at the upper left corner of the image, the labeling pixel  $X(i, j)$  starts moving from left to right scanning each

row of the image matrix. By reaching every non-zero value in the image a label is assigned to the pixel and this label is propagated to its right and below neighbors in an 8-connectivity manner. The information about the labels assigned to every pixel and also the pixels above and to the left of every pixel is stored and using this information, in the second pass every label is translated to its equivalent class, which represents the label of the connected components that form every individual shape. In the final result, every microcalcification is assigned a label that is used for identifying it for further processing. The largest label represents the number of microcalcifications on the mammogram [25].

In order to evaluate the characteristics and the location of every object. Since the mammograms are digitized to a spatial resolution of 0.05mm pixel size, we can discard objects smaller than 0.1 mm. Also, objects bigger than 2 cm in diameter are discarded. As a second object reduction step, objects not located within a 1 cm radius region of another object, are also discarded, since malignant microcalcifications are typically clustered..

## 5.3 Feature Extraction

### 5.3.1 Shape based features

One of the characteristics of a microcalcification, which can be a sign of their malignancy, is their sizes. We have used the number of pixels in each microcalcification as a measure of its size.

One of the shape features that has proven to be a good measure for classifying microcalcifications by their shape is compactness.

Compactness (C) [25] is defined as the ratio of the squared perimeter (P) to the area (A), i.e.,

$$C = \frac{P^2}{A} \quad (5.1)$$

Compactness represents the roughness of an object's boundary relative to its area. The smallest value of compactness is 12.56, which is for circle. As circle deviates towards a more complicated shape, compactness becomes larger.

The various shape features extracted from each region are:

1. Area
2. Eccentricity

3. Major axis length
4. Minor axis length
5. Compactness

### 5.3.2 Texture based features

Texture information is used in a wide range of applications including natural scene, remotely sensed data, and biomedical modalities. The perception of texture is believed to play an important role in human visual system for recognition and interpretation. Texture features have been utilized for many medical image applications, including mammography, and have proven to be useful in discriminating texture classes in X-ray mammography, as well as in other modalities for breast cancer detection. So here we have extracted texture information from first as well as second order statistics of intensity histogram of mammogram.

#### First order histogram based features

Statistical properties of intensity histogram is used for texture analysis. Following properties of histogram is extracted for texture analysis.

1. Mean -

$$m = \sum_{i=0}^{L-1} z_i * p(z_i) \quad (5.2)$$

gives a measure of average intensity.

2. Standard Deviation -

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)} \quad (5.3)$$

a measure of average contrast of image.

3. Smoothness -

$$R = 1 - [1/(1 + \sigma^2)] \quad (5.4)$$

measures relative smoothness of the intensity in region.

4. Skewness -

$$\mu_3 = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)} \quad (5.5)$$

measures skewness of a histogram.

5. Uniformity -

$$U = \sum_{i=0}^{L-1} p^2(z_i) \quad (5.6)$$

measures uniformity of the image.

6. Entropy -

$$E = - \sum_{i=0}^{L-1} p(z_i) * \log_2 p(z_i) \quad (5.7)$$

a measure of randomness of intensity in the image histogram

where  $z_i$  = random variable indicating intensity

$p(z_i)$  = intensity histogram of an image

### Co-occurrence matrix based features

In this work, like most published work on microcalcifications classification used co-occurrence matrices to describe textural properties. The spatial gray level dependence (SGLD) matrix or co-occurrence matrix (GLCM) gives us an estimation of the second order joint probability density function (i.e. The probability that 2 pixels which are located with an intersample distance  $d$  and direction  $\theta$  having a gray level  $i$  &  $j$  respectively). In other words, the co-occurrence matrix is a tabulation of how often different combinations of pixel brightness values (gray levels) occur in an image.

$$C(i, j) = \sum_{p=1}^n \left\{ \sum_{q=1}^m \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta(x), q + \Delta(y)) = j \\ 0, & \text{otherwise} \end{cases} \right. \quad (5.8)$$

$C(i, j)$  = entry of GLCM for intensity pair  $i$  and  $j$  at distance depending on value of  $\Delta$  specified.

$m, n$  = dimensions of original matrix

let

$\theta$  = direction  $\theta$  having a gray level  $i$  &  $j$  respectively

$d$  = intersample distance

As shown in figure 5.1 arrows shows direction of computing of GLCM in angle 0, 45, 90, 135 degrees. Distance is no. of pixels from the considered central pixel shown by black box.

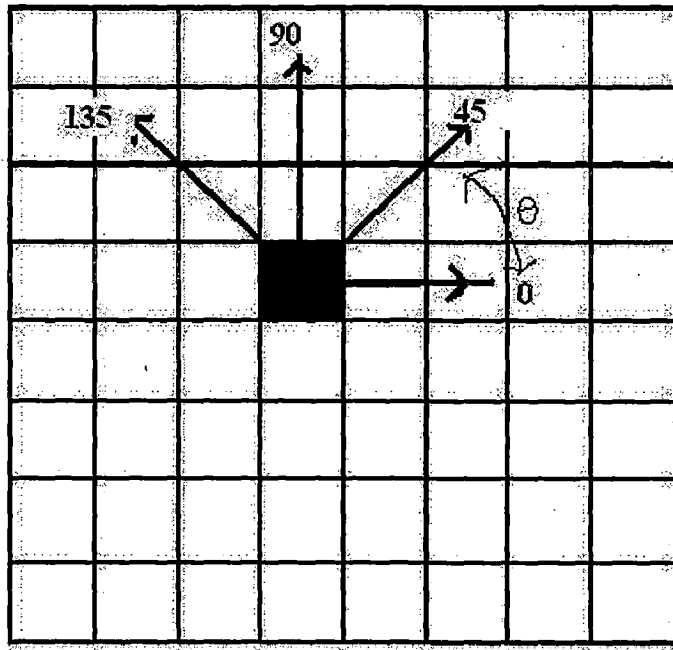


Figure 5.1: Co-occurrence matrix-angle  $\theta$  and distance  $d$

The parameter values are  $d = 1$ , and  $\theta = 0, 45, 90$  and  $135$  degree. Haralick [11] proposed following features to be extracted from the occurrence matrix to get textural information about an image. Following are various texture features extracted for each  $(d, \theta)$  pair.

1. Angular second moment
2. Contrast
3. Correlation
4. Difference entropy
5. Difference variance
6. Entropy
7. Inverse difference moment -
8. Sum average
9. Sum entropy
10. Sum variance
11. Variance

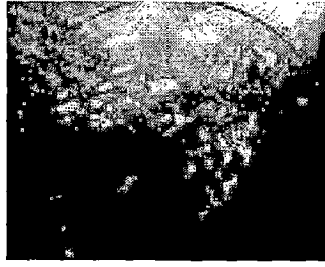


Figure 5.2: Original image: case2.jpg (McGill Database)

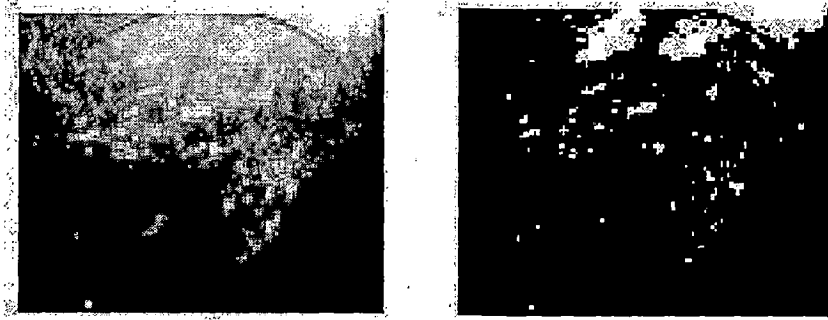


Figure 5.3: Cropped, segmented image of case2.jpg

Difference variance and entropy are measured from difference second-order histogram statistics. The difference second-order histogram represents the probability of occurrence of differences,  $I(p + \Delta(x), q + \Delta(y)) - I(p, q) = i$ , in the gray-level values of two pixels separated by a specific distance vector  $d$ .

The angular second moment gives a strong measure of uniformity. Higher nonuniformity values provide evidence of higher structural variations. This measure provides evidence of how sharp the structural variations in the image are. The correlation feature is a measure of gray-level linear dependency of the image. The entropy computed from the second order histogram provides a measure of nonuniformity. High values of uniformity measures will indicate less structural variations while lower values can be interpreted as indicating a higher probability of microcalcification related structures. The inverse difference moment is a measure of local homogeneity [24].

### 5.3.3 Results and discussions

As seen from last chapter, we transformed original image to cropped image first and then by applying enhancement & segmentation algorithms stated in last chapters, transformed cropped image into segmented image.

As shown in figures 5.2 & 5.3, we get binary image showing microcalcifications. Now next task is to extract each microcalcification from binary image for further analysis.



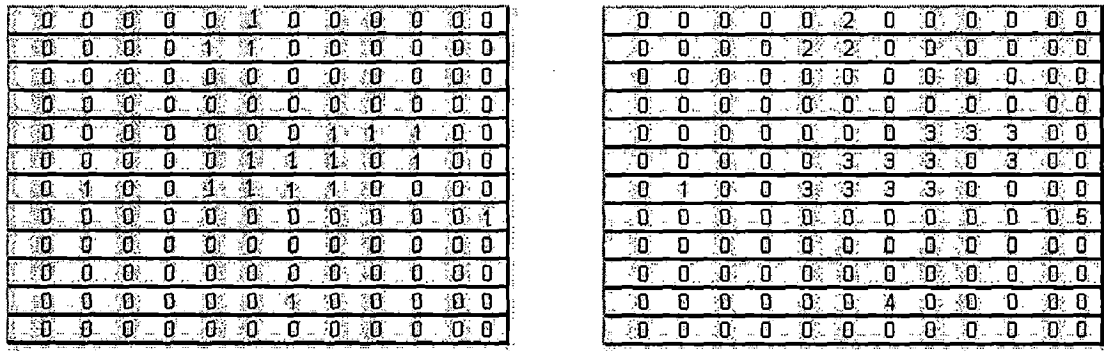


Figure 5.4: Binary image matrix & labeled image matrix

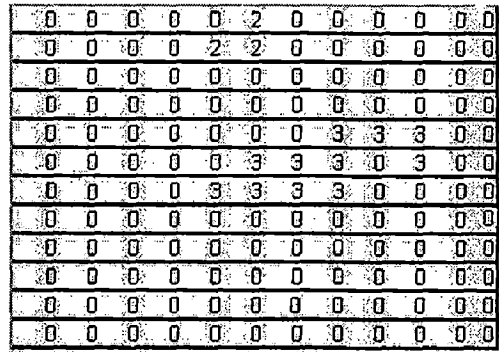


Figure 5.5: Labeled image matrix after object reduction

### Labeling

As stated in section 5.2 after applying labeling procedure we get labeled microcalcifications in image which is explained in following figures.

As shown in figure 5.4 , it can be clearly seen this binary image contains five objects. After labeling operation it gives labeled image. In labeled image we can see we obtained five different classes with numbering.

### Object reduction

Since the mammograms are digitized to a spatial resolution of 0.05 mm pixel size, we can discard objects smaller than 0.1 mm. Also, objects bigger than 2 cm in diameter are discarded. So after applying object reduction steps to above labeled figures, it can be seen that class labels which are having less number of pixels than 2 are discarded. So we get image matrix after object reduction step as follows.

It can be clearly seen that object number 1,4,5 class as having only one pixel eliminated from the labeled image.

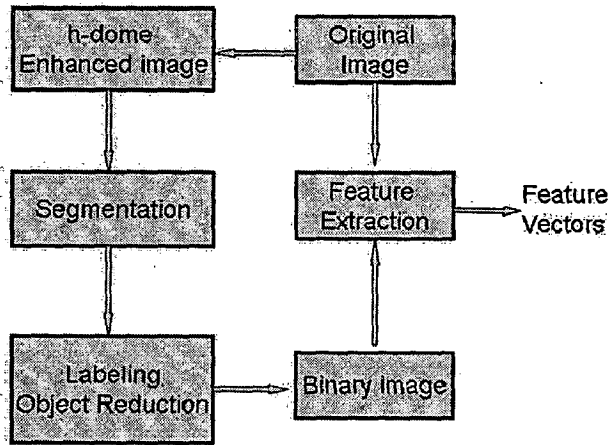


Figure 5.6: Schematic block diagram of feature extraction process

### Features extraction

The binary image obtained after labeling and object reduction, is used only for the extraction of the exact location and area of every object. However, the original image is used for feature extraction, using the binary image as a mask for each object.

Figure 5.6 shows systematic flow of different operations used for feature extraction. Here original image is cropped image as we are interested in region of interest marked by clinician. After enhancement as explained in section 2.4, then applying minimum cross entropy thresholding algorithm for segmenting image into binary image with microcalcifications. After that labeling and object reduction removes objects of size which are of no interest in microcalcifications point of view and labeling remaining for further analysis.

Resulted binary image is used as mask to extract region representing microcalcifications from original image. Finally we extracted different shape and texture based features from those regions.

Features extracted are enlisted as follows:

a) Shape based features

1. Area
2. Eccentricity
3. Major axis length
4. Minor axis length
5. Compactness

b) First order intensity histogram based features

1. Mean
2. Standard deviation
3. Smoothness
4. Skewness
5. Uniformity
6. Entropy

c) Co-occurrence matrix based features

1. Angular second moment
2. Contrast
3. Correlation
4. Difference entropy
5. Difference variance
6. Entropy
7. Inverse difference moment -
8. Sum average
9. Sum entropy
10. Sum variance
11. Variance

In co-occurrence matrix we calculated above listed properties for 0, 45, 90, 135 degrees. Thus overall we get 55 feature vectors.

All these features calculated on one sample image from our McGill Database is shown in Table 5.1, Table 5.2 & Table 5.3.

We have chosen  $d = 1$  as for  $d$  greater than 1 is not showing any useful information in our case.

The next step is to incorporate the features obtained for different directions (angles 0, 45, 90, 135) into one summary feature for each offset distance. This is accomplished because different directions should be treated the same for microcalcification feature extraction. For deriving these summary features, there are two possibilities: averaging the co-occurrence matrices over the angles and then calculating the features for each distance; or calculating the features for each of the four angles and then averaging the results. We have done with averaging the co-occurrence matrices over the angles. Minimum and maximum value of these summary vector is also calculated.

In Table 5.1, Table 5.2, Table 5.3 first column contains the sample image name from which features are extracted. In our case we have enlisted for case2.jpg image from our standard database of mammograms i.e. McGill University Database. Region in the second column in this tables is the region which is labeled and remained after object reduction step. In our case distance between central pixel and other is taken as  $d=1$  for calculation of co-occurrence matrix as explained in section 5.3.2.

Table 5.1 enlists all first order statistics based features i.e, mean, standard deviation, smoothness, skewness, uniformity, entropy.

Table 5.2 enlists one of the feature extracted from co-occurrence matrix, here i.e. Angular Second Moment (ASM), which is calculated for four different values of  $\theta$ . As explained earlier max, min and average values are calculated to make these features independent of direction.

Table 5.3 enlists the features based on shape of microcalcifications i.e. area, Major axis length, Minor axis length, eccentricity, perimeter & compactness as detailed in section 5.3.1.

**Table 5.1: Sample feature calculation table -- Angular Second moment based on Co-occurrence matrix with distance=1**

| Nameoffile     | region | d | 0 deg  | 45 deg | 90 deg | 135 deg | max_ASM_d1 | min_ASM_d1 | avg_ASM_d1 | range  |
|----------------|--------|---|--------|--------|--------|---------|------------|------------|------------|--------|
| case17rml0.jpg | 1      | 1 | 0      | 0      | 0.3333 | 0       | 0.3333     | 0          | 0.083325   | 0.3333 |
| case17rml0.jpg | 2      | 1 | 0.0125 | 0.0133 | 0.0124 | 0.0137  | 0.0137     | 0.0124     | 0.012975   | 0.0013 |
| case17rml0.jpg | 3      | 1 | 0.125  | 0.1667 | 0.1111 | 0.1667  | 0.1667     | 0.1111     | 0.142375   | 0.0556 |
| case17rml0.jpg | 4      | 1 | 0.1111 | 0.1667 | 0.125  | 0.1667  | 0.1667     | 0.1111     | 0.142375   | 0.0556 |
| case17rml0.jpg | 5      | 1 | 0      | 0      | 0.3333 | 0       | 0.3333     | 0          | 0.083325   | 0.3333 |
| case17rml0.jpg | 6      | 1 | 0.5    | 1      | 0.5    | 1       | 1          | 0.5        | 0.75       | 0.5    |
| case17rml0.jpg | 7      | 1 | 0.1667 | 0.3333 | 0.25   | 0.3333  | 0.3333     | 0.1667     | 0.270825   | 0.1666 |
| case17rml0.jpg | 9      | 1 | 0.0128 | 0.0139 | 0.013  | 0.0154  | 0.0154     | 0.0128     | 0.013775   | 0.0026 |
| case17rml0.jpg | 10     | 1 | 0.1    | 0.125  | 0.0833 | 0.125   | 0.125      | 0.0833     | 0.108325   | 0.0417 |
| case17rml0.jpg | 11     | 1 | 0.07   | 0.0667 | 0.0679 | 0.0667  | 0.07       | 0.0667     | 0.067825   | 0.0033 |
| case17rml0.jpg | 13     | 1 | 0.5    | 1      | 0.5    | 1       | 1          | 0.5        | 0.75       | 0.5    |
| case17rml0.jpg | 14     | 1 | 0.1    | 0.125  | 0.0833 | 0.125   | 0.125      | 0.0833     | 0.108325   | 0.0417 |
| case17rml0.jpg | 15     | 1 | 0.0238 | 0.0293 | 0.0238 | 0.0293  | 0.0293     | 0.0238     | 0.02655    | 0.0055 |
| case17rml0.jpg | 17     | 1 | 0.5    | 1      | 0.5    | 1       | 1          | 0.5        | 0.75       | 0.5    |
| case17rml0.jpg | 18     | 1 | 0.3333 | 0.5    | 0.25   | 0.5     | 0.5        | 0.25       | 0.395825   | 0.25   |
| case17rml0.jpg | 22     | 1 | 0.25   | 0.5    | 0.3333 | 0.5     | 0.5        | 0.25       | 0.395825   | 0.25   |
| case17rml0.jpg | 23     | 1 | 0.1111 | 0.1667 | 0.125  | 0.1667  | 0.1667     | 0.1111     | 0.142375   | 0.0556 |
| case17rml0.jpg | 24     | 1 | 0.5    | 1      | 0.5    | 1       | 1          | 0.5        | 0.75       | 0.5    |
| case17rml0.jpg | 26     | 1 | 0.3333 | 0.5    | 0.25   | 0.5     | 0.5        | 0.25       | 0.395825   | 0.25   |
| case17rml0.jpg | 28     | 1 | 0.0625 | 0.0972 | 0.0667 | 0.0833  | 0.0972     | 0.0625     | 0.077425   | 0.0347 |
| case17rml0.jpg | 31     | 1 | 0.0833 | 0.1667 | 0.1429 | 0.1667  | 0.1667     | 0.0833     | 0.1399     | 0.0834 |
| case17rml0.jpg | 35     | 1 | 0.3333 | 0.5    | 0.25   | 0.5     | 0.5        | 0.25       | 0.395825   | 0.25   |
| case17rml0.jpg | 38     | 1 | 0.0227 | 0.0303 | 0.0309 | 0.034   | 0.034      | 0.0227     | 0.029475   | 0.0113 |
| case17rml0.jpg | 39     | 1 | 0.0269 | 0.0321 | 0.0293 | 0.0321  | 0.0321     | 0.0269     | 0.0301     | 0.0052 |
| case17rml0.jpg | 42     | 1 | 0.5    | 0      | 0      | 0       | 0.5        | 0          | 0.125      | 0.5    |
| case17rml0.jpg | 49     | 1 | 0.0057 | 0.006  | 0.0058 | 0.006   | 0.006      | 0.0057     | 0.005875   | 0.0003 |

Table 5.2: Sample feature calculation table – Intensity Histogram Based features of image case17rmlo.jpg

| Nameoffile     | region | mean    | std dev | smoothness | skewness | uniformity | entropy |
|----------------|--------|---------|---------|------------|----------|------------|---------|
| case17rmlo.jpg | 1      | 228     | 2.5495  | 0.0001     | 0        | 0.25       | 2       |
| case17rmlo.jpg | 2      | 229.677 | 15.4519 | 0.0037     | -0.0622  | 0.0334     | 5.1456  |
| case17rmlo.jpg | 3      | 238.167 | 9.822   | 0.0015     | -0.0207  | 0.1111     | 3.2516  |
| case17rmlo.jpg | 4      | 228.833 | 8.153   | 0.001      | -0.01    | 0.0972     | 3.4183  |
| case17rmlo.jpg | 5      | 230     | 2.2361  | 0.0001     | 0        | 0.25       | 2       |
| case17rmlo.jpg | 6      | 217.25  | 1.6394  | 0          | 0.0001   | 0.375      | 1.5     |
| case17rmlo.jpg | 7      | 226.875 | 2.2604  | 0.0001     | -0.0001  | 0.1563     | 2.75    |
| case17rmlo.jpg | 9      | 233.211 | 18.4026 | 0.0052     | -0.2197  | 0.0316     | 5.1362  |
| case17rmlo.jpg | 10     | 223.8   | 7.4762  | 0.0009     | -0.0012  | 0.1022     | 3.4566  |
| case17rmlo.jpg | 11     | 223.333 | 4.6248  | 0.0003     | -0.0009  | 0.1111     | 3.4346  |
| case17rmlo.jpg | 13     | 227.5   | 2.5981  | 0.0001     | -0.0001  | 0.375      | 1.5     |
| case17rmlo.jpg | 14     | 225.2   | 6.5747  | 0.0007     | -0.0045  | 0.0933     | 3.5069  |
| case17rmlo.jpg | 15     | 224.184 | 12.2302 | 0.0023     | -0.0403  | 0.0437     | 4.6805  |
| case17rmlo.jpg | 17     | 213.5   | 5.5     | 0.0005     | 0.0012   | 0.25       | 2       |
| case17rmlo.jpg | 18     | 196     | 24.097  | 0.0089     | 0        | 0.1667     | 2.585   |
| case17rmlo.jpg | 22     | 215     | 9.5568  | 0.0014     | -0.0041  | 0.1667     | 2.585   |
| case17rmlo.jpg | 23     | 220.5   | 6.5     | 0.0006     | -0.0015  | 0.0972     | 3.4183  |
| case17rmlo.jpg | 24     | 211.75  | 4.3229  | 0.0003     | -0.0006  | 0.25       | 2       |
| case17rmlo.jpg | 26     | 213     | 9.2916  | 0.0013     | -0.0059  | 0.2222     | 2.2516  |
| case17rmlo.jpg | 28     | 209.7   | 7.963   | 0.001      | -0.0019  | 0.08       | 3.7219  |
| case17rmlo.jpg | 31     | 185.786 | 9.8356  | 0.0015     | -0.0046  | 0.0816     | 3.6645  |
| case17rmlo.jpg | 35     | 216.667 | 2.7487  | 0.0001     | 0        | 0.2222     | 2.2516  |
| case17rmlo.jpg | 38     | 224.208 | 7.3171  | 0.0008     | -0.0029  | 0.0582     | 4.3082  |
| case17rmlo.jpg | 39     | 218.708 | 12.3659 | 0.0023     | -0.0456  | 0.0521     | 4.4728  |
| case17rmlo.jpg | 42     | 244     | 3.559   | 0.0002     | 0.0005   | 0.3333     | 1.585   |
| case17rmlo.jpg | 49     | 222.624 | 21.4702 | 0.007      | -0.1332  | 0.0222     | 5.7902  |

Table 5.3: Sample feature calculation table- Shape based Features of image case17rmlo.jpg

| Nameoffile     | region | area | majorAL | minorAL | eccentricity | perimeter | compactness |
|----------------|--------|------|---------|---------|--------------|-----------|-------------|
| case17rmlo.jpg | 1      | 4    | 4.6188  | 1.1547  | 0.9682       | 6         | 9           |
| case17rmlo.jpg | 2      | 27   | 14.2925 | 5.4759  | 0.9237       | 37.5563   | 52.23984    |
| case17rmlo.jpg | 3      | 9    | 4.0537  | 3.0245  | 0.6658       | 8.2426    | 7.548939    |
| case17rmlo.jpg | 4      | 9    | 4.3897  | 3.2033  | 0.6837       | 10.2426   | 11.65676    |
| case17rmlo.jpg | 5      | 4    | 4.6188  | 1.1547  | 0.9682       | 6         | 9           |
| case17rmlo.jpg | 6      | 4    | 2.3094  | 2.3094  | 0            | 4         | 4           |
| case17rmlo.jpg | 7      | 6    | 4.1633  | 2       | 0.8771       | 6.8284    | 7.771174    |
| case17rmlo.jpg | 9      | 43   | 10.0825 | 6.5503  | 0.7602       | 31.2132   | 22.6573     |
| case17rmlo.jpg | 10     | 8    | 6.12    | 2.0525  | 0.9421       | 10.2426   | 13.11386    |
| case17rmlo.jpg | 11     | 13   | 5.8723  | 3.5608  | 0.7952       | 13.6569   | 14.34699    |
| case17rmlo.jpg | 13     | 3    | 2.582   | 1.7638  | 0.7303       | 3.4142    | 3.885587    |
| case17rmlo.jpg | 14     | 9    | 5.7866  | 2.3511  | 0.9137       | 10.2426   | 11.65676    |
| case17rmlo.jpg | 13     | 3    | 2.582   | 1.7638  | 0.7303       | 3.4142    | 3.885587    |
| case17rmlo.jpg | 14     | 9    | 5.7866  | 2.3511  | 0.9137       | 10.2426   | 11.65676    |
| case17rmlo.jpg | 15     | 26   | 7.3992  | 7.1825  | 0.2402       | 28.1421   | 30.46068    |
| case17rmlo.jpg | 17     | 3    | 2.582   | 1.7638  | 0.7303       | 3.4142    | 3.885587    |
| case17rmlo.jpg | 18     | 3    | 3.4641  | 2.2111  | 0.7698       | 5.6569    | 10.66684    |
| case17rmlo.jpg | 22     | 3    | 3.8541  | 1.4264  | 0.929        | 4.8284    | 7.771149    |
| case17rmlo.jpg | 23     | 8    | 3.9259  | 3.0007  | 0.6448       | 8.2426    | 8.492557    |
| case17rmlo.jpg | 24     | 3    | 2.582   | 1.7638  | 0.7303       | 3.4142    | 3.885587    |
| case17rmlo.jpg | 25     | 1    | 1.1547  | 1.1547  | 0            | 0         | 0           |
| case17rmlo.jpg | 26     | 3    | 3.8541  | 1.4264  | 0.929        | 4.8284    | 7.771149    |
| case17rmlo.jpg | 28     | 10   | 6.8035  | 2.5098  | 0.9295       | 12.2426   | 14.98813    |
| case17rmlo.jpg | 31     | 7    | 8.1301  | 1.9581  | 0.9706       | 13.6569   | 26.64442    |
| case17rmlo.jpg | 35     | 3    | 3.8541  | 1.4264  | 0.929        | 4.8284    | 7.771149    |
| case17rmlo.jpg | 38     | 30   | 13.2289 | 3.8804  | 0.956        | 30.4853   | 30.97845    |
| case17rmlo.jpg | 39     | 25   | 13.2753 | 3.9839  | 0.9539       | 33.8995   | 45.96704    |
| case17rmlo.jpg | 42     | 3    | 3.4641  | 1.1547  | 0.9428       | 4         | 5.333333    |
| case17rmlo.jpg | 49     | 68   | 18.9902 | 6.2786  | 0.9438       | 46.5269   | 31.83459    |

For each suspected microcalcification, four  $(d, \theta)$  pairs yields 44 texture features. Then the range and maximum and minimum values are taken as features to be used for classification of microcalcifications. So the number are features derived from co-occurrence matrix are reduced to 33.

*This chapter analysed different regions present on the mammogram. Different features of image calculated to extract diagnostic valuable information from the mammogram. This feature sets are ready to classier for classification of region into benign or malignant case.*



# Chapter 6

## Classification of microcalcifications

### 6.1 Introduction

One of the main aim when implementing a robust classifier for recognizing breast tissue is the selection of the appropriate features that describe and highlight the differences between the abnormal and the normal tissue in an ample way. A great number of features and classification methods have already been used to detect and classify the masses as malignant or benign. Recently support vector machines (SVMs) have been used for classification for mammographic microcalcifications [26]. In the present work, SVM based classifier has been implemented that uses features which result from the shape descriptors extracted from ROI, statistical texture analysis capable of distinguishing between benign and malignant microcalcifications.

### 6.2 Support Vector Machines (SVMs)

Burges [27] has provided an excellent tutorial for Support Vector Machines and explains the key concepts in a very lucid language. The description that follows has been taken from the same. SVM differ radically from comparable approaches such as neural networks: SVM training always finds a global minimum and has simple geometric interpretation. The problem which drove the initial development of SVMs occurs in several guises- the bias variance trade-off, capacity control, over fitting - but the basic idea is the same. For a given learning task with a given amount of training data, the best generalization performance will be achieved if the right balance is struck between the accuracy attained on that particular training set and the "capacity" of the machine. Machine with too much capacity is like a botanist with a photographic

memory who when presented with a new tree concludes that it is not a tree because it has a different number of leaves from anything he has seen before; a machine with too little capacity is like his lazy friend who says that if it is green then it must be a tree. Neither can generalize well. The formulation and exploration of these concepts led to a path breaking development in the theory of statistical learning theory. This theory is the principle of structural risk minimization (SRM) (Vapnik, 1979). It grew out of considerations of under what circumstances, and how quickly the mean of some empirical quantity converges uniformly to the true mean (that which would be calculated from an infinite amount of data), as the number of data points increases. We will briefly try to cover this principle since this forms the backbone of SVMs.

Suppose we have a machine whose task is to learn the mapping  $x_i \rightarrow y_i$ . The machine is actually defined by a set of possible mappings  $x \rightarrow f(x, \alpha)$  where the functions  $f(x, \alpha)$  themselves are labeled by the adjustable Parameter  $\alpha$ . The expectation of the test error for a trained machine is therefore:

$$R(\alpha) = \int \frac{1}{2} |y - f(x, \alpha)| dP(x, y) \quad (6.1)$$

The quantity  $R(\alpha)$  is called the expected risk or just the risk and this is the quantity that we are ultimately interested in. The "empirical risk"  $R_{emp}(\alpha)$  is defined to be just the measured error rate on the training set (for a fixed, finite number of observations):

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^l |y - f(x_i, \alpha)| \quad (6.2)$$

where  $l$  is the number of observations. The quantity  $\frac{1}{2} |y - f(x_i, \alpha)|$  is called the loss. Suppose we define our problem such that the loss can only be either 0 or 1. Now choose some  $\eta$  such that  $0 \leq \eta \leq 1$ . Then for losses taking these values with probability  $1 - \eta$ , the following bound holds (Vapnik, 1995):

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\left( \frac{h((\log 2l/h) + 1) - \log(\eta/4)}{l} \right)} \quad (6.3)$$

where  $h$  is a non-negative integer called the Vapnik Chervonenkis (VC) dimension and is a measure of the notion of capacity above. The second term on the right is called the "VC confidence". It is a monotonically increasing function of  $h$  for any value of  $l$ . Given some selection of learning machines whose empirical risk is zero, one wants to choose that machine whose associated set of functions has minimal VC dimension.

We can now summarize the principle of SRM. The VC confidence term in equation

6.3 depends on the chosen class of functions whereas the empirical risk and actual risk depend on the one particular function chosen by the training procedure. We would like to find that subset of the chosen functions, such that the risk bound for that subset is minimized. One cannot arrange things so that the VC dimension varies smoothly, since it is an integer. Instead introduce a “structure” by dividing the entire class of functions into nested subsets. For each subset we must be able to compute  $h$ , or to get a bound on  $h$  itself. SRM then consists of finding that subset of functions which minimizes the bound on the actual risk. This can be done by simply training a series of machines one for each subset where for a given subset the goal of training is simply to minimize the empirical risk. One then takes that trained machine in the series whose sum of empirical risk and VC confidence is minimal.

### 6.2.1 Linear SVM

Consider the problem of separating the set of training vectors belonging to two separate classes. Suppose there exist a set  $S$  of  $l$  observations. Each observation consists of a pair: a vector is associated with,  $x_i \in R^n, i=1,2,\dots,l$  and the associated truth labels  $y_i \in \{-1, 1\}$  provided by a trusted source. If the two classes are linearly separable, the objective is to find a hyperplane, which separates the all the points with the same labels on the same side of the hyperplane. The points  $x$  which lie on the hyperplane satisfy  $w \cdot x + b = 0$  where  $w$  is normal to the hyperplane,  $|b|/||w||$  is the perpendicular distance from the hyperplane to the origin, and  $||w||$  is the Euclidean norm of  $w$ . Let  $d_+$  ( $d_-$ ) be the shortest distance from the separating hyperplane to the closest positive (negative) example. The “margin” of a separating hyperplane is defined to be  $d_+ + d_-$ . For the linearly separable case, the support vector algorithm simply looks for the separating hyperplane with largest margin. This can be formulated as follows: suppose that all the training data satisfy the following constraints:

$$x_i \cdot w + b \geq +1 \text{ for } y_i = +1 \quad (6.4)$$

$$x_i \cdot w + b \leq -1 \text{ for } y_i = -1 \quad (6.5)$$

These can be combined as:

$$y_i(x_i \cdot w + b) - 1 \geq 0 \text{ for } \forall i \quad (6.6)$$

If the points for which the equality in Eq.6.4 holds are considered then these points

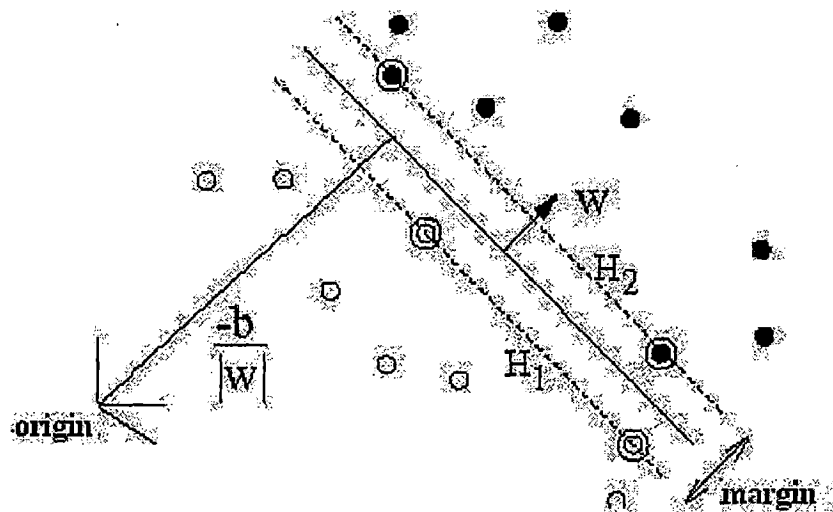


Figure 6.1: Linear separating hyperplanes for the separable case. (The support vectors are circled.) [27]

lie on the hyperplane  $H_1 : x_i w + b = 1$  with normal  $w$  and perpendicular distance from the origin  $|1 - b|/\|w\|$ . Similarly, the points for which the equality 6.5 holds lie on the hyperplane  $H_2 : x_i w + b = -1$ , with normal again  $w$ , and perpendicular distance from the origin  $|-1 - b|/\|w\|$ . Hence  $d_+ = d_- = 1/\|w\|$  and the margin is simply  $2/\|w\|$ . The margin is a measure of generalizability. The larger the margin, the better the generalization. Among the separating hyperplanes, there must be one from which the distance to the closest point is maximal- the optimal separating hyperplane (OSH) and which will maximize the margin sometimes also known as maximum margin hyperplane (MMH). Note that  $H_1$  and  $H_2$  are parallel (they have the same normal) and that no training points fall between them. Thus a pair of hyperplanes can be found which gives the maximum margin by minimizing  $\|w\|$ , subject to constraints in Eq. 6.6.

Mathematically, such hyperplane can be found by minimizing the following cost function, subject to constraints in Eq. 6.6.

$$J(w) = \frac{1}{2} w \cdot w = \frac{1}{2} \|w\|^2 \quad (6.7)$$

Thus the expected solution for a typical two dimensional case would have the form shown in Fig. 6.17. Those training points for which the equality in Eq. 6.6 holds (i.e. those which wind up lying on one of the hyperplanes  $H_1, H_2$ ), and whose removal would change the solution found, are called support vectors; they are indicated in Fig. 6.7 by the extra circles. This is a quadratic programming problem, solved by the Karush-Kuhn-Tucker (KKT) theorem.

## 6.2.2 Lagrangian Formulation of the Problem

According to the property that  $\|w\|^2$  is convex, this can be minimized under the constraints in Eq.6.6 by means of classic method of Lagrange multipliers. There are two reasons for doing this. The first is that the constraints in Eq.6.6 will be replaced by constraints on the Lagrange multipliers themselves, which will be much easier to handle. The second is that in this reformulation of the problem, the training data will only appear (in the actual training and test algorithms) in the form of dot products between vectors.

Hence if  $\alpha_i, i = 1, \dots, l$  are  $l$  non negative Lagrange multipliers associated with constraints in Eq.6.6, the solution to the problem is equivalent to determining the solution of the Wolfe dual problem:

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (6.8)$$

$$\sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0 \quad (6.9)$$

The solution for  $w$  reads

$$w = \sum_i \alpha_i y_i x_i \quad (6.10)$$

Support vector training (for the separable, linear case) therefore amounts to maximizing LD with respect to the  $\alpha_i$ , subject to constraints (Eq. 6.9) and positivity of the  $\alpha_i$ , with solution given by (Eq. 6.10). Notice that there is a Lagrange multiplier for every training point. In the solution, those points for which  $\alpha_i > 0$  are called "support vectors", and lie on one of the hyperplanes  $H_1, H_2$ . These vectors are termed support vectors and they are the only vectors of  $S$  needed to determine the maximum margin hyperplane. All other training points have  $\alpha_i = 0$  and lie either on  $H_1, H_2$  (such that the equality in Eq. 6.6 holds), or on that side of  $H_1$  or  $H_2$  such that the strict inequality in Eq.6.6 holds. For these machines, the support vectors are the critical elements of the training set. They lie closest to the decision boundary; if all other training points were removed (or moved around, but so as not to cross  $H_1$  or  $H_2$ ), and training was repeated, the same separating hyperplane would be found. Once a Support Vector Machine has been trained, the problem of classifying a new data vector  $x$  is now simply to determine on which side of the decision boundary (that hyperplane lying half way between  $H_1$  and  $H_2$  and parallel to them) a given test pattern  $x$  lies and assign the corresponding class

label, i.e. the class of  $x$  is taken to be with  $b$  obtained from the Karush-Kuhn-Tucker (KKT) conditions.

### 6.2.3 Non-Separable Case

In case the set  $S$  cannot be separated by any hyper surface, due to the partial overlapping of the two classes, the previous analysis can be generalized by introducing non negative slack variables  $\xi = (\xi_1, \xi_2, \dots, \xi_l)$  such that

$$y_i(x_i \cdot w + b) \geq 1 - \xi_i, \text{ for } \forall(i = 1, 2, \dots, l) \quad (6.11)$$

The solution to minimize the following modified cost function subject to the constraints in Eq. 6.11 is called Soft Margin Separating Hyperplane (SMSH).

$$J(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (6.12)$$

where  $C$  is the regularization parameter. If the parameter  $C$  is small the separating hyperplane (i.e. SMSH) tends to maximize the distance  $1/\|w\|$ , while the larger  $C$  will cause SMSH to minimize the number of misclassified points. Again the vectors satisfying the constraints above with the equality sign are termed as support vectors and the only vectors needed to determine the decision surface. Similarly to the linearly separable case, the dual formulation requires the solution of a quadratic problem with linear constraints:

Maximize:

$$L_D \equiv \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (6.13)$$

subject to:

$$0 \leq \alpha_i \leq C \quad (6.14)$$

$$\sum_i \alpha_i y_i = 0 \quad (6.15)$$

The solution is again given by

$$w = \sum_i^{N_S} \alpha_i y_i x_i \quad (6.16)$$

where  $N_S$  is the number of support vectors. The only difference from the optimal

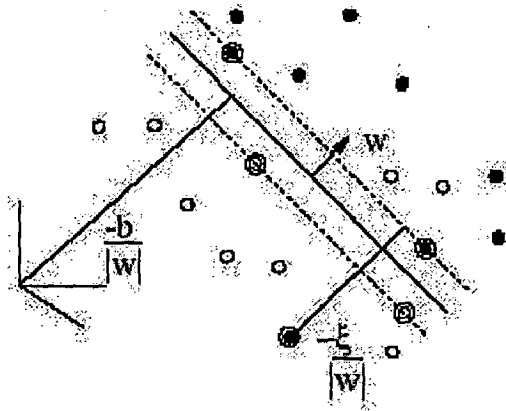


Figure 6.2: Linear separating hyperplanes for the non-separable case. (The support vectors are circled) [27]

hyperplane case is that  $\alpha_i$  now have the upper bound of  $C$ . The Fig. 6.9 depicts an example of a set of non-separable vectors belonging to two classes.

### 6.2.4 Nonlinear SVM Classifier

The linear SVM can be readily extended to a nonlinear classifier by first using a nonlinear operator or mapping function  $\phi(x)$  to map the input pattern  $x$  into higher dimensional space  $H$ . The nonlinear SVM classifier so obtained is defined as

$$f(x) = w \cdot \Phi(x) + b \quad (6.17)$$

which is linear in terms of the transformed data  $\Phi(x)$ , but non linear in terms of the original data  $x_i \in R^n, i = 1, \dots, l$ .

Following nonlinear transformation, the parameters of decision function  $f(x)$  the minimization of following cost function:

$$J(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (6.18)$$

subject to

$$y_i(\phi(x_i) \cdot w + b) \geq 1 - \xi_i, \quad \xi_i \geq 0; \quad i = 1, 2, \dots, l. \quad (6.19)$$

The solution to minimize above cost function subject to the constraints in Eq. 6.19 from the following dual:

Maximize:

$$L_D \equiv \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (6.20)$$

where  $K(\mathbf{x}, \mathbf{y})$  known as kernel is a nonlinear function and is defined as

$$K(\mathbf{x}, \mathbf{y}) \equiv \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}) \quad (6.21)$$

subject to:

$$0 \leq \alpha_i \leq C \quad (6.22)$$

$$\sum_i \alpha_i y_i = 0 \quad (6.23)$$

The solution is again given by

$$\mathbf{w} = \sum_i^{N_s} \alpha_i y_i \Phi(\mathbf{x}_i) \quad (6.24)$$

where  $N_s$  is the number of support vectors. The only difference from the optimal hyperplane case is that now have the upper bound of  $C$ . The KKT optimality conditions for (Eq. 6.20) will lead to the following three cases for each  $\alpha_i$ :

1.  $\alpha_i = 0$ . This corresponds to  $y_i f(\mathbf{x}_i) > 1$ . In this case, the data element  $x_i$  is outside the decision margin of the function  $f(x)$  and is correctly classified.
2.  $0 < \alpha_i < C$ . In this case,  $y_i f(\mathbf{x}_i) = 1$ . The data element  $x_i$  is strictly located on the decision margin of the function  $f(x)$ . Hence  $x_i$  called a margin support vector of  $f(x)$ .
3.  $\alpha_i = C$ . In this case,  $y_i f(\mathbf{x}_i) < 1$ . The data element  $x_i$  is inside the decision margin (though it may still be correctly classified). Accordingly,  $x_i$  called error support vector of  $f(x)$ .

### 6.2.5 SVM Kernel Functions

The kernel function in an SVM plays the central role of implicitly mapping the input vector (through an inner product) into a high-dimensional feature space.

1. Polynomial kernel:



$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^p \quad (6.25)$$

2. Gaussian RBF kernel:

$$K(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2) \quad \text{where } \gamma = \frac{1}{2\sigma^2} \quad (6.26)$$

## 6.3 Results and discussions

### 6.3.1 Input data

Mammograms containing microcalcifications from McGill University database have been used in this work. Total number of 1444 samples are obtained from mammograms containing microcalcifications from McGill University database [1]. Out of which 544 samples are from malignant class and remaining 900 samples are from benign class. These are divided into two sets: training and testing. The training-set contained 900 samples of which 600 are benign and 300 are of malignant nature respectively. Remaining 544 samples are used for testing of SVM classifier. Testing-set contained 300 cases of benign and 244 cases of malignant calcifications respectively.

### 6.3.2 Feature selection

Of course not all the features are important in the classification and some of them are of no significance. Features are selected by observing variation in their values and ability of feature to discriminate between two class. Each feature vector is applied to SVM and accuracy is observed and those features giving better accuracy of classification are taken for final classification. The feature's values for true and false clusters of microcalcifications are analysed separately [28].

The final feature set chosen for classification are Table 6.1.

### 6.3.3 Parameter Selection and Training

Once the training samples are obtained, the next step is to determine the optimal parametric settings of SVM. In this process, the following variables: the type of kernel

|                             |                                  |
|-----------------------------|----------------------------------|
| 1. Min ASM                  | 16. Min entropy                  |
| 2. Max ASM                  | 17. Max entropy                  |
| 3. Entropy                  | 18. Min IDM                      |
| 4. Area                     | 19. Max IDM                      |
| 5. Major Axis length        | 20. Min information correlation1 |
| 6. Minor axis length        | 21. Min information correlation2 |
| 7. Eccentricity             | 22. Max information correlation1 |
| 8. Min contrast             | 23. Max information correlation2 |
| 9. Max contrast             | 24. Min Sum average              |
| 10. Min correlation         | 25. Max Sum average              |
| 11. Max correlation         | 26. Min Sum entropy              |
| 12. Min difference entropy  | 27. Max Sum entropy              |
| 13. Max difference entropy  | 28. Min Sum variance             |
| 14. min difference variance | 29. Max Sum variance             |
| 15. max difference variance | 30. Min variance                 |
|                             | 31. Max variance                 |

Table 6.1: List of extracted features

function, its associated parameter, and the regularization parameter  $C$  must be decided. To optimize these parameters, sixfold cross validation has been applied to the training-set. This procedure consists of the following steps. First, divide randomly all the available samples in training-set into six equal-sized subsets. Second, for each model-parameter setting, train the SVM classifier six times; during each time one of the subsets is held out in turn while all the rest of the subsets are used to train the SVM. The trained SVM classifier is then tested using the held-out subset, and its classification error.

In this work we used Radial Basis Function kernel (RBF). Various parameters for the SVM like regularization parameter  $C$ , sigma of RBF etc. are varied as:  $C$  from 1 to 10000, and Gamma from 0.1 to 4 to choose the best parameters for SVM. After observing accuracy of classification it is found that for  $C=10$ ,  $\gamma=2$  it gives 82 % accuracy of classification.

*This chapter shows application of SVM for classification of mammograms using feature vectors extracted in last chapter. In initial training and testing SVM gave 82 % classification accuracy.*

# Chapter 7

## Conclusions and scope of future work

### 7.1 Conclusions

This work proposed a new algorithm for the detection of microcalcifications on mammograms. The basic idea was to evaluate the dynamics of microcalcifications, since each calcification appears as an elevation if we view each mammogram as a topographic map. Every suspicious object was marked using a binary image, which was used as a mask for object extraction from the original image. The features of the extracted objects were classified using support vector machine.

This thesis work proposed enhancement algorithms based on morphological operators & traditional Contrast Limited Adaptive Histogram Equalization(CLAHE). CLAHE is used for comparison purpose. CLAHE when applied on synthetic test images clearly indicated good contrast enhancement, but major problem is, this technique enhances noise equally along with microcalcifications. As per literature tophat enhancement is used by many researcher for enhancement of mammograms, but we proved that how this technique is structuring element dependent & how, the choice of structuring element becomes crucial in this enhancement algorithm. The h-dome makes use of morphological reconstruction for extracting high intensity peaks i.e. microcalcifications from the mammogram. The proposed h-dome filtering technique clearly indicated a better performance than other two techniques.

Qualitative analysis is done by presenting all enhanced images to expert radiologist. Radiologist gave positive feedback for our proposed method. In some of the cases tophat outperformed h-dome enhanced image. Which is clearly due to matching of structuring element in that case. From quantitative analysis, which used two evaluation indices i.e.CII and DV/BV, it can be concluded that h-dome is better technique for enhancement of mammograms as compared to other stated techniques.

In next part of thesis work, we presented minimum cross entropy based thresholding algorithm for segmentation of microcalcifications. This algorithm is tested first on synthetic images with known simulated microcalcifications. It is also gave a good segmentation results on McGill University Database as validated by radiologist.

Shape, textural statistical features have been extracted from each region containing suspected microcalcification. This feature sets are used for classification. In last part of thesis work we used these features for classification using SVM. SVM for Radial Basis Function kernel with  $C=10$ ,  $\alpha=2$  gave 82 % accuracy of classification.

## 7.2 Scope of future work

The performance of enhancement method depends on the imaging properties of database. As the imaging properties of the selected mammograms may affect the performance of stated method, further evaluation of the performance of the proposed algorithm using different databases is prompted.

In feature selection we used our judgement and variation in values of feature vectors for choosing particular feature set for classification. Further statistical analysis may improve selection of relevant features for classification. We tried our feature sets on SVM for classification with radial basis function kernel, there is scope for classification using SVM with different kernels and with selected relevant features.

# Bibliography

- [1] Interactive Mammography Analysis Web Tutorial-  
<http://sprojects.mmi.mcgill.ca/mammography>
- [2] Understanding cancer- <http://www.cancer.org/>
- [3] H.D Chang, Xiaopeng Cai, Xiaowei Chen, Liming Hu, Xueling Lou, "Computer-aided detection and classification of microcalcifications in mammograms:survey", The journal of the pattern recognition society, May 2003, pp.2967-2991.
- [4] W.M.Morrow, R.B.Paranjape, R.M.Rangayyan, J.E.L.Desautels, "Region-based contrast enhancement of mammograms", IEEE Trans. on Medical Imaging, vol. 11, no. 3, 1992, pp. 392-406.
- [5] Dongming Zhao, M. Shridhar, David G. Dau, "Morphology on detection of calcifications in Mammograms", IEEE conference ,1992, pp. 129-132.
- [6] T.Stojic, I.Reljin, B.Reljin, "Local contrast enhancement in digital mammography by using mathematical morphology", International Symposium on Signals, Circuits and Systems, 2005 , Vol. 2, pp-609-612.
- [7] Jong Kook Kim, Jeong Mi Park, Koun Sik Song and Hyun Wook Park,"Adaptive mammographic Image enhancement using First Derivative and Local Statistics", IEEE Transaction on medical imaging, October 1997 ,pp.-495-502.
- [8] Michael Wirth, Matteo Fraschini, Jennifer Lyon,"Contrast Enhancement of microcalcifications in mammograms using morphological enhancement and non-flat structuring elements", Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems, CBMS, 2004.
- [9] Moti Melloul, Leo Joskowicz , "Segmentation of microcalcification in X-ray mammograms using entropy thresholding", CARS 2002.

- [10] Li C.H., Lee C.K., "Minimum cross entropy thresholding", *Pattern Recognition* 26, 1993, pp-617-625.
- [11] Robert M., Haralick, K. Shanmugam, Itshak Dinstein, "Textural features for image classification", *IEEE Trans. on Systems, Man, and Cybernetics*, Vol-SMC-3, No.6, 1973, pp-610-621.
- [12] S.Yu, L.Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films", *IEEE Trans Med. Imaging*, 2000, 19(2), pp-115-126.
- [13] B.Zheng, W.Qian, L.P.Clarke, "Digital mammography: mixed feature neural network with spectral entropy decision for detection of microcalcifications", *IEEE Trans. Med. Imaging*, 1996, 15(5), pp-589-597.
- [14] J.K.Kim, H.W.Park, "Statistical textural features of microcalcifications in digitized mammograms", *IEEE Trans. Med. Imaging*, 1999, 18(3), pp-231-238.
- [15] Sihem Bouyahia, Jerome Mbainibeye, Nouredine Ellouze, "Computer aided diagnosis of mammographic images", *First International Symposium on Control, Communications and Signal Processing*, IEEE, 2004, pp-259-262.
- [16] I.El-Naqa, Y.Yang, M.N.Wernick, N.P.Galatsanos, R.M.Nishikawa, "A support vector machine approach for detection of microcalcifications", *IEEE Trans. on Medical Imaging*, Vol 21, no.12, 2002, pp-1552-1563.
- [17] J.C.Fu, S.K.Lee, S.T.C.Wong, J.Y.Yeh, A.H.Wang, H.K.Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications", *Elsevier: Computerised Medical Imaging and Graphics* 29, 2005, pp-419-429.
- [18] J.Dengler, S.Behrens, J.F.Desaga, "Segmentation of microcalcifications in mammograms", *IEEE Trans. on Medical Imaging*, IEEE, 1993, pp-634-642
- [19] Stelios Halkiotis, Taxiarchis Botsis, maria Rangoussi, "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks", *Elsevier: Signal processing* 87, 2007, pp-1559-1568.
- [20] L.Vincent, "Morphological grayscale reconstruction in image analysis : applications and efficient algorithms", *IEEE Trans. Image Processing*, 1993, pp-176-201.

- [21] Laine AF, Schuler S, Fan J and Huda W, "Mammographic feature enhancement by multiscale analysis", IEEE Trans. Med. Imag., 1994, pp-725-740.
- [22] A.Vanzo, G.Ramponi, G.L.Siearanza , "An Image enhancement technique using polynomial filters", in IEEE International Conference on Image processing , IEEE, 1994, pp.477-481.
- [23] C.H.Li, P.K.S. Tam,"An iterative algorithm for minimum cross entropy thresholding",Pattern recognition Letters 19, 1998,pp-771-776.
- [24] A.P Dhawan, Y.Chitre, C.Kaiser-Bonaso, Moskowitz,"Analysis of mammographic microcalcification using gray-level image structure features",IEEE Trans. Med. Imag.,Vol.15,No.3, June 1996, pp-246-259.
- [25] H.S.Zadeh, S.P.ourabdollah-Nezhad, F.R.Radc,"Shape-based and texture-based feature extraction for classification of microcalcifications in mammograms", Proc. SPIE Vol. 4322, pp- 301-310 .
- [26] L.Wei, Y.Yang, R.M.Nishikawa, Y.Jiang, "A study on several machine learning methods for classification of malignant and benign clustered microcalcifications", IEEE Trans. on Medical Imaging, Vol-24, No.3, 2005, pp-371-380.
- [27] Burges C. J. C," A tutorial on Support Vector Machines for Pattern Recognition".
- [28] S. C.Cheran, R.Cataldo, P.Cerello, F.De Carlo, F.Fauci, G.Forni, B.Golosio, A.Lauria, E.Lopez Torres, I. De Mitri, G.Masala, G.Raso, A. Retico, A.Tata, "Detection and classification of microcalcifications clusters in digitized mammograms "Nuclear Science Symposium Conference Record, IEEE, 2004, pp-4136-4140.