DETECTION AND CLASSIFICATION OF MICROCALCIFICATION IN MAMMOGRAMS

A DISSERTATION

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

MASTER OF TECHNOLOGY

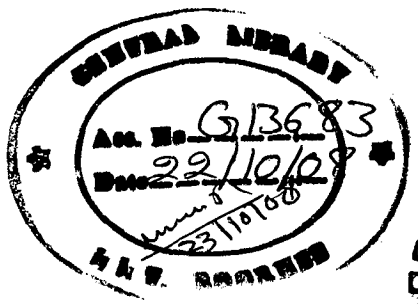
in

ELECTRICAL ENGINEERING

(With Specialization in Measurement & Instrumentation)

By

ANAND PATIDAR



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE - 247 667 (INDIA)

JUNE, 2008

I.D. No.-M.T. 430/2008-14 VK

CANDIDATE'S DECLARATION

I hereby declare that the work presented in this dissertation entitled "**DETECTION AND CLASSIFICATION OF MICROCALCIFICATION IN 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 2007 to June 2008 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 :27 June,2008

Place : Roorkee

Patidar

(ANAND PATIDAR)

CERTIFICATE

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

14.7.16

(Dr. Vinod Kumar)
Professor
Department of Electrical Engineering
Indian Institute of Technology, Roorkee
Roorkee-247667, Uttrakhand, 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.

I am grateful to all my teachers of M&I group for their suggestions and constant encouragement.

Timely assistance and help from laboratory staff of Biomedical Image and Signal Processing is sincerely acknowledged.

My special thanks to Vikas Gupta, Manish Kumar Varshney, Prashant Ashok Gupta for their valuable suggestion and discussions(tech and non-tech). My special thanks to my hostel group Varun Jain, Laxmikant Zunke 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.

(ANAND PATIDAR)

ABSTRACT

Breast cancer is becoming a major cause of death among women. As its cause is still unknown, early detection of breast cancer has become important to improve breast cancer treatment. Mammography plays an important part in early detection of breast cancers. Mammography is the X-ray imaging that uses a low-dose x-ray system for examining the breasts. Among all the mammogram abnormalities, microcalcications are the most difficult type of tumor to detect. Area of thesis is microcalcification detection in mammograms and classification on the basis of diagnostic feature information extracted form mammograms.

Thesis work proposed a computer aided diagnostic system for detection and classification of microcalcification in mammograms. CAD system consist of four stages - Enhancement, Segmentation, Feature extraction, and Classification of microcalcification into benign and malignant case.

Enhancement algorithm developed is local contrast enhancement based on morphological operation. For the evaluation of the performance of the enhancement algorithm, contrast improvement index and detail variance to background variance ratio are used.

After enhancement, algorithm for segmentation of microcalcification based on edge detection and morphological operation is implemented. After segmentation of each microclacification from mammogram, diagnostic features are calculated. shape based and texture based features are extracted. Total 28 features are calculated.

Feature set extracted is used for classification of microcalcification between benign and malignant case. Feature set is divided into two groups - training data set and testing data set. SVM with radial basis function is trained first using training data set and then tested using testing data set.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Introduction

Breast cancer is one of the major causes for the increase in mortality among women, especially in developed and under developed countries. In order to detect early onset of cancers in breast screening, Radiologists reading mammograms should be trained in the recognition of the signs of early onset of, which may be subtle and may not show typical malignant features. As its cause is still unknown, early detection of breast cancer has become important to improve breast cancer treatment. Leading experts, the National Cancer Institute, the American Cancer Society, and the American College of Radiology now recommend annual mammograms for women having age over 40. Mammography is the X-ray imaging that uses a low-dose x-ray system for examining the breasts. As for mammography low dose x-rays are used, the image quality is poor and many time it happens that radiologist misses the detection of breast cancer. Mammograms are the X-ray images of breast.

Our aim is to have an early detection of breast cancer. Early detection helps for better Diagnosis of the breast cancer. The chapter gives brief introduction about the current research in the area of mammogram analysis and also about the overview of the thesis work.

1.2 Objective of Thesis Work

The presence of microcalcification clusters (MCCs) is an important sign for the detection of early breast carcinoma. Computer Aided Diagnosis (CAD) systems have improved diagnosis of abnormalities in mammogram images. Microcalcifications are small calcium deposits on breast tissue, presence of which shows high correlation with breast cancer. Detection of microcalcification on mammogram is most important for diagnosis of breast cancer.

An early Detection of Breast Cancer using mammography 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 show that reading of mammograms by a second radiologist improves the accuracy of mammogram interpretation or in other words better diagnosis of breast cancer. Development of Computer aided detection system will replace second radiologist and helps in separate out clearly mammogram

1.3 Literature Survey

Diagnostic Features in the Mammogram i.e. microcalcification are vary widely in shape n size. Several studies show that conventional methods for the enhancement of mammograms cannot adapt to varying characteristics features in the mammogram. Conventional enhancement techniques for the enhancement of mammograms are discussed in this paper [3]. Conventional techniques include contrast stretching based enhancement, histogram equalization based technique, convolution mask enhancement and fixed and adaptive neighboring enhancement are discussed in this paper [3].

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 [3] [6].

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 [4].

Region-based image processing technique discussed in this paper [3] [5] enhances the contrast of the mammographic features of ROIs with various sizes and shapes according to the change of their surroundings. The extent and shape of the grown region adapt to local variation of the gray levels.

Convolution masking [3] is commonly used for mammography enhancement. The unsharp masking [5] [3] [6] and sobel gradient operators are two examples. The unsharp masking method reduces the low-frequency information while amplified the high frequency information. These processes could change the images dramatically.

An algorithm for both local contrast enhancement and background texture suppression in digital mammograms is proposed in [7]. Morphological operations can be employed for many image processing purposes, including edge detection, segmentation, and enhancement of images. 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 [7]. There are number of enhancement algorithm are implemented and reviewed in [3] [6]. 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. Further Michael wirth et al. [8]

proposed an algorithm which is based on morphological operations and non-flat structuring elements. It suggests 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.

The improvement in mammogram images after enhancement is often very difficult to measure. In mammograms, the improved perception is difficult to quantify. Laine et al. [9] proposed the contrast improvement index as a measure of the enhancement performance. A.Vanzo et al. [10] proposed statistical measure based on the separate estimate of the detail variance and background variance.

Various threshold techniques are implemented for segmentation of microcalcification. Melloul et al. [10] proposed an algorithm for the segmentation of microcalcification in the x-ray mammograms using entropy thresholding. It uses a three dimensional co-occurrence matrix for threshold calculation. C.H. li et al. [11] proposed a method for solving the problem of threshold selection by using minimum cross entropy.

J.C.Fu et al. [12] proposed an algorithm for segmentation of microcalcification based on edge detector algorithm. Gradient operators Canny and sobel are applied for the enhancement of edges around the microcalcification. It is followed by flood filling operation and morphological closing operation to eliminate noise.

Most systems extract features from the texture, spatial and spectral domains. It is the most important stage before the classification. J.C.Fu et al. [12] discussed feature extraction and selection for microcalcification classification based on texture information , *shape related information, and spectral domain based features. 61 features are extracted* from each Region of Interest (ROI) from textural, spatial and spectral domains.

Back-propagation neural networks (BPNN) were once commonly used for microcalcification classification [14]. BPNNs has a major disadvantage related to choosing training parameters such as the transfer function, number hidden layer nodes, and convergence algorithm.

The performance of two types of classifiers, GRNN and SVM is compared in the paper [12]. General regression neural network employs a base of radial functions for functional approximation. The support vector machine (SVM) approach also known as structural risk minimization is based on intuitive geometric principles. SVM based computer aided system is also developed in this paper [13]. In this paper comparison between two types of kernel i.e. RBF kernel and polynomial kernel is made.

1.4 Overview of Thesis Work

The basic objective of Computer Aided Diagnosis (CAD) systems is having improved diagnosis of abnormalities in mammogram image. It provides a second opinion on radiologist's image readings. It also improves the quality and productivity by improving the accuracy of radiological diagnosis and reducing the image reading time. As the en-

hancement and detection of microcalcification are most important and crucial part of 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 and also provides better accuracy as compare to the CAD systems already developed.

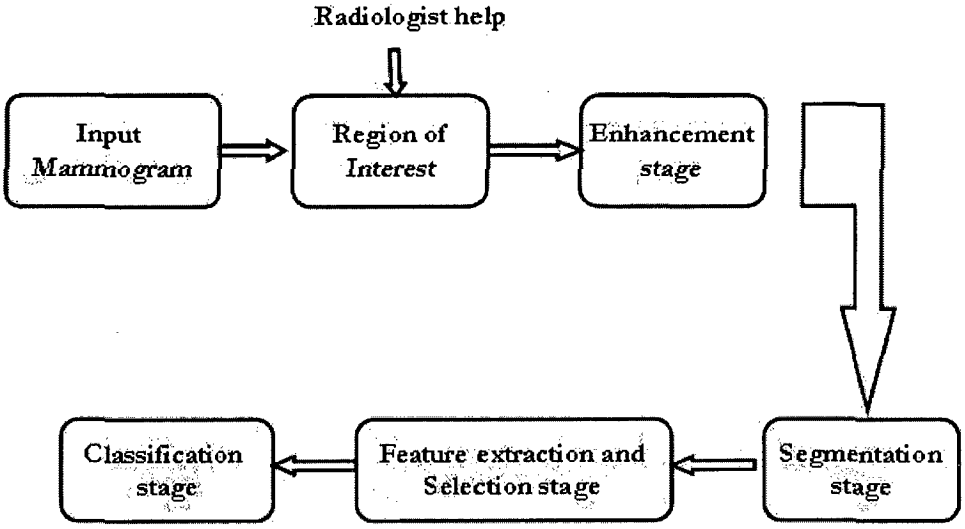


Figure 1.1: Proposed CAD system for detection and classification of microcalcification in mammograms.

The first step is extraction of region of interest from the input mammogram to process it further. Region of interest is extracted using the help of radiologist. After ROI extraction mammogram is processed using the developed algorithm for the enhancement. Algorithm is based on morphological top hat and bottom hat transformation. 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 the MCC detected are extracted and selected. Selected features are further used for the classification to classify MCCs into malignant and benign.

1.5 Organization of Thesis

In chapter one, brief introduction about the objective, literature review and overview has been given. In next chapter, an introduction to basics of mammography, different views of mammogram scanning, mammogram abnormalities, Problems in interpretation of the mammogram has been discussed in detail. Chapter three describes preprocessing of mammograms, morphological enhancement algorithms studied and implemented and finally formation of synthetic images for evaluation and results are discussed. In chapter four, different evaluation indices such as Contrast Improvement Index (CII), and Detail variance to Background variance (DV/BV) ratio, used for evaluating enhancement techniques has been discussed and calculated.

Chapter five describes an algorithm for the segmentation of microcalcification. Edge based detection and morphological operation based algorithm is applied to synthetic mammogram images for segmentation of the microcalcifications. After segmentation of microcalcification, next step is feature extraction. Chapter six describes labeling and object reduction techniques and, shape based and texture based features used in the process of feature extraction of the microcalcifications. After feature extraction, classification techniques based on support vector machine is described in chapter seven. Chapter eight summarizes the results, conclusions and describes directions of future work.

Chapter 2

Basic of Mammography

2.1 Introduction

Mammography is the X-ray imaging that uses a low-dose x-ray system for examining the breasts. It is acknowledged as a single most effective method for screening for breast cancer. The image captured using this technique is known as mammogram which are used to find potential signs of breast cancer like tumors, small clusters of calcium deposits and abnormal changes in the skin.

This chapter provides a brief overview of mammography. Important definitions and terminology used in mammography are described. Anatomy of female breast is discussed in short. Later an attempt is made to give an idea to reader about the problems involved in interpretation of mammograms. Mammogram abnormalities are discussed in detail.

2.2 Anatomy of the Female Breast

The breast is a mass of glandular, fatty, and fibrous tissues positioned over the pectoral muscles of the chest wall and attached to the chest wall by fibrous strands called Cooper's ligaments. The anatomy of the adult female breast consists of 12 - 20 conical lobes or grape bunches. From outside, the lobes feel like little nodes or lumps. The space in between the lobes is filled with connective and fatty tissue. Fat also surrounds the whole system of milk ducts and glands. 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 intra lobular connective tissue and adipose tissue [1].

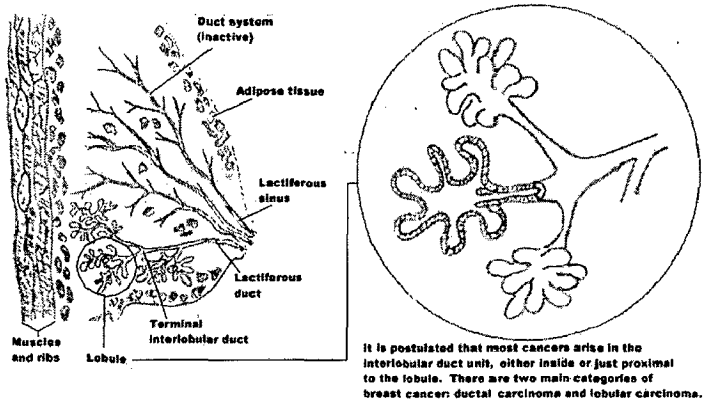


Figure 2.1: Schematic Diagram of the Female Breast[1]

There is also extra lobular connective tissue which binds the lobes together as well as extra lobular 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.

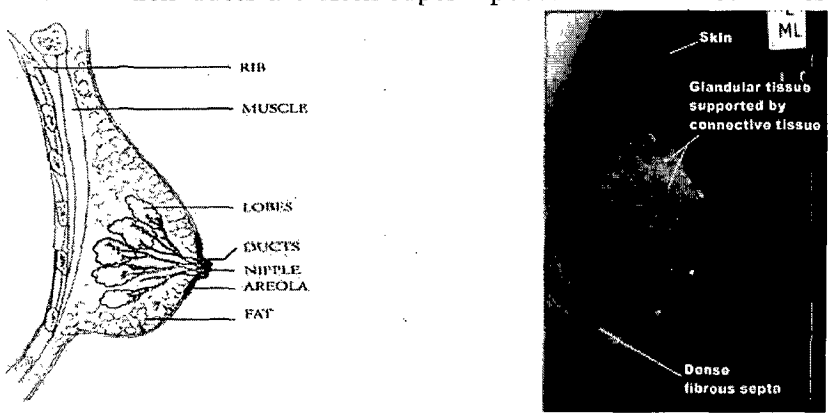


Figure 2.2: breast anatomy vs mammogram[1]

2.3 Mammogram Views For Breast Scanning

The breast can be imaged from a number of angles. During a routine mammogram, each of your breasts will be imaged separately with two different views. Each view shows somewhat different details and territory[2].

- Cranio-caudal (CC) view is taken from above a horizontally-compressed breast.
- Mediolateral-oblique (MLO) is taken from the side and at an angle of a diagonally-compressed breast.

2.3.1 Cranio-caudal (CC) view

A CC view of your breast may be taken during a routine mammogram as well as during a diagnostic mammogram. It will show as much as possible of your glandular tissue (ducts and lobes), the surrounding fatty tissue and the outermost edge of your chest muscle. The CC view can't capture much of the breast tissue that is in your armpit and upper chest.

2.3.2 Mediolateral-oblique (MLO)

An MLO view of your breast may be taken during a routine mammogram. The angle of an MLO allows more of your breast tissue to be imaged (it covers the main area of your breast) as well as the tissue in your armpit. If no oblique projection is taken, the mediolateral position may be preferable to the latero-medial view It will show glandular as well as fatty tissue, and it gives a larger area than a CC view.

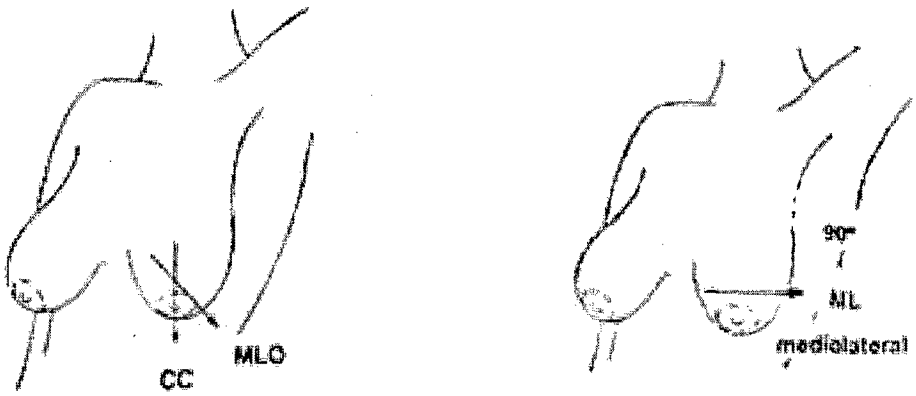


Figure 2.3: Different angles for breast scanning[2]

2.3.3 Other views

- Latero-Medial view : LM view images the breast from the outer side of the breast inward toward the center of the chest) since the lateral side of the breast, where pathological changes are most commonly found, is then closest to the film. However, if the physician wants to include as much of the medial side of the breast as possible, the LM view may be chosen.
- Medio-lateral view: The ML view is taken from the center of the chest outward.
- Spot compression : This view involves compression on only a small area, to get more detail.
- Cleavage view : In this view both breast compressed, to see tissue nearest center of chest.
- Magnification : This view is used to see borders of structures and calcifications.

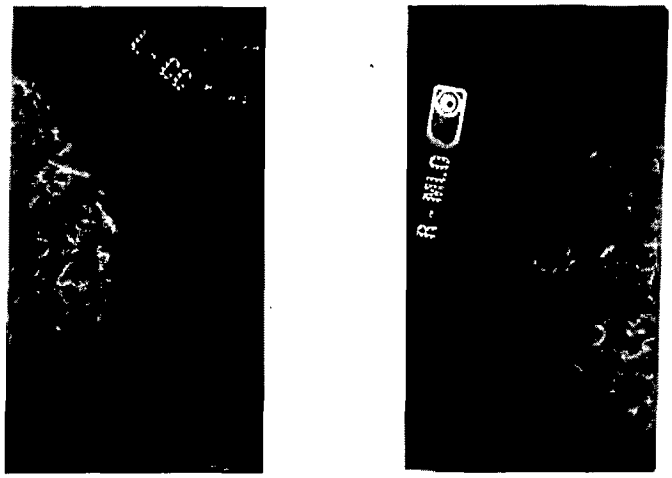


Figure 2.4: Mammograms for CC and MLO view[1]

2.4 Viewing the Mammograms

The detection of breast cancer involves the screening of breast tissues thought to be normal. The x-ray image of the breast exhibits a pattern within a spectrum of "normal variation." Mammographic interpretation involves careful scrutiny of this "normal pattern" and any abnormalities which present itself as a disruption of the "normal pattern"[1].

The first step is to compare the right breast to the left breast on the current examination. Since the breasts are seen as symmetric organs. Viewing the mammogram should proceed from a distance to closer scrutiny of the particular suspicious areas. From a distance, the images should be compared area for area, and the respective regions of the left and right breasts should look similar.

This evaluation for symmetry is followed by a close-up individual view of each image, looking for disruption in the "normal pattern" of the breast such as abnormal densities, areas of architectural distortion, masses and calcifications.

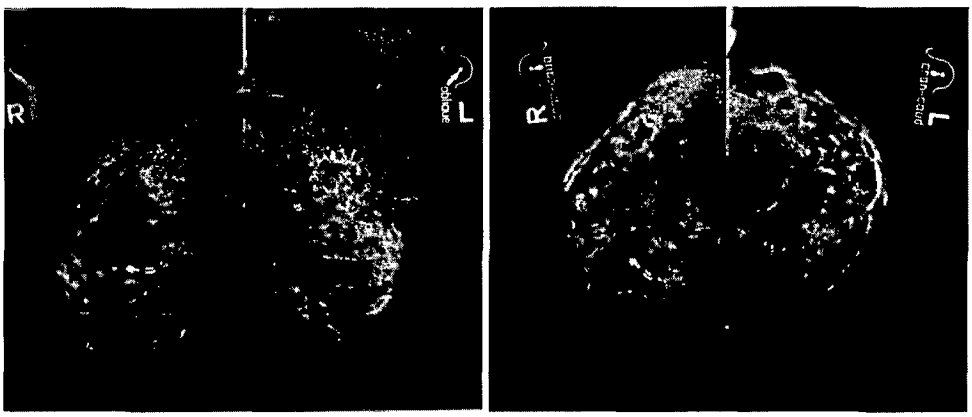


Figure 2.5: Comparision of right breast to left breast for Oblique and CC view[1]

2.5 Symptoms of Breast Cancer

In the early stage many symptoms don't involve any noticeable discomfort or pain.Even symptoms vary from woman to woman. In fact, when breast cancer first develops, there may be no symptoms at all. But as the cancer grows, the following symptoms generally occur [1].

- A lump or thickening in or near the breast or in the underarm area.
- 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).
- Pain in the breast, anything from stabbing pains through to a constant ach.
- Pink, red or dark colored area that has a texture similar to orange skin.

2.6 Sign of Breast Cancer

2.6.1 Asymmetric density

The breasts are seen as symmetric structures and should be compared as such. While exact mirror images are not to be expected, from an overall vantage point the tissue patterns within each breast should be similarly distributed. An asymmetric area may be indicative of a developing mass, a variation of normal breast tissue, postoperative change from a previous biopsy, or merely poor positioning and compression during imaging. The appearance of asymmetries due to positioning and compression during imaging is often the result of superimposition of normal breast structures. True breast asymmetry, on the other hand, is three-dimensional and should be present on both MLO and CC views[1]

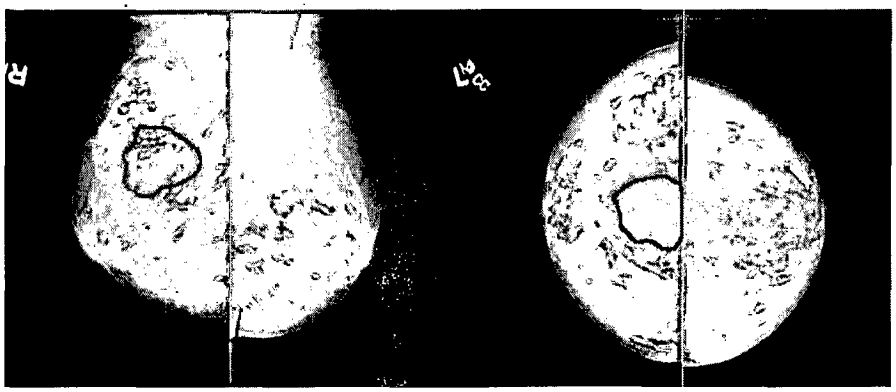


Figure 2.6: Mammograms with MLO and CC view showing asymmetric density[1].

Once an asymmetry is determined to be three-dimensionally real, the interpreter must determine whether the asymmetry is a benign variation of asymmetric breast tissue or a focal asymmetric density that may represent a significant mass.

The central difference between a focal asymmetric density and normal asymmetric breast tissue is the appearance of the former's density to be concentrated toward its center and to be forming a mass with a dense central zone. Confirmation is made by its recognition on more than two views, and the density tends to be concentrated rather than diffuse. Closer inspection, spot compression, and magnification may confirm that the focal asymmetric density is a mass. Benign asymmetric tissue, on the other hand, tends to be larger and more diffuse with a morphology and pattern similar to breast tissue in the contralateral breast. Closer inspection will reveal the absence of a radio dense center, diffused margins, and no associated mass formation or calcifications.

In conclusion, asymmetric density is often the first impression that the interpreter notices. It requires closer scrutiny and possibly further evaluations because asymmetric density may be a localizing sign of breast cancer. The management of a case, i.e. the decision to follow-up or to biopsy, will depend upon the interpretation of the asymmetric density.

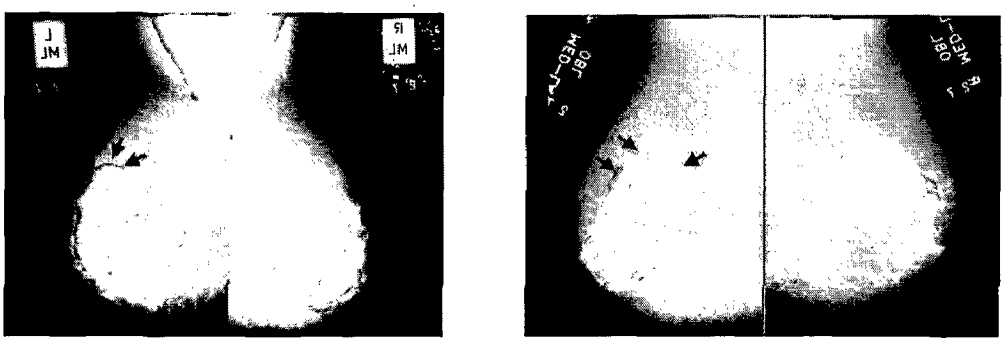


Figure 2.7: Mammogram showing Malignant and benign asymmetric density[1].

2.6.2 Masses

A mass commonly described as a "spot" or "density" may appear on the mammogram and has certain features which suggest a breast cancer. 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). Masses are quite subtle, and often occurred in the dense areas of the breast tissue, have smoother boundaries than microcalcifications, and have many shapes such as circumscribed, speculated (or stellate), lobulated or ill-defined. Presence of mass may associate with benign or malignant process. Depending

on the morphologic criteria of the mass, the likelihood of malignancy can be established [1][16].

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 distortion in the normal breast architecture, including speculations 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.

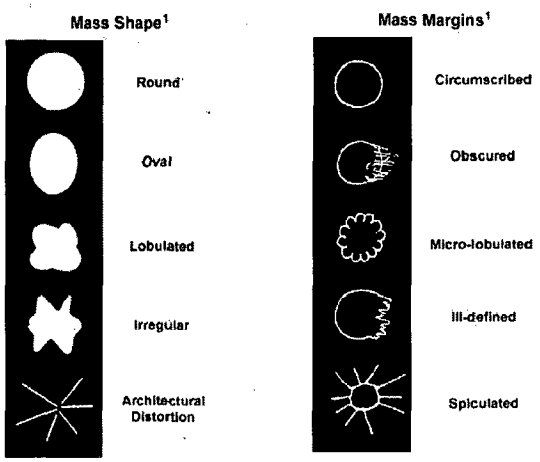


Figure 2.8: Mass shape and Margins[1].

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

Figure 2.9: Mammogram showing oval mass with well-defined circumscribed margin[1].

Effect on Surrounding Tissue and Associated Findings: These are descriptions associated with the mass such as architectural distortion, enlarged duct, skin changes, nipple and areolar abnormalities, etc.

2.6.3 Calcifications

Calcifications are often important and common findings on a mammogram. They can be produced from cell secretion or from necrotic cellular debris. They may be intramammary, within and around the ducts, within the lobules, in vascular structures, in interlobular connective tissue or fat. Alternatively, they may be found in the skin. 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. Certain calcification patterns are almost always pathognomic of a benign process, and in such cases no further analysis is needed.

Size: Microcalcifications may associated with a malignant process or a benign process. The problem with this general rule is that there is no fine line of measurement that could enable one to distinguish between micro and macro. All calcifications start out imperceptibly small and radiographically invisible. Most radiologists place calcifications 0.5 mm or less to have a high probability of association with cancer; and calcifications of

2.0 mm or larger are typical of a benign process. The smallest visible calcifications on a mammogram is approximately 0.2 - 0.3 mm.

Number: Microcalcification generally appears in group called clusters. 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. Any number of calcifications less than four will rarely lead to the detection of breast cancer in and of itself.

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.

2.6.4 Classification of Calcification

As discuss earlier the morphology of calcifications is considered to be the most important indicator in differentiating benign from malignant. The American College of Radiology (ACR) Breast Imaging Reporting and Data System (BIRADS) has classified findings of calcifications into three categories[1]:

1. Typically benign
2. Intermediate concern
3. Higher probability of malignancy

Different types of calcification falls into these categories are as follow

- 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: These calcifications are typically found in involuting fibroadenomas. Fibroadenomas usually regress with menopause and microcalcifications will develop into coarse macrocalcifications.
- 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.

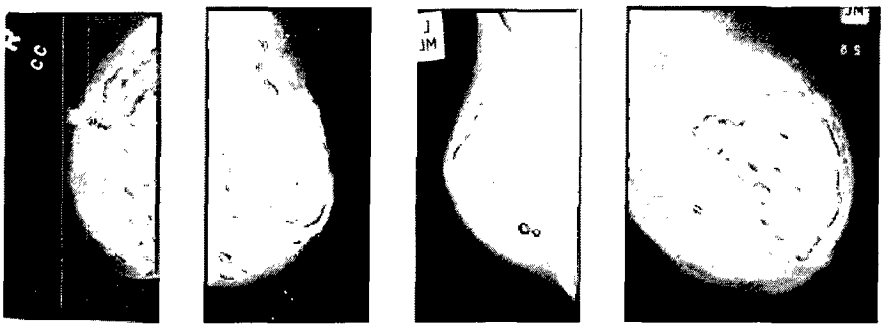


Figure 2.10: Mammogram showing Skin calcifications, Vascular calcifications, Popcorn-like calcification, Rod-shaped calcifications[1].

- Round Calcifications: Smooth round calcifications are associated with a benign process. They may vary in size in a cluster. When <1mm they are often found in the acini of lobules. When <0.5mm the term punctuate is used.
- Punctuate Calcifications: Round or oval calcifications <0.5mm are called punctuate. They appear as sharply defined, pinpoint deposits.
- Spherical or Lucent-Centered Calcifications: Spherical or lucent-centered calcifications can range from <1mm to >1cm. They may be found as debris collected in a duct, in areas of fat necrosis, and occasionally in fibroadenomas.
- Rim or Egg-shell Calcifications: These are thin calcifications that surround all or part of the margin of a mass. They are typically found in the wall of cysts. Breast cancer rarely produces this type of calcifications.

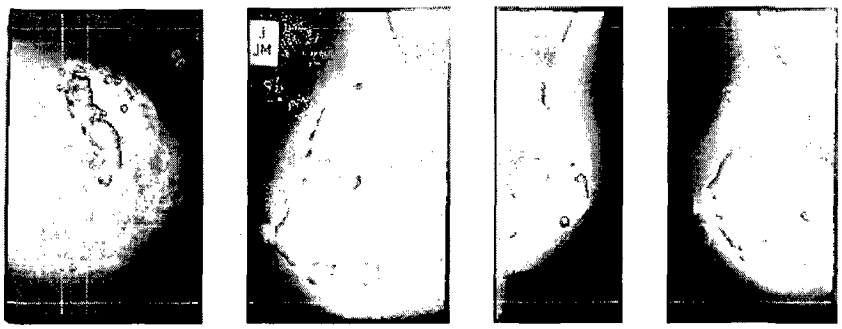


Figure 2.11: Mammogram showing Round calcifications, Punctuate calcifications, Spherical calcifications,Rim calcifications[1].

- Milk of Calcium Calcifications: This form of calcium precipitates and settles in the bottom of cysts. They appear appear as smudgelike amorphous particles on vertical beam (CC image) mammography and appear as sharply defined, semilunar, crescent-shaped on the horizontal beam view.

- Dystrophic Calcifications: These calcifications are irregular in shape but they are usually large, i.e. >0.5mm in size. They often form in the irradiated breast or after trauma to the breast.

- Pleomorphic or Heterogeneous Calcifications: Heterogeneous or pleomorphic calcifications in and of themselves are not associated with a benign or malignant process. However, a cluster of calcifications irregular in shape, size, and tend to be <0.5mm 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 <1mm in width.

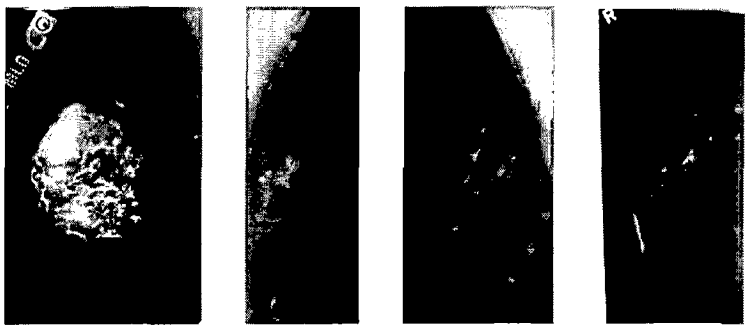


Figure 2.12: Mammogram showing Milk of calcium calcifications, Dystrophic calcifications, Pleomorphic calcifications, Fine linear calcifications[1].

Due to small size and low contrast to the background of microcalcification it becomes too difficult to detect microcalcifications even for experienced radiologist. Consequently, computer assisted detection of microcalcification has aroused a great deal of interest. In next five chapter an algorithm for detection and classification on microcalcification is explained.

Chapter 3

Mammogram Enhancement

3.1 Introduction

Image enhancement is used to refine a given image, so that the desired image features become easier to perceive for the human visual system or more likely to be detected by automated image analysis. Mammogram images have limited contrast because of the nature and superimposition of the soft tissues of the breast, which is compressed during imaging procedure.

This chapter explains the Morphological based algorithm developed for the enhancement of the mammogram images. Basic morphological operations are discussed in brief.

3.2 Need of Enhancement

Breast cancer is a leading cause of death among all cancers for middle-aged and older women, its incidence is rising as well. Hence, prevention and early diagnosis are very important. Unfortunately, primary prevention is not possible since the cause of this disease is not yet understood. For the detection of breast cancer mammograms are taken. Radiologist looks for certain signs and characteristics indicative of cancer when evaluating a mammogram. But due to it's fuzzy nature, low contrast, low differentiability from the surroundings, it becomes too difficult for radiologist to analyze mammogram.

To deal with above stated problems, it is very important to suppress the noise, to enhance the contrast between the region of interest (ROI) and background.

3.3 Contrast Enhancement

Image Enhancement technique are the mathematical techniques that are aimed at realizing improvement in the quality of the image. These techniques include remove noise, enhance contrast, and sharpen the details in a given image.

A digital image is defined as two-dimensional array of numbers that represent the real, continuous intensity distribution of a spatial signal. The continuous spatial signal

is sampled at regular intervals and the intensity is quantized to a finite number of levels. Each element of array is referred to pixel value. Digital is defined as a spatially distributed intensity signal *f(m,n)*, where f ia the intensity of pixel , and m and n define position of the pixel.

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix}$$

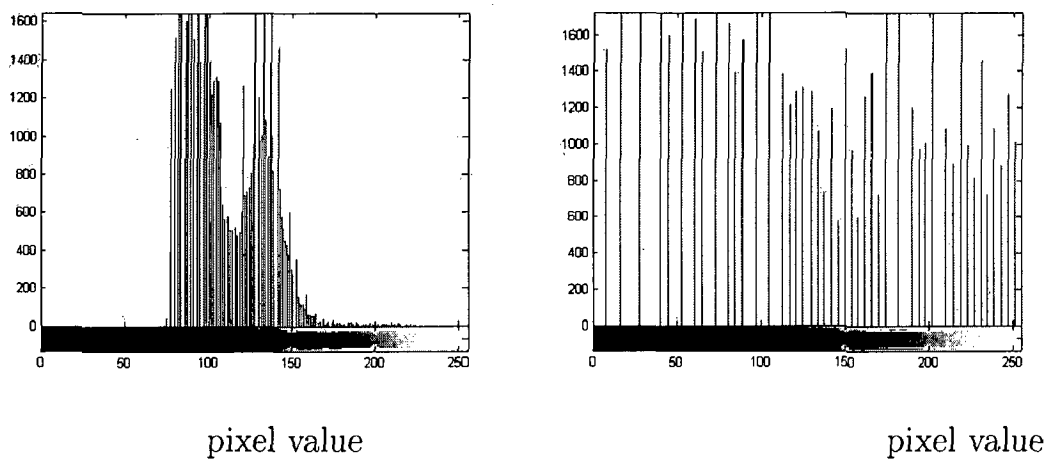


Figure 3.1: Low contrast and High contrast image histogram[15]

3.4 Preprocessing of Mammograms

Before enhancement of mammogram image , they are preprocessed i.e. region of interest (ROI) is extracted. Images from Mcgill University database where region of Interest are marked by experienced radiologist are cropped. Now the cropped region is further processed.

Due to above operation of cropping, unwanted part of the image is removed. Removing of unwanted or unnecessary information from the image helps number of background pixels from the images, storage requirements, I/O time, and image processing time to reduced significantly.

3.5 Morphological Based Enhancement

There are number of traditional methods are developed but they are fail to provide the accuracy regarding the shape n size of the microcalcification. Here the method developed

is based on the morphological enhancement. This method is known as Top hat filtering.

Mathematical morphology is a tool for extracting image components that are useful in the representation and description of region shape, such as boundaries, skeleton, and the convex hull. Morphological contrast enhancement is based on the notion of morphological top-hats which were first proposed by Meyer. 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 surrounding background in the images with uneven background intensity.

3.5.1 Basic of morphological enhancement

(1) Structuring Element: All morphological algorithms are based on operations between the two sets, one is a whole image I, or some of its part and the second one is the structuring element S. The size and shape of the structuring element plays a crucial role in the morphology-based image processing since they define a neighborhood of the actual image pixel under processing. Typical structuring elements are cross, square, disk, diamond, of different shapes.

0	0	1	0	0		1	1	1	1	1
0	1	1	1	0		1	1	1	1	1
1	1	1	1	1		1	1	1	1	1
0	1	1	1	0		1	1	1	1	1
0	0	1	0	0		1	1	1	1	1

Figure 3.2: structuring element with diamond and square shape

(2) Dilation of an image I: The dilation of a gray-scale digital image I(m,n) by a structuring element S(i,j) is defined as

$$(I \oplus S)_{(m,n)} = max\{I(m-i, n-j) + s(i,j)\}$$

with $[(m-i, n-j)] \in D_I, (i,j) \in D_s$, where $D_I$and $D_S$are the domains of I and S [7].

The origin of structuring element S is assumed on the actual pixel position (m,n) of the input image. From eq 3.1 pixel element at point (m, n) is replaced with maximum value of a sum of image pixels and structuring element within their common domain i.e. a union of the two sets with at least one overlapped element. If all members of a structuring element are positive, the output image tends to be brighter than the input, and dark details are reduced or completely removed, depending on how their values and shapes relate to the

structuring element used. Usually, the structuring elements have values of unity, when the dilation can be described as[7][15]

$$(I \oplus S)_{(m,n)} = max\{I(m-i, n-j)\}$$

In this case the pixel element at point (m,n) is simply replaced by maximum value of image pixels in the domain of operation.

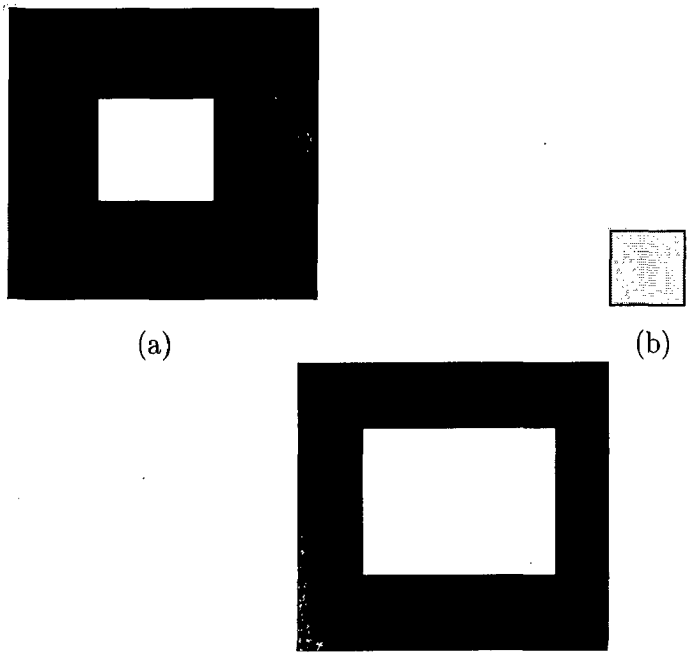


Figure 3.3: (a) Initial Image I (b) Structuring element (c) Output image after dilation with structuring element

3) Erosion of an image I: The erosion of gray scale digital image I(m,n) by a structuring element S(i,j) is defined as

$$(I \otimes S)_{(m,n)} = min\{I(m+i, n+j) - S(i,j)\}$$

with $\quad [(m+i),(n+j)] \in D_I, (i,j) \in D_S$

This operation is performed over all points of I totally covered by S. If all terms of the structuring element are positive, the output image tends to, be darker than the input. The bright details in the input image are reduced or removed, depending on, how their values and shapes relate to the structuring element. As by dilation, if the nonzero structuring elements are unity-valued, the erosion can be rewritten in simpler form

$$(I \otimes S)_{(m,n)} = min\{I(m+i, n+j)\}$$

In this case the pixel element at point (m,n) is simply replaced by minimum value of image pixels covered by structuring element located into the image object. After erosion output objects are shrinked [7][15].

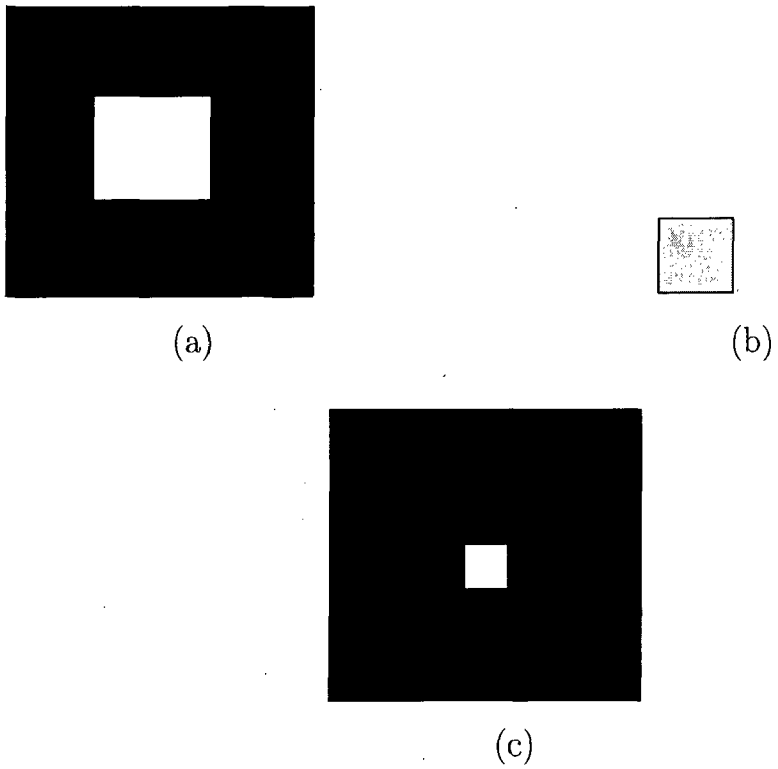


Figure 3.4: (a) Initial input image I (b) Structuring element (c) Output image after erosion with structuring element

By combining above two operation i.e. dilation and erosion we can derive various useful operations for image processing. Two of such operations used in the algorithm are opening and closing of the image is described below.

(4) Opening of an image I: The opening of image I by a structuring element S is defined as erosion followed by dilation. After applying gray scale opening light details smaller than structuring elements are removed [7][15]. It is defined as

$$I \circ S = (I \otimes S) \oplus S$$

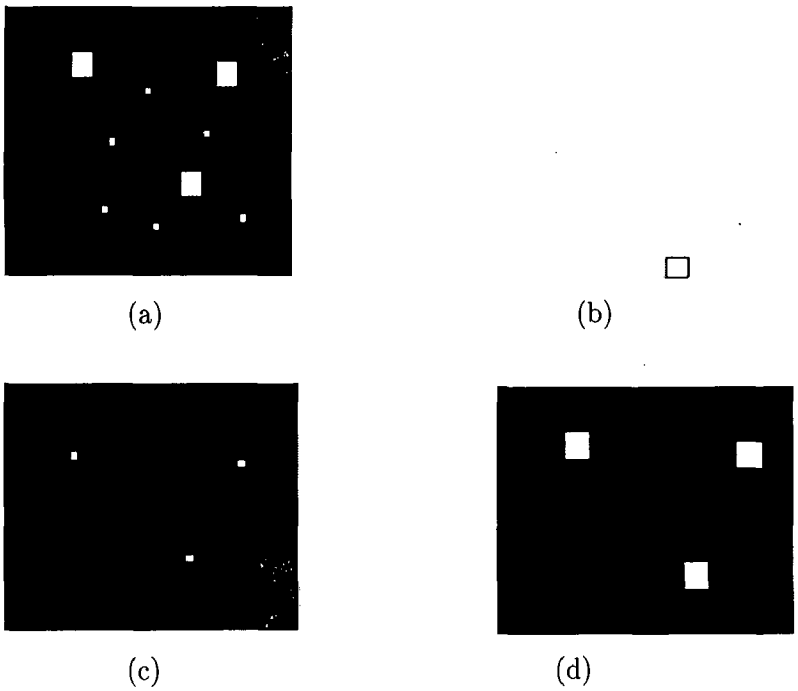
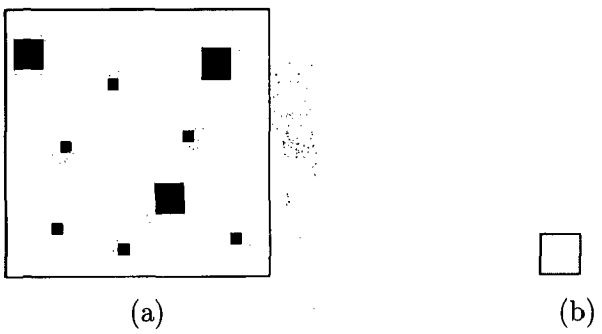
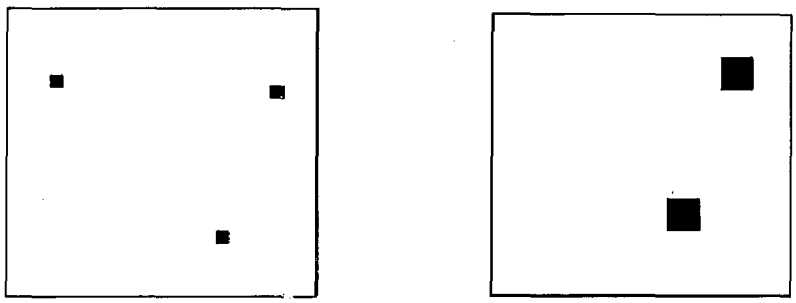


Figure 3.5: (a) Initial image with different size bright spots. (b) Structuring element. (c) Image after erosion. (d) Opened image .

(5) Closing of an Image: The closing of image I by a structuring element S is defined as dilation followed by erosion. After applying gray scale closing dark details smaller than structuring elements are removed [7][15]. It is defined as

$$I \bullet S = (I \oplus S) \otimes S$$





(c) (d)

Figure 3.6: (a) Initial image with some dark spots. (b) Structuring element. (c) Dilated image. (d) Closed image obtained after eroding the dilated image.

3.6 Local Contrast Enhancement

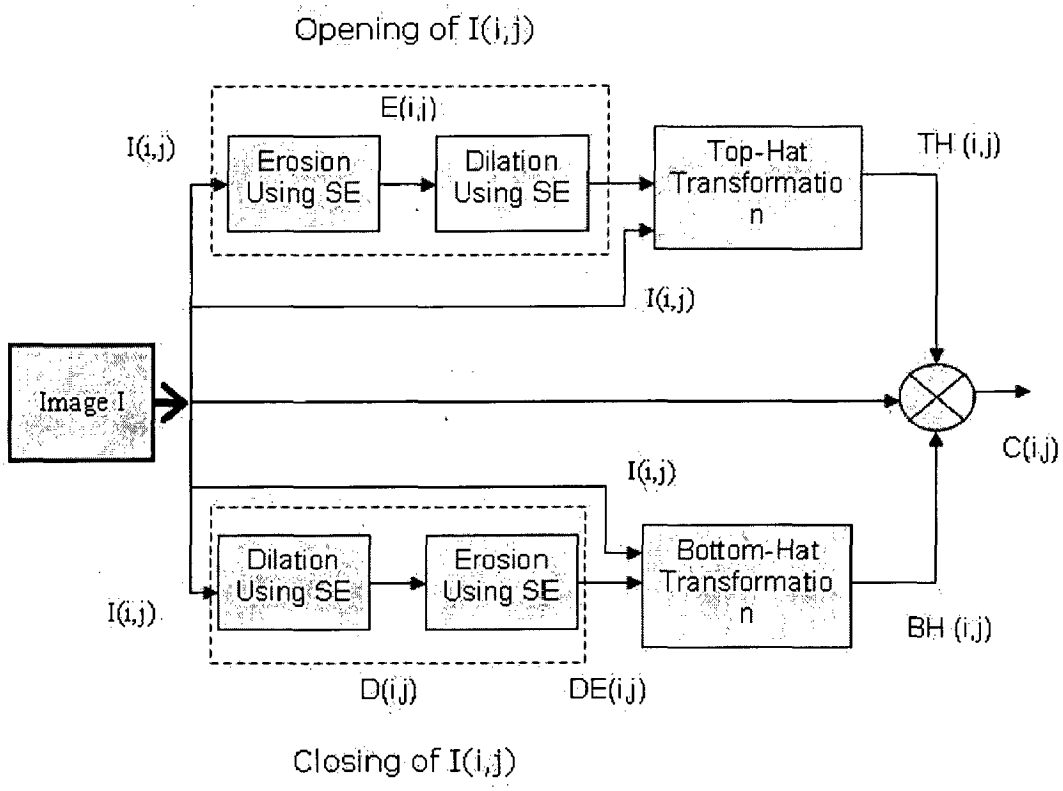


Figure 3.7: Block diagram for local contrast enhancement

Top hat transformation: Top hat transformation is obtained by subtracting a morphologically opened image from an original one (see figure 3.7)

$$TH = I - (I \circ S)$$

Top hat transformation is an excellent tool for enhancing small bright details from the background, and has proved its efficiency in the image fingerprint segmentation.

Bottom hat transformation: Bottom hat transformation is defined as a difference between a morphologically closed image and an original image (see figure 3.7) .

$$BH = (I \bullet S) - I$$

Bottom hat transformation helps in extracting dark features from a brighter background.

Local Contrast Enhancement: for obtaining local contrast enhancement first we add an original image to the top-hat transformed image, and making the difference between this sum and the bottom-hat image. Output image we will call a contrast image, C (see figure 3.7) [7]

$$C = I + TH - BH$$

By forming the difference (TH - BH), light details, smaller than the structuring element, are strongly emphasized, while dark details are suppressed. Furthermore, by adding an original image to this difference some kind of high-frequency filtering is achieved. Consequently, the enhancement of bright details smaller than structuring element S is reinforced, and uneven surrounding texture is highly equalized.

Additional background suppression without affecting already extracted bright details may be obtained by forming the difference between the contrast image C and its complement (negative image) C' = 1-C,assuming normalized levels (1 =white ,0 = black). An image obtained after this operation is denoted as a difference image D(see fig 3.8).

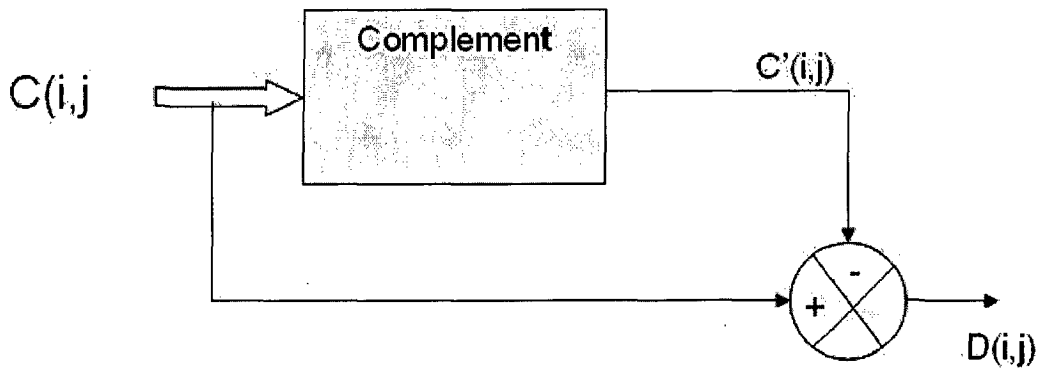


Figure 3.8: Additional Background Suppression of Enhanced Image

If the pixel value are normalized in the range og 0 and 1 the expression for additional background suppression is written as

$$D = C - C' = 2C - 1$$

3.7 Result and Discussion

3.7.1 Test images

A . Synthetic Image : One of the major problem with the mammogram images is uncertainty of the presence of microcalcification as well as if a false positive is not an actual calcification. This is the only reason radiologist are prone to human errors.

Therefore synthetic images are created with known microcalcifications on them. While preparing different synthetic images we have taken care of shapes and sizes of microcalcifications.

First our algorithm is applied on synthetic images with known number of microcalcification and, with known shape and size. Some of the test images are shown below :

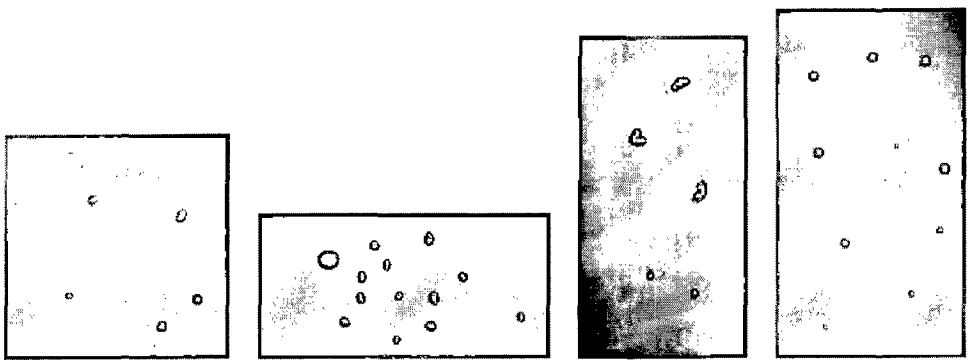


Figure 3.9: Synthetic images with known number of microcalcification

A.2 Image Database : Standard mammogram images from the Mcgill University database with known region of interest ROI are used [1]. First original image is cropped into ROI image marked by specialist. The advantage of cropping is that we can discard portions of the image that we don't want or that do not contribute to the composition. But cropping of images is strictly based on the information available in the mini mammographic database and also after concerning with the radiologist.

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, ill dened masses, architectural distortion, asymmetry and normal.

3.7.2 Result

Figure 3.10, 3.11, 3.12, 3.13 shows result of the top hat enhancement algorithm. Each figure shows 3 images - (a) original Image (b) Top Hat enhanced Image (c) Image after additional background suppression after subtracting the compliment of the enhanced image from itself. Results are divided into two parts. First part shows result of enhancement algorithm on synthetic images and second part shows result of enhancement algorithm on original mammogram images obtained from Mcgill University database.

Part 1 : Results on synthetic images

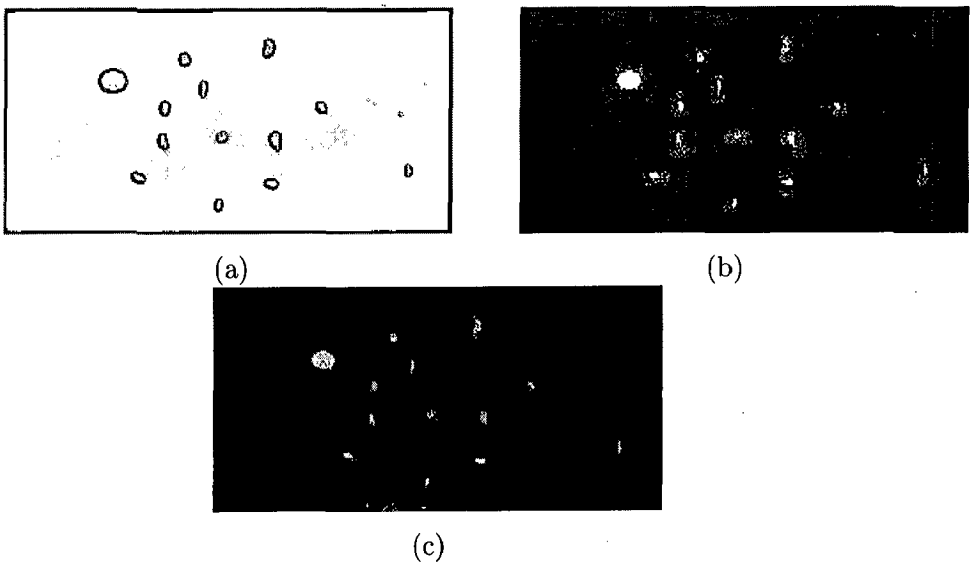


Figure 3.10: Result on synthetic test image 1 : (a) input image (b) Enahnced image (c) output after additional background suppression

In the above figure 3.10, image (a) is the snthetic input image with known number of calcification. The image contain 13 simulated calcifications. Image (b) shows the enhance output of input image...where the ratio of foreground gray level to background level is increased as compare to original image. Similarly after applying additional background suppression technique, gray level of the background is decreased further.

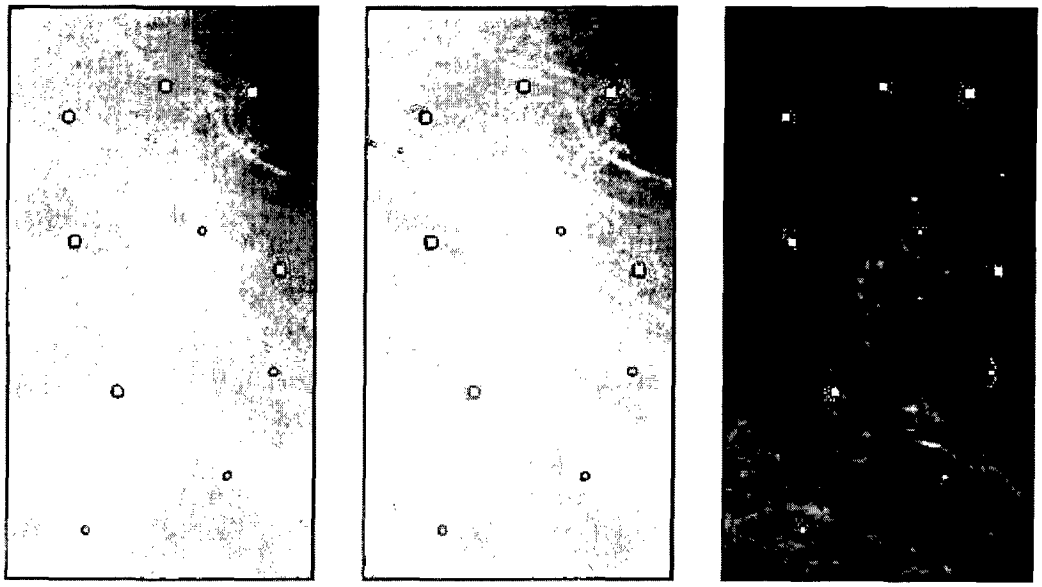


Figure 3.11: Result on synthetic test image 2 : (a) input image (b) Enahnced image (c) output after additional background suppression

In figure 3.11, synthetic image with simulated microcacification is shown. After enhancement, the contrast of foreground is increased as well the background is also increased. Our aim is to increase the ratio of foreground to background contrast level. After applying additional background suppression, background gray level is decreased while foreground gray level is increased. The advantage of applying this technique is ratio of the froeground gray level to backgorund gray level in increased and the calcification spots are clearly detected.

Part 2 : Result on original mammogram ROI

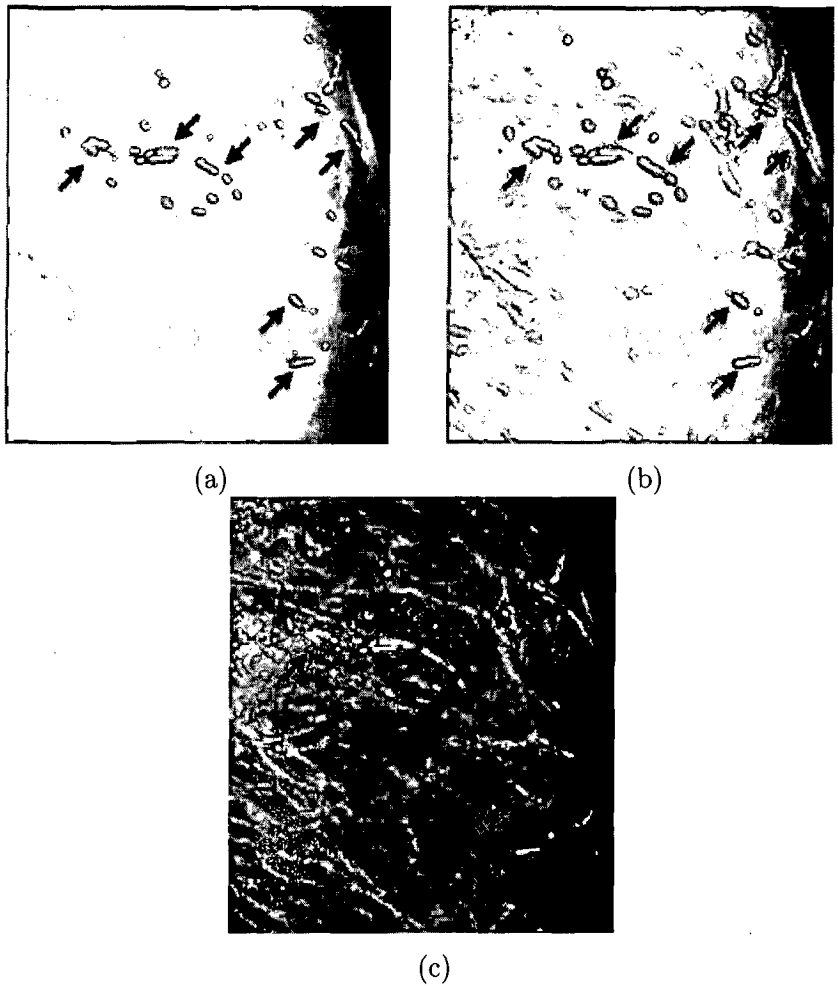


Figure 3.12: Result on case25_Lmlo : (a) input image (b) Enahnced image (c) output after additional backgraound suppression

An ROI of original mammogram is shown in figure 3.12 (a). Micorcalcification are shown by arrows. If we closely observe, contrast of the input image is enhanced after applying the enahmcement algorithm, and the calcification are clearly spotted in the enhanced image. But background contast is also enhanced which is not desirable. For this we applied background supression technique. The advantage of this technique is ratio of forground gray level to background gray level is increased and the calcification are clearly differentiated from background region.

Chapter 4

Quantitative Analysis of Enhancement Algorithm

4.1 Introduction

It is very difficult to measure the improvement of the enhancement objectively. If the enhanced image can make observer perceive the region of interest better, then we can say that the original image has been improved. In mammograms, the improved perception is difficult to quantify. In order to measure the improvement in contrast after applying enhancement algorithms, it is better to design some methods for the evaluation of enhancement objectively. In the previous chapter, Local Contrast Enhancement algorithm is discussed and applied on number of images both sythetic and original database from Mcgill University.

The statistical measurements such as variance or entropy can always measure the local contrast enhancement, however, that show no consistency for the mammograms and have not been particularly meaningful for mammogram images.

In this Chapter , two evaluation indices i.e. contrast improvement index (*CII*) and DV/BV ratio is explained. Evaluation of enhancement algorithm is done based on the contrast improvement index and detail to background variance ratio. Finally we used image proles of a mammogram for visual inspection of performance of enhancement.

4.2 Contrast Improvement Index (*CII*)

A quantitative measure of contrast improvement can be defined by a Contrast Improvement Index (*CII*)

$$C = \frac{C_{processed}}{C_{original}}$$

where Cprocessed and $C_{original}$ are the contrasts for a region of interest in the processed and original images[9].

Morrow et al.[5] has defined the contrast of the image as

$$C = \frac{f - b}{f + b}$$

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. With the aid of the mathematical phantom, we computed local masks to separate the foreground and background regions of each feature included in the blended mammogram.

4.3 Detail and Background Region Variance

Vanzo et al. proposed an evaluation index based on the separate estimate of the local variance on the image details (detail variance) and in uniform areas (background variance).These parameters are used for comparative analysis of different enhancement algorithms[17].

steps involved in evaluating average detail variance and average background variance are as follow

1. The first step is to define detail and background region in the ideal image. for this we evaluate the variance in a n x n window over each pixel (here n is taken as 3).

2. Now we compare the variance calculated with a fixed threshold value. the threshold value is calculated as the average value of the minimum and maximum pixel value in the image.

3. If the variance is greater than the threshold value, the pixel is said to belongs to detail region, otherwise to the background region. In this way we divided the whole image into two regions i.e. detail and background regions.

4. Pixels belong to detail variance are assigned a value of '1' (maximum) and pixel belongs to background variance are assigned a value of '0' (assume the pixel value in the image ranges between 0 and 1). In this way , a binary map image is generated where, say, white pixels indicate details and black ones indicate background.

5. Now we calculate the variance in n x n window of the processed image centered on each pixel.

6. If the pixel belongs to white pixel in the binary image then the variance is accumulated in the detail region variance register (DVR) otherwise to the background region variance register (BVR).

7. After Processing the whole image, detail region variance register contains the sum of the variance in n x n window of all the pixels belong to the detail region and background variance register contains the sum of the variance in n x n window of all the pixels belong to the background region.

8. Now total number of pixels in the detail region and the background region is calculated, say *DPN* is total number of detail region pixel and *BPN* is total number of background region pixel.

9. Next step is to calculate the average detail variance (DV) by dividing detail region variance register by total number of detail region pixels.

$$DV = \frac{DVR}{DPN}$$

10. Similarly average background variance is calculated by dividing background region variance register by total number of background region pixels.

$$BV = \frac{BVR}{BPN}$$

11. After Calculating DV and BV value, ratio of detail variance to background variance is calculated as an estimation of evaluation index for the enhancement algorithm.

$$Ratio = \frac{DV}{BV}$$

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 39 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, indi- cating that they represent two reliable quality factors when it is needed to compare the performance of different operators.

4.4 Results and Discussion

Two Indexes are calculated for the evaluation of enhancement algorithms is Contrast Improvement Index (CII) and Detail Variance to background variance ratio.

Table 4.1 shows the value of CII for the enhanced images as compare to original images of Mcgill Unversity mammogram database. A graph is drawn (figure 4.1) between CII and

Mcgill Images, which shows that Contrast of the enhanced image is greater than original image as CII is greater than 1.

Table 4.2 and figure 4.2shows the value of DV/BV ratio for the enhanced images as compare to original images of Mcgill Unversity mammogram database.

Table 4.1 : Contrast Improvement Index for Top Hat enhancement applied on ROIs of mammograms containing microcalcification

Name of File	Original Image Contrast	Enhanced Image Contrast	CII
case11Lcc_small_ans.jpg	0.089	0.035	0.3967
case11Lmlo_small_ans.jpg	0.026	0.005	0.1984
case11Rcc_small_ans.jpg	0.152	0.241	1.5829
case11Rmlo_small_ans.jpg	0.036	0.025	0.6893
case11_1Lcc_small_ans.jpg	0.02	0.029	1.5067
case11_1Rmlo_small_ans.jpg	0.121	0.168	1.3797
case15Lcc_small_ans.jpg	0.341	0.138	0.4056
case15Lmlo_small_ans.jpg	0.134	0.074	0.5554
case15Rcc_small_ans.jpg	0.174	0.202	1.1641
case15Rmlo_small_ans.jpg	0.152	0.099	0.6514
case15_1Lcc_small_ans.jpg	0.16	0.124	0.7764
case15_1Lmlo_small_ans.jpg	0.31	0.134	0.4325
case15_1Rmlo_small_ans.jpg	0.083	0.066	0.7946
case22Lcc_small.jpg	0.515	0.117	0.2261
case25Lcc_small_ans.jpg	0.062	0.005	0.0737
case25Lmlo_small_ans.jpg	-0.059	-0.03	0.4443
case25Rcc_small_ans.jpg	0.084	0.019	0.2245
case25Rmlo_small_ans.jpg	-0.038	-0.02	0.6412
case37Lcc_ans_small.jpg	-0.014	-0.04	3.2554
case37Lmlo_ans_small.jpg	0.146	0.028	0.189
case37Rcc_ans_small.jpg	0.026	0.031	1.1988
case37Rmlo_ans_small.jpg	0.188	0.044	0.234
case46Lcc_small.jpg	0.092	0.027	0.2884
case46Lmlo_small.jpg	0.081	0.053	0.6475
case46Rcc_small.jpg	0.372	0.139	0.375
case46Rmlo_small.jpg	0.084	0.057	0.6818
case49Lcc_small.jpg	0.035	0.099	2.849
case49Lmlo_small.jpg	0.126	0.128	1.0112
case49Rcc_small.jpg	0.247	0.119	0.4802
case49Rmlo_small.jpg	0.13	0.143	1.0982
case6Lcc_ans_small.jpg	0.054	0.038	0.7114
case6Lmlo_ans_small.jpg	0.054	0.027	0.4939
case7Lcc_small_ans.jpg	0.085	-0.04	-0.456
case7Lmlo_small_ans.jpg	0.149	-0.03	-0.184
case7Rcc_small_ans.jpg	0.169	-0.01	-0.078
case7Rmlo_small_ans.jpg	0.118	-0.02	-0.15
case17rmlo.jpg	0.04	0.138	3.4712
case19Rcc.jpg	-0.029	0.28	-9.767
case2.jpg	0.359	0.546	1.5205
case21Lcc.jpg	0.051	0.137	2.6645
case21Lmlo.jpg	0.034	0.215	6.3115

case24Lcc.jpg	0.111	0.567	5.1124
case24Lmlo.jpg	0.041	0.174	4.2261
case26Lcc.jpg	8E-05	0.218	2700.7
case26Lmlo.jpg	0.067	0.325	4.8502
case27Rcc.jpg	0.166	0.386	2.3312
case27Rmlo.jpg	0.076	0.222	2.9217
case2lmlo.jpg	0.331	0.508	1.5334
case36Lcc.jpg	0.026	0.305	11.612
case36Lmlo.jpg	-0.039	0.149	-3.809
case39Lcc.jpg	0.089	0.189	2.1171
case39Lmlo.jpg	0.147	0.281	1.904
case44Rcc.jpg	0.076	0.247	3.2484
case44Rmlo.jpg	0.056	0.212	3.7892
case5lcc.jpg	0.05	0.266	5.303
case5lmlo.jpg	0.684	0.913	1.3335

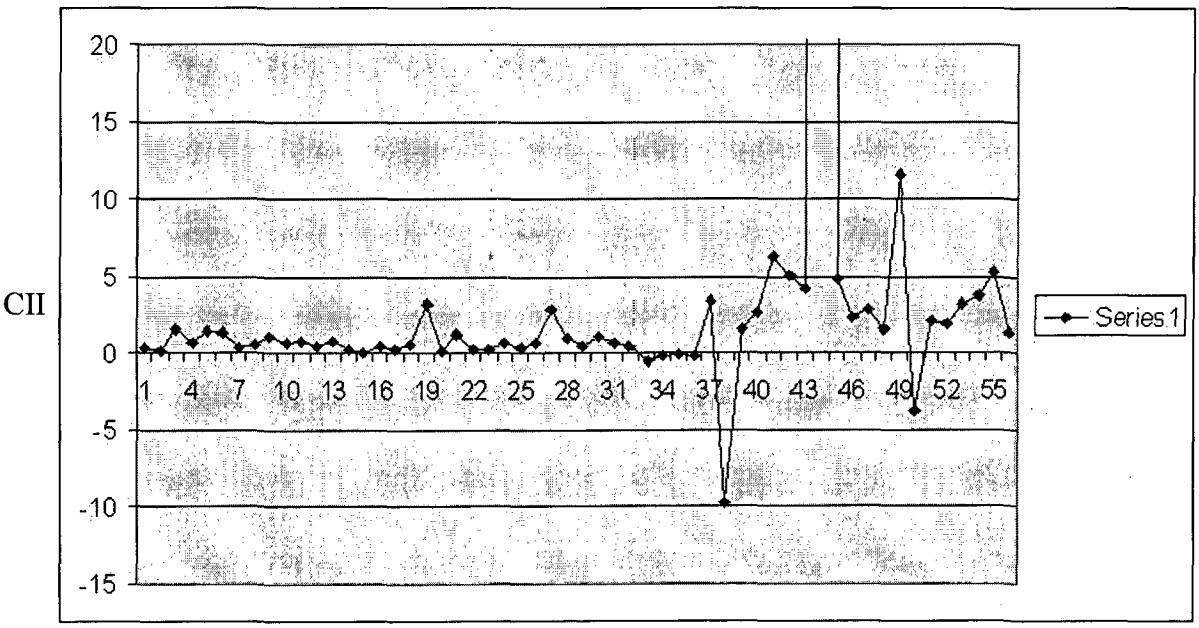


Figure 4.1: Contrast Improvement Index for Top Hat enhancement applied on ROIs of mammograms containing microcalcification

Table 4.2 : DV/BV ratio for Top Hat enhancement applied on ROIs of mammograms containing microcalcification

Name of file	Detail Variance	Background Varince	DV/BV ratio
case11Lcc_small_ans.jpg	112.789	42.0147	2.6845
case11Lmlo_small_ans.jpg	121.417	65.1424	1.8639
case11Rcc_small_ans.jpg	129.839	92.7756	1.3995
case11Rmlo_small_ans.jpg	137.591	96.5753	1.4247
case11_1Lcc_small_ans.jpg	127.428	71.7481	1.776
case11_1Rmlo_small_ans.jpg	125.299	90.3718	1.3865
case15Lcc_small_ans.jpg	124.799	48.8374	2.5554
case15Lmlo_small_ans.jpg	96.4773	47.2637	2.0413
case15Rcc_small_ans.jpg	122.197	72.7615	1.6794
case15Rmlo_small_ans.jpg	108.619	77.9999	1.3926
case15_1Lcc_small_ans.jpg	118.122	71.4966	1.6521
case15_1Lmlo_small_ans.jpg	109.695	46.2515	2.3717
case15_1Rmlo_small_ans.jpg	103.006	56.5842	1.8204
case22Lcc_small.jpg	74.6488	5.06401	14.741
case25Lcc_small_ans.jpg	91.1823	12.3154	7.4039
case25Lmlo_small_ans.jpg	100.365	16.2316	6.1833
case25Rcc_small_ans.jpg	103.018	15.8848	6.4853
case25Rmlo_small_ans.jpg	102.253	13.2831	7.698
case37Lcc_ans_small.jpg	137.584	60.4158	2.2773
case37Lmlo_ans_small.jpg	97.3543	44.4081	2.1923
case37Rcc_ans_small.jpg	74.9959	46.7747	1.6033
case37Rmlo_ans_small.jpg	83.8559	35.4755	2.3638
case46Lcc_small.jpg	38.2776	12.943	2.9574
case46Lmlo_small.jpg	72.2603	32.3076	2.2366
case46Rcc_small.jpg	95.1724	11.9214	7.9833
case46Rmlo_small.jpg	62.8594	20.6982	3.037
case49Lcc_small.jpg	64.1699	39.7372	1.6149
case49Lmlo_small.jpg	114.52	27.1375	4.22
case49Rcc_small.jpg	87.8973	13.9745	6.2899
case49Rmlo_small.jpg	114.311	24.3076	4.7027
case6Lcc_ans_small.jpg	104.873	47.8654	2.191
case6Lmlo_ans_small.jpg	94.8173	59.2539	1.6002
case7Lcc_small_ans.jpg	89.6641	11.4122	7.8569
case7Lmlo_small_ans.jpg	77.7004	14.5601	5.3365
case7Rcc_small_ans.jpg	96.1412	18.3839	5.2296
case7Rmlo_small_ans.jpg	97.3397	26.0569	3.7357
case17rmlo.jpg	91.2343	62.5248	1.4592
case19Rcc.jpg	119.3	57.3138	2.0815

case2.jpg	85.4364	20.9217	4.0836
case21Lcc.jpg	87.5438	61.8988	1.4143
case21Lmlo.jpg	82.6506	48.6177	1.7
case24Lcc.jpg	105.235	33.2119	3.1686
case24Lmlo.jpg	128.833	61.7913	2.085
case26Lcc.jpg	97.4088	62.4301	1.5603
case26Lmlo.jpg	70.9226	34.6238	2.0484
case27Rcc.jpg	86.0027	33.9993	2.5295
case27Rmlo.jpg	83.4195	42.2745	1.9733
case2lmlo.jpg	92.6029	28.4325	3.2569
case36Lcc.jpg	101.881	39.9747	2.5486
case36Lmlo.jpg	95.5003	63.8149	1.4965
case39Lcc.jpg	74.4926	41.5302	1.7937
case39Lmlo.jpg	91.9005	55.9508	1.6425
case44Rcc.jpg	87.9241	38.4506	2.2867
case44Rmlo.jpg	109.041	55.2037	1.9752
case5lcc.jpg	91.0447	41.6364	2.1867
case5lmlo.jpg	83.864	3.62764	23.118

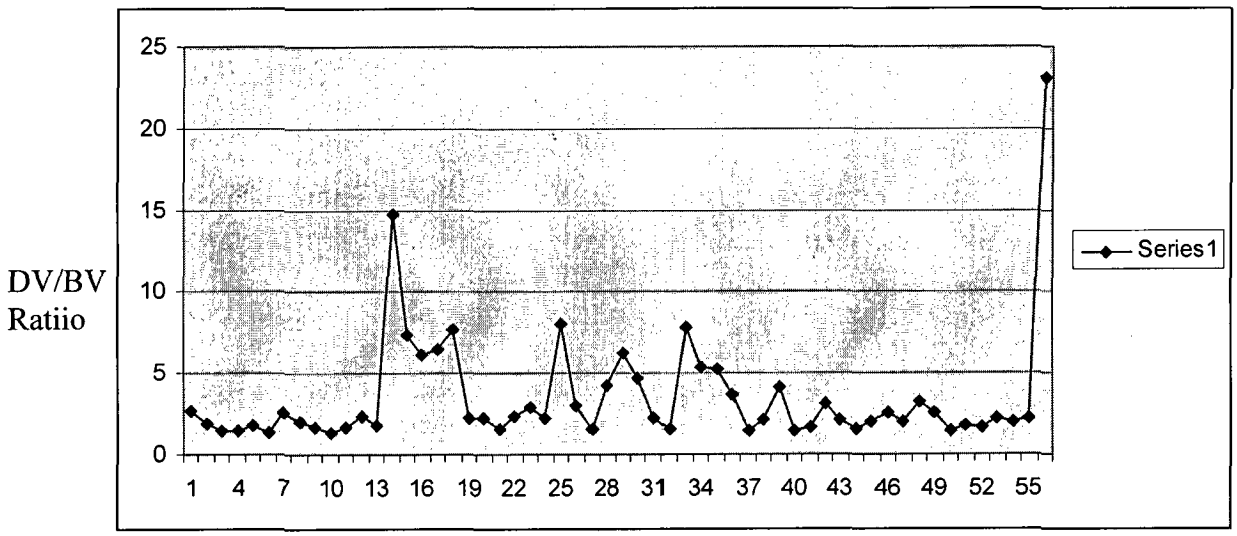


Figure 4.2: DV/BV ratio for Top Hat enhancement applied on ROIs of mammograms containing microcalcification

Chapter 5

Segmentation of Microcalcification

5.1 Introduction

Segmentation refers to the process of partitioning a digital image into multiple regions (sets of pixels). The result of image segmentation is a set of regions that collectively cover the entire image, or a set of contours extracted from the image. Now here our aim is extracting the microcalcification from mammograms for further calculation of features and classification of microcalcification among benign and malignant .

In previous chapters, enhancement algoithm is discussed. The result of enhancement is used as an input for the segmentation stage.

In this chapter , an algorithm for segmentation is discussed. We begin with Basics of segmentation algorithm which discusses the two major categories of segmentation algorithms. Then detection of discontinuities are discuss in short. Edge Detection is discussed in detail with details of different gradient operators. Then finally the block diagram of the algorithm is discussed in detail.

5.2 Basics of Segmentation

Image segmentation algorithms are generally based on one of the two basic properties of intensity values i.e. discontinuity and similarity. In the first category, the approach is to partition an image based on abrupt changes in intensity, such as edges in an image. Point detection, line detection, and edge detection are the example of detection of discontinuity based algorithms[15].

The principal approaches in the second category are based on partitioning an image into regions that are similar according to a set of predefined criteria. Thresholding, region growing, and region splitting and merging are examples of methods in this category.

Algorithm developed for the detection of microcalcification is based on detection of edges and therefore detection of discontinuity is discussed in detail in the chapter.

5.3 Detection of Discontinuity

There are three basic types of gray-level discontinuities that can be detected in the digital image

1. Points.
2. Lines.
3. Edges.

The most common way to look for the discontinuities is to run a mask through the images. A 3 x 3 mask is shown below

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

Figure 5.1: General 3 x 3 mask.

The response of a mask at any point at any point in the image is given by

$$R = w_1 z_1 + w_2 z_2 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots w_9 z_9$$
$$= \sum w_i z_i \qquad \text{where i varies from 1 to 9}$$

where z_i is the gray level of the pixel associated with mask coefficient w_i. As usual the response of the mask is defined with respect to its centre location.

1. Point detection

The detection of isolated points in an image is straightforward in principle. Using the mask shown in figure 5.2

-1	-1	-1
-1	8	-1
-1	-1	-1

Figure 5.2: 3 x 3 point detection mask

A point has been detected at the location on which the mask is centered if

$$|R| \geqslant T$$

Where T is a non negative threshold and R is the result of application of mask on the pixel of the image.This formulation measures the weighted differences between the centre point and its neighbors.

The idea is that an isolated point (a point whose gray level is significantly different from its background and which is located in a homogeneous or nearly homogenous area) will be quite different from its surroundings, and thus be easily detectable by this type of mask.

2. Line detection

After the point detection , next level of complexity is line detection. line in the image can be detected in four direction i.e. horizontal , +45°, vertical, -45°. The mask for detection of line at different angle are shown below.

-1	-1	-1
2	2	2
-1	-1	-1

-1	-1	2
-1	2	-1
2	-1	-1

-1	2	-1
-1	2	-1
-1	2	-1

2	-1	-1
-1	2	-1
-1	-1	2

Figure 5.3: 3 x 3 Line mask : a) Horizonta, b) +45°, c) Vertical, d) -45^0

3. Edge detection

Edge is a set of connected pixels that lie on the boundary between two regions. Edge Detection is the most common approach for the detecting meaningful discontinuities in the gray level. Edges are more closely modeled as having a "ramplike" profile. The slope of the ramp is inversely proportional to the degree of blurring in the edge [15].

First derivative of the ramp like gray profile is positive at the points of transition into and out of the ramp as we move from left to right along the profile. It is constant for points in the ramp, and zero in the areas of constant gray level.

5.4 Gradient Operators

First-order derivatives of digital image are based on various approximations of the 2-D gradient[15][18]. The Gradient of an image f(x,y) at location (x,y) is defined as the

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

Gradient vector is defined by two quantities - magnitude of the gradient vector and direction of the gradient vector. Magnitude of the gradient vector is given by

$$mag(\nabla f) = \left[G_x^2 + G_y^2\right]^{\frac{1}{2}}$$

The direction $\alpha(x, y)$of the gradient vector ∇f at (x,y) is given by

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y}{G_x}\right)$$

Computation of gradiant of an image is based on obtaining the partial derivatives $\frac{\partial f}{\partial x}$and $\frac{\partial f}{\partial y}$at every pixel location.

Roberts cross gradiant operator : One of the simplest way to implement first order derivative

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

-1	0
0	1

0	-1
1	0

Figure 5.4: A 3 x 3 region of an image (the Z's are the gray values) and Roberts gradient mask

$$G_x = (z_9 - z_5) \quad and \quad G_y = (z_8 - z_6)$$

Prewitt gradiant operator:

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

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

Figure 5.5: Prewitt gradient operator

$$G_x = (z_7 + z_8 + z_9) - (z_1 + z_2 + z_3) \quad and \quad G_y = (z_3 + z_6 + z_9) - (z_1 + z_4 + z_7)$$

Sobel Gradiant operator:

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Figure 5.6: Sobel Gradient operator

$$G_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3) \quad and \quad G_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)$$

5.5 Algorithm

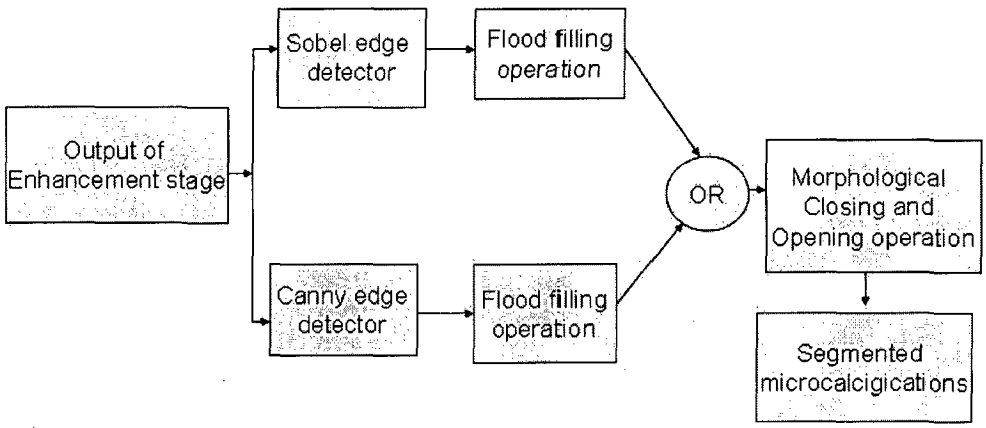


Figure 5.7: Block diagram of mammogram segmentaion algorithm[12]

In the above Block Diagram, segmentation algorithm is divided into 4 steps - edge detection , flood filling operation , logically ORed operation, and noise reduction algorithm.

Step 1 : After top-hat enhancement, the image undergoes Sobel and Canny edge detection to segment the enhanced borders from the background image. The Sobel edge detector applies Sobel approximation to the derivative of the image and detects edges whenever the gradient of reconstructed input image is at its maximum. The Canny edge detector finds edges by looking for local maximum of the gradient of unprocessed input image. In each edge detection algorithm, the gradient is calculated using the derivative of a Gaussian filter, and the output is a binary image, where 1 represents edges and 0 represents background [12].

In each edge detection algorithm, the gradient is calculated using the derivative of a Gaussian filter, and the output is a binary image, where 1 represents edges and 0 represents background.

step 2 : The Sobel and Canny binary output images are post-processed by the flood-filling operation to fill all objects with closed borders. For both binary and grayscale images, the boundary of the fill operation is determined by the connectivity you specify. Here fill operation is done with assuming 8-connectivity.

step 3 : The next step is logically ORed operation. The result of flood filling operation on the output of sobel and canny edge detectors are logically ORed. The result of logically ORed operation is a new image that possibly includes false positive but discourages false negatives.

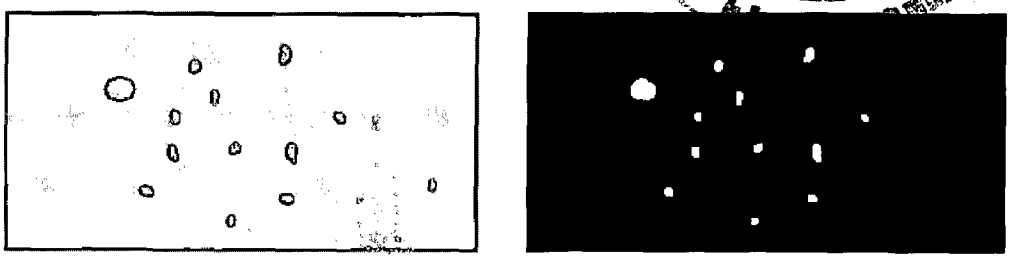
step 4 : The output of logically ORed operation is undergo morphological closing and opening to eliminate noise.

5.6 Results and Discussion

Segmentation algorithm is first applied on synthetic image with known number of micro-calcification and then algorithm is tested on original mammogram ROIs.

Results are divided into two parts, First part shows result of algorithm applied on synthetic images and second part shows result of algorithm applied on original mammogram ROIs.

Part I - Result on Synthetic image

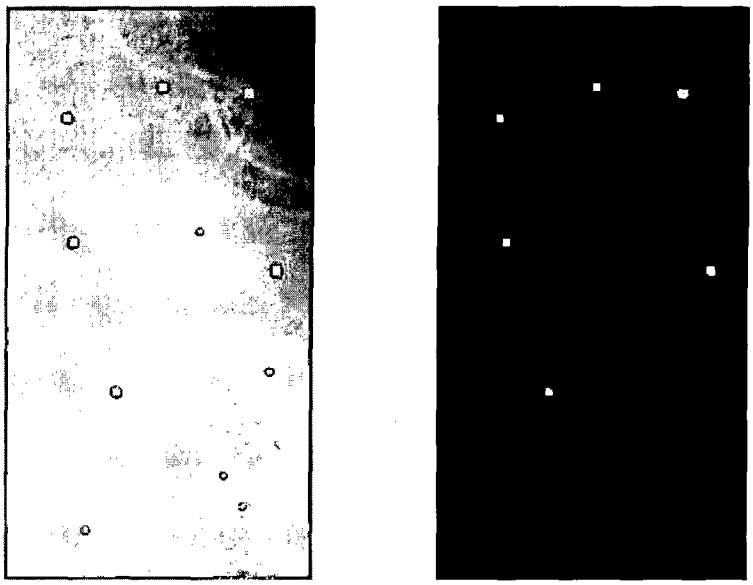


(a) (b)

Figure 5.8: Result on synthetic test image 1 : (a) input image (b) Result of segmentation.

Figure 5.8 (a) shows synthetic image with known microcalcification. Total number of 13 calcification spots are simulated in the image. After application of segmentation algorithm, output is shown in figure 5.8 (b). If we cosely observe the output, total number of simulated calcification in input image and output image is same i.e. 13. The calcifications

are differentiated from background as calcification region is represented by pixel value '1' (maximum) and background region is represented by pixel value of '0' (minimum).



(a) (b)

Figure 5.9: Result on synthetic test image 2 : (a) input image (b)Result of segmentation

In figure 5.9, synthetic image is shown. Total number of 9 calcification spots are simulated in the input image, out of which 6 spots are belong to calcification category as per the size and 3 sopts are very small in size and are treated as noise. In the output image, calcification spots are segmented from the background region. At the same time, the 3 spots which are not belong to calcification category are removed due to application of morphological opening and closing operation.

Part 2 : Result on Original mammogram ROI.

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.

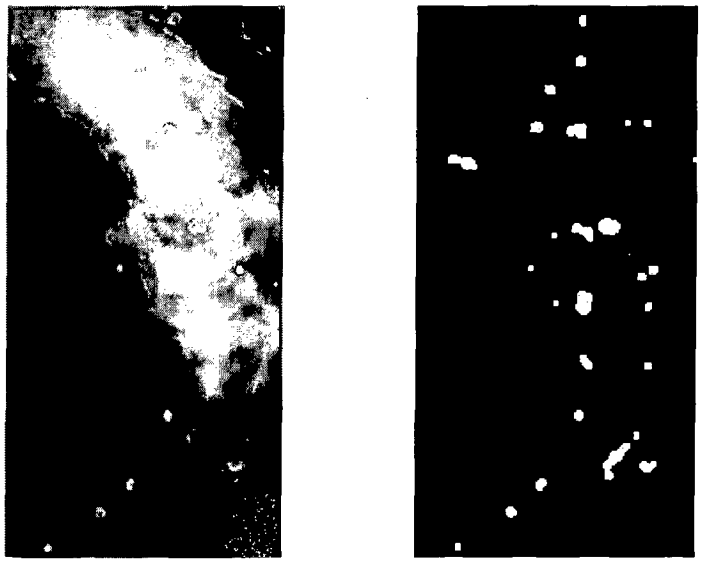


Figure 5.10: Result on original image : (a)case46_Lcc input image (b)Result of segmentation

Figure 5.10 shows ROI of original mammogram image having microcalcification which appear as a bright spots in the image. Some parts of the background , such as dense tissue, appears brighter than the microcalcications in the fatty part of the breast in the ROI, because of this reason it is difficult to differentiate microcalcification from the background region. After applying segmentation algorithm microcalcifications are clearly seen without any noticeble distortion in shape and size.

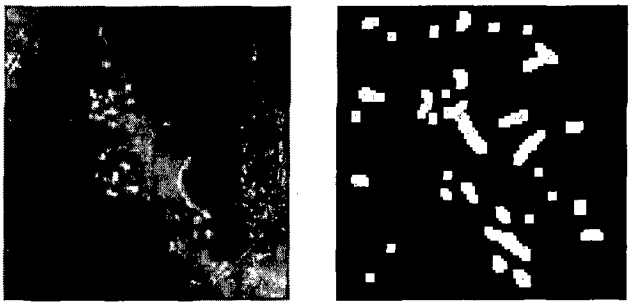


Figure 5.11: Result on original image : (a)case39_Lmlo image (b)Result of segmentation

Similarly in fig. 5.11 microcalcifications are not appear clearly due to bright region because of dense surrounding tissue. After applying algorithm, segmented image is shown in fig. 5.11 (b) showing microcalcification clearly with respect to background.

Chapter 6

Feature extraction

6.1 Introduction

Feature extraction is most important step of the computer aided diagnostic system for classification of microcalcification between benign and malignant. In the segmentation stage, Microcalcificaton are detected as a binary image ('1' represent microcalcification).

This microcalcification varies in size, shape and morphology widely as discussed in detail in chapter 2. The Detected Microcalcification may belongs to one of the two classes benign and malignant. Benign calcification are related to non-cancerous and malignant calcification are related to cancerous.

Now our aim is to extract relevent features from the segmented calcification for better classification among benign and malignant. For this some algorithms are implemented for calculation of features from the mammogram image.

In this chapter, features related to shape, texture, and co-occurrence matrix are discussed in details and also the procedure of labeling of seperated microcalcification.

6.2 Review of Microcalcification

Calcifications are analyzed according to their size, shape, number, and distribution. Detail of various forms of microcalcification classification is shown below [1]

1. ***Size :*** 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.

2. ***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.

3. ***Morphology:*** The morphology of calcifications is considered to be the most important indicator in differentiating benign from malignant. 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.

6.3 Feature Extraction Process

The major steps of feature extraction process are enhancement , segmentation, labeling, and feature extraction [19] as shown in figure 6.1.

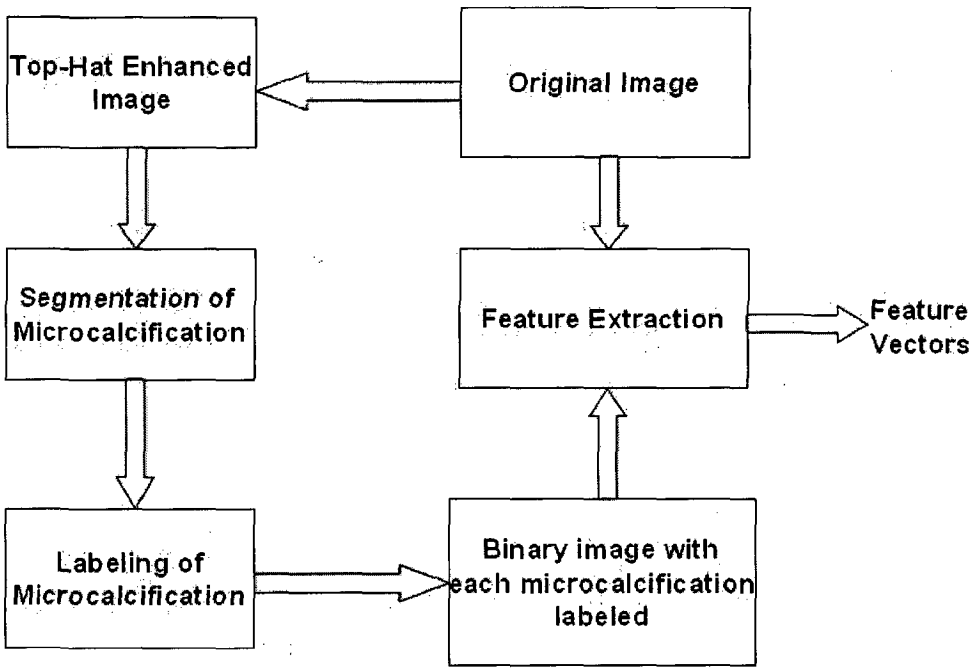


Figure 6.1: Feature extraction process

Steps involved in Feature extraction process are as follow

1. Enhancement stage involves the enhancement of the original image for increasing the contrast of the calcification and at the same time suppressing the background contrast. The algorithm developed for this stage is discussed in chapter 3.

2. Segmentation stage provides the binary image based on the output of the enhanced image. output of segmentation image is such that the calcification are represented by pixel value '1' and background information is represented by pixel value '0'. The algorithm developed for this stage is discussed in chapter 5.

3. After segmentation the next task is to labeled each calcification in the image with a labeled number start from a value of 1 and at the same time object reduction are applied such that those calcification whose area is less than 2 (that is calcification consist of only 2 pixels) are neglected and are not used in feature extraction process further.

4. After labeling , feature for each labeled calcification are calculated.

6.4 Labeling of Microcalcification and Object Reduction

The next step in our feature extracting procedure is labeling of the separate microcalcifications. This is done for two reasons: determining the number of calcifications in a cluster, and gaining the ability to work with individual microcalcifications. For this step, the gray-level segmented images, obtained by the above method, are transformed to binary. For extracting features related to the shape of the calcifications, we work on these binary images. The steps for labeling procedure is described below [20].

1. Starting from the upper left corner of the image, the labeling pixel x(i,j) moves from left to right scanning each row of the image matrix.

2. The first non-zero pixel reached by the labeling pixel is labeled 1.

3. Now when reaching next pixel, x(i,j), with a value of one (part of a microcalcification), the pixels x(i,j-1), x(i-1,j) and x(i-1,j-1) are checked out, if they all have a zero value, the label variable is incremented and a new label is assigned to x(i,j), otherwise x(i,j) is labeled the same as them.

4. When the labeling pixel reaches the bottom right corner of the image, all of the non-zero pixels in the image have been labeled.

5. If all the islands in the image have a convex shape, the labeling procedure can end here. However, due to the irregular shapes of some of the microcalcifications, parts of some of them might be labeled differently.

6. For overcoming this problem, a new labeling pixel starts moving from the lower right corner of the image.

7. When it reaches a labeled pixel, sets its label to the maximum label value found in the 8 neighboring pixels. This procedure is repeated from the other two corners of the image.

Considering all possible shapes for the micrcalcifications, this procedure ensures that all of the pixels belonging to one microcalcification have been labeled the same. Finally, the

label numbers are sorted and renamed so that the labels are sequential and in order. The largest label represents the number of microcalcifications in the cluster.

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 microcalcications are typically clustered.

0	0	0	1	1	0	0	0	0	0
0	0	0	1	1	1	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	0	1	1	0	0
0	1	0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0	0	0
0	0	1	1	1	1	0	0	0	1
0	0	0	1	1	0	0	0	0	1

0	0	0	1	1	0	0	0	0	0
0	0	0	1	1	1	0	0	0	0
0	2	0	0	0	0	0	0	0	0
0	2	2	0	0	0	0	0	0	0
0	0	0	0	0	3	3	3	0	0
0	0	0	0	0	0	3	3	0	0
0	4	0	0	0	0	0	0	0	0
0	0	0	5	5	5	0	0	0	0
0	0	5	5	5	5	0	0	0	6
0	0	0	5	5	0	0	0	0	6

Figure 6.2: Binary image matix and Labeled matirx

0	0	0	1	1	0	0	0	0	0
0	0	0	1	1	1	0	0	0	0
0	2	0	0	0	0	0	0	0	0
0	2	2	0	0	0	0	0	0	0
0	0	0	0	0	3	3	3	0	0
0	0	0	0	0	0	3	3	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	5	5	5	0	0	0	0
0	0	5	5	5	5	0	0	0	0
0	0	0	5	5	0	0	0	0	0

Figure 6.3: Labeled image matrix after object reduction: pixel label 4 & 6 are ignored.

6.5 Feature Extraction

Features of the micorcalcification are broadly divided into three categories

1. Shape based features
2. Texture based features

3. Gray level Co-occurrence matrix based features

6.5.1 Shape based feature

As discussed in section 6.2, 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 mirocalcification as a measure of its size[20].

1. ***Compactness :*** One of the shape features that has proven to be a good measure for classifying microcalcifications by their shape is compactness.12 Compactness (C) is defined as the ratio of the squared perimeter (P) to the area (A), i.e.,

$$C = \frac{P^2}{A}$$

   A 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.

2. ***Area :*** Area is defined as the actual number of pixels in the region.

3. ***Equivalent diameter :*** Equivalent diameter is defined as the diameter of a circle with the same area as the region. It is computed as

$$Equidiameter = \sqrt{(4 * area)/\pi}$$

   where *area* is the actual number of pixels in the region.

4. ***Major axis length :*** The length of the major axis of the ellipse that has the same normalized second central moments as the region. In other words, it is defined as the major axis length of an ellipse which is fit to the region.

5. ***Minor axis length :*** The length of the minor axis of the ellipse that has the same normalized second central moments as the region. In other words, it is defined as the minor axis length of an ellipse which is fit to the region.

6. ***Eccentricity :*** The eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. An ellipse is fit to the region of calcification and the eccentricity of of that ellipse is used as feature. The eccentricity e is defined as :

$$e = \sqrt{1 - \frac{b^2}{a^2}}$$

   where b is the minor axis and a is the major axis lenght. Now mojor and minor axis length is equivalent to the major and minor axis length of the ellipse whose secod central moments are same as that of the region.

6.5.2 Texture based features

Texture information plays an important role in image analysis and understanding, with potential applications in remote sensing, quality control, and medical diagnosis. Texture is one of the important characteristics used in identifying an object or a region of interest (ROI) in an image[15].

First order histogram based features

One of the simplest approaches for describing texture is to use statistical moments of the gray-level histogram of an image or region. Let z be a random variable denoting gray levels and $p(z_i)$, i=0,1,2,.....L-1, be the corresponding histogram, where L is the number of distinct gray levels.

the n^{th}moment of Z about the mean is given by

$$\mu_n(z) = \sum (z_i - m)^n p(z_i)$$

where m is the mean value of z (the average gray level) and is given by

$$m = \sum z_i p(z_i)$$

1. ***Standard Deviation :*** It is a measure of gray level contrast. It is defined as square root of second order moment and is given by

$$\sigma = \sqrt{\sum (z_i - m)^2 p(z_i)}$$

2. ***Relative smoothness :*** Relative smoothness is defined as

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

it is 0 for areas of constant intensity (the variance is zero there) and approaches 1 for larger values of $\sigma^2(z)$. As variance values tends to be large for gray scale images with values in the range of 0 to 255, it is a good idea to normalize variance to the interval [0,1] by dividing $\sigma^2(z)$ by (L-1)^2.

3. ***Skewness :*** Third moment is a measure of skewness of the histogram.

$$\mu_3(z) = \sum (z_i - m)^3 p(z_i)$$

This measure is 0 for symmetric histogram, and have positive values for histogram skewed to the right (about the mean) and negative for histogram skewed to the left. Values of this measure are brought in the range comparable to other five measures $\mu_3(z)$ by (L-1)^2.

4. ***Uniformity :*** It is defined as

$$U = \sum p^2(z_i)$$

This measure is 0 for symmetric histogram, and have positive values for histogram skewed to the right (about the mean) and negative for histogram skewed to the left. Values of this measure are brought in the range comparable to other five measures $\mu_3(z)$ by (L-1)^2.

5. ***Uniformity :*** It is defined as

$$e = -\sum p(z_i) \log_2 p(z_i)$$

6.5.3 Co-occurrence matrix based texture feature

Most of the published work on microcalcification classification used co-occurrence matrices to describe textural properties. The gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The co-occurrence matrix is based on the repeated occurrence of gray-level configuration in the texture.The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image [22].

Process to create GLCM

A gray-level co-occurrence matrix (GLCM) contains information about how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. The spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but other spatial relationships between the two pixels can also be specified. Each element (i,j) in the resultant glcm is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image [22].

The number of gray levels in the image determines the size of the GLCM .The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image. For example, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse with respect to the specified offset.

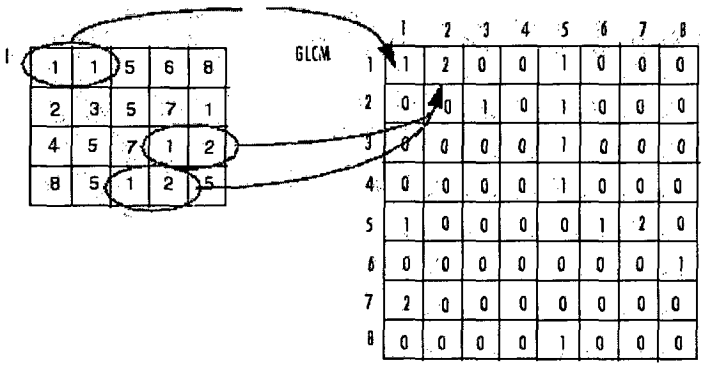


Figure 6.4: Gray level co-occurrence matrix [22]

Other spatial relationships

As GLCM specifies the spatial relationship between the pixels. we have different offsets for specifying different spatial relationships. These offsets define pixel relationships of varying direction and distance. For example, you can define an array of offsets that specify four directions (horizontal, vertical, and two diagonals) and four distances. In this case, the input image is represented by 16 GLCMs. When you calculate statistics from these GLCMs, you can take the average.

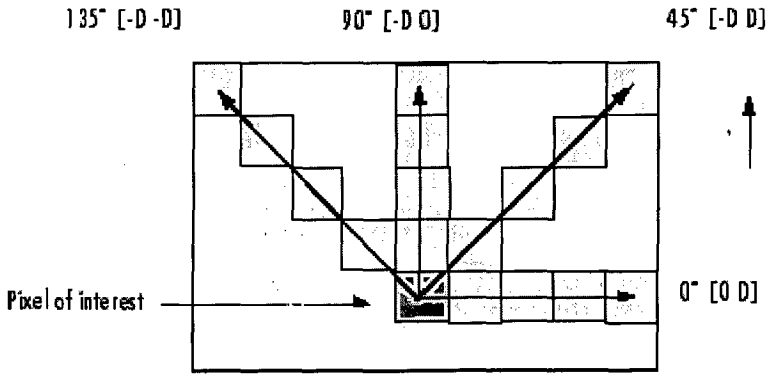


Figure 6.5: GLCM offsets (4 direction and 4 distances for each direction)[22]

GLCM Based Properties

Based on the Gray Level Co-occurrence matrix defined above some properties are derived from the matrix i.e. *Contrast, Homogeneity, Energy, Correlation* [22].

1. Contrast : It measures the local variations in the gray-level co-occurrence matrix i.e a measure of the intensity contrast between a pixel and its neighbor over the whole

image. Contrast is 0 for a constant image.

$$contrast = \sum |i-j|^2 p(i,j)$$

where i,j are the x and y position in GLCM.

2. Correlation : It measures the joint probability occurrence of the specified pixel pairs. It is a measure of how correlated a pixel is to its neighbor over the whole image.

$$\sum \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j}$$

Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN for a constant image.

3. Energy : It provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment.

$$\sum p(i,j)^2$$

Energy is 1 for a constant image.

4. Homogeneity : It measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$\sum \frac{p(i,j)}{1+|i-j|}$$

Homogeneity is 1 for a diagonal GLCM.

6.6 Results and Discussions

As feature extraction process is discussed in section 6.3, original image is first cropped as per the region of interest and then applying the enhancement and segmentation algorithm discussed in chapter 3 and 5 respectively.

From the segmented image following features are calculated

a) Shape based features

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

6. Eccentricity

b) First order histogram based texture feature

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

c) Gray co-occurrence matrix based texture features (offset - 0°, 45°, 90°, 135°)

1. Contrast
2. Correlation
3. Energy
4. Homogeneity

Table 6.1 enlists all shape based features i.e. Compactness, area, equivalent diameter, major axis length, minor axis length, eccentricity.

Table 6.2 enlists all first order histogram based texture features i.e. mean, standard deviation, relative smoothness, skewness, uniformity, entropy.

Table 6.3 , 6.4 ,6.5, 6.6 enlists gray co-occurrence matrix based texture features i.e contrast, correlation, energy, homogeinity respectively. Features calculated for all four direction of the gray level spatial relationship i.e. 0°, 45°, 90°, 135°

Table : 6.1 Shape based Features - area, eccentricity, equivalent diameter, major axis length, Minor axis length, compactness

Name of file	Label no.	Area	Eccen.	Equi. Dia.	Major AL	Minor AL	Compactness
case11Lcc_small	1	10363	0.51475	114.8677	128.0889	109.8159	37.477768
case11Lcc_small	2	4	0.94629	2.256758	4.820546	1.558527	11.656854
case11Lcc_small	3	21	0.89681	5.170883	8.935666	3.9532	24.075613
case11Lcc_small	5	5	0.57602	2.523133	3.40196	2.780887	9.3254834
case11Lcc_small	7	4	0	2.256758	2.309401	2.309401	4
case11Lcc_small	9	40	0.93623	7.136496	13.94441	4.900001	27.745079
case11Lcc_small	12	3	0.94281	1.95441	3.464102	1.154701	5.3333333
case11Lcc_small	13	5	0.76509	2.523133	3.306559	2.129163	5.8627417
case11Lcc_small	15	8	0.86603	3.191538	4.618802	2.309401	8
case11Lmlo_small	1	4538	0.50514	76.0129	83.68567	72.22363	23.332567
case11Lmlo_small	2	3	0.92899	1.95441	3.854129	1.426385	7.7712362
case11Lmlo_small	3	9	0.89399	3.385138	6.804941	3.049216	25.237454
case11Lmlo_small	4	3	0.94281	1.95441	3.464102	1.154701	5.3333333
case11Rcc_small	1	48	0.59122	7.81764	8.856048	7.142468	11.899705
case11Rcc_small	3	3	0.7303	1.95441	2.581989	1.763834	3.8856181
case11Rcc_small	5	10	0.58741	3.568248	4.085748	3.306559	7.7941125
case11Rcc_small	6	3	0.7303	1.95441	2.581989	1.763834	3.8856181
case11Rcc_small	7	9	0.90794	3.385138	5.399179	2.262845	9.8474908
case11Rcc_small	8	6	0.98601	2.763953	6.928203	1.154701	16.666667
case11Rmlo_small	1	804	0.28279	31.99507	33.58353	32.2127	15.049751
case15Lcc_small	1	9	0.98225	3.385138	9.031889	1.69394	24.431361
case15Lcc_small	2	71	0.88672	9.507892	17.32351	8.008861	32.97305
case15Lcc_small	3	1084	0.55088	37.15093	45.65043	38.09928	65.803294
case15Lcc_small	4	74	0.5192	9.706685	11.43246	9.770809	25.519059
case15Lcc_small	5	6	0.86367	2.763953	4.687529	2.362758	13.714045
case15Lcc_small	6	8	0.87821	3.191538	5.518244	2.639252	15.321068
case15Lcc_small	8	3	0.7698	1.95441	3.464102	2.211083	10.666667
case15Lcc_small	9	42	0.71373	7.312733	10.68747	7.485753	37.254875
case15Lcc_small	13	63	0.92805	8.956232	17.50278	6.51906	36.265769
case15Lcc_small	14	8	0.73488	3.191538	5.062474	3.43337	21.356602
case15Lcc_small	16	18	0.94923	4.787307	9.237604	2.905933	18.983646
case15Lmlo_small	1	34	0.68362	6.579525	8.053221	5.877573	14.105526
case15Lmlo_small	2	41	0.90574	7.225152	13.68545	5.800457	40.25822

Table : 6.2 First order histogram based texture feature- mean, Variance relative smoothness, skewness, uniformity, entropy

Name of file	Label no.	Mean	variance	Relative smoothness	Skewness	Uniformity	Entropy
case11Lcc_small	1	0.588	0.0177	0.01739	6.457E-05	0.0088	7.1043
case11Lcc_small	2	0.475	0.00021	0.00021	-1.72E-06	0.1563	53.375
case11Lcc_small	3	0.504	0.00109	0.00109	-9.44E-06	0.0679	22.764
case11Lcc_small	5	0.466	0.00015	0.00015	9.978E-07	0.1852	50.442
case11Lcc_small	7	0.553	1.9E-05	1.9E-05	0	0.25	58.5
case11Lcc_small	9	0.488	0.0018	0.00179	-2.43E-05	0.0366	14.857
case11Lcc_small	12	0.469	3.1E-05	3.1E-05	1.192E-07	0.5556	72.493
case11Lcc_small	13	0.467	0.00022	0.00022	-1.94E-06	0.2222	62.234
case11Lcc_small	15	0.482	0.00012	0.00012	-8.94E-08	0.1875	52.25
case11Lmlo_small	1	0.715	0.0169	0.01662	-0.000911	0.009	7.0536
case11Lmlo_small	2	0.499	0.00016	0.00016	4.327E-07	0.1667	57.731
case11Lmlo_small	3	0.517	0.00089	0.00089	-1.42E-05	0.0926	33.805
case11Lmlo_small	4	0.581	6.4E-05	6.4E-05	-1.24E-07	0.3333	58.644
case11Rcc_small	1	0.541	0.0092	0.00911	-0.000593	0.0269	17.621
case11Rcc_small	3	0.5	7.6E-05	7.6E-05	0	0.25	66
case11Rcc_small	5	0.533	0.00089	0.00089	-3.99E-05	0.1111	42.387
case11Rcc_small	6	0.495	8.7E-05	8.7E-05	1.509E-07	0.25	66.5
case11Rcc_small	7	0.565	0.00195	0.00195	-4.86E-05	0.1	44.514
case11Rcc_small	8	0.556	0.00029	0.00029	-1.96E-06	0.2778	63.773
case11Rmlo_small	1	0.736	0.0104	0.01029	0.0002686	0.0117	6.8676
case15Lcc_small	1	0.561	0.00293	0.00292	1.637E-05	0.0703	34.375
case15Lcc_small	2	0.505	0.00278	0.00278	-8.4E-06	0.0271	10.319
case15Lcc_small	3	0.579	0.03785	0.03647	0.0052388	0.0081	7.3236
case15Lcc_small	4	0.495	0.00095	0.00095	3.673E-06	0.045	11.59
case15Lcc_small	5	0.508	0.00162	0.00162	-4.13E-05	0.1111	44.312
case15Lcc_small	6	0.474	8.4E-05	8.4E-05	-3.42E-07	0.1406	36.031
case15Lcc_small	8	0.533	0.00118	0.00118	2.015E-05	0.1667	56.438
case15Lcc_small	9	0.515	0.00343	0.00342	4.167E-05	0.0318	14.974
case15Lcc_small	13	0.518	0.00291	0.0029	9.668E-05	0.03	11.641
case15Lcc_small	14	0.521	0.00219	0.00218	-1.13E-05	0.0938	50.75
case15Lcc_small	16	0.464	0.00157	0.00157	-7.44E-06	0.0672	28.911
case15Lmlo_small	1	0.508	0.00185	0.00184	2.231E-05	0.0408	19.017
case15Lmlo_small	2	0.494	0.00246	0.00245	-4.48E-05	0.0302	15.292
case15Lmlo_small	3	0.609	0.00039	0.00039	-1.09E-05	0.44	51.502
case15Lmlo_small	4	0.479	0.00692	0.00687	0.0004772	0.0155	6.8783
case15Lmlo_small	5	0.471	0.0006	0.00059	2.57E-06	0.1563	55.125
case15Lmlo_small	6	0.483	1E-05	1E-05	1.676E-08	0.375	67
case15Lmlo_small	7	0.486	0.00568	0.00565	0.0003359	0.0237	10.925
case15Lmlo_small	8	0.505	0.00152	0.00152	-2.24E-05	0.0933	40.182

Table : 6.3 GLCM based texture feature- Contrast for four directions 0 deg, 45 deg, 90 deg, 135 deg

Name of file	Label no.	Distance	0 deg	45 deg	90 deg	135 deg
case11Lcc_small	1	1	75.1273	140.675	100.185	134.79
case11Lcc_small	2	1	19	43	49.75	57
case11Lcc_small	3	1	35.75	196.208	171.407	165.833
case11Lcc_small	5	1	6.5	24.5	18.6667	24.5
case11Lcc_small	7	1	1	9	4	1
case11Lcc_small	9	1	56.013	115.682	121.472	160.833
case11Lcc_small	12	1	9	0	0	0
case11Lcc_small	13	1	23.25	8	5.66667	48.5
case11Lcc_small	15	1	4.5	0.33333	6	19
case11Lmlo_small	1	1	153.885	249.315	192.795	231.894
case11Lmlo_small	2	1	32.25	9	12.6667	50.5
case11Lmlo_small	3	1	142.5	104.5	49.6667	129.4
case11Lmlo_small	4	1	0	0	14.5	0
case11Rcc_small	1	1	340.393	750.939	330.018	361.469
case11Rcc_small	3	1	16	4	4	36
case11Rcc_small	5	1	29.75	127	72.4444	37.5
case11Rcc_small	6	1	22.5	25	2.5	16
case11Rcc_small	7	1	30.6	130.25	146.75	217.25
case11Rcc_small	8	1	0	0	48.6	0
case11Rmlo_small	1	1	317.399	465.398	353.284	522.298
case15Lcc_small	1	1	580.75	457.571	51.6429	687.714
case15Lcc_small	2	1	116.294	168.465	96.7875	121.882
case15Lcc_small	3	1	385.675	691.394	506.349	775.589
case15Lcc_small	4	1	29.7545	28.5657	35.8704	68.6869
case15Lcc_small	5	1	144.667	248	124.5	29
case15Lcc_small	6	1	3.91667	9.44444	5	3.66667
case15Lcc_small	8	1	100.667	100	206.5	113
case15Lcc_small	9	1	88.7375	288.957	286.636	343.571
case15Lcc_small	13	1	163.75	311.952	234.5	287.838
case15Lcc_small	14	1	176.667	156.111	258	180.556
case15Lcc_small	16	1	157.5	230.375	56.0833	148.688
case15Lmlo_small	1	1	43.5238	124.429	41.8	49.8571
case15Lmlo_small	2	1	158.785	202.45	95.7361	125

Table : 6.4 GLCM based texture feature- Correlation for four directions 0 deg, 45 deg, 90 deg, 135 deg

Name of file	Label no.	Distance	0 deg	45 deg	90 deg	135 deg
case11Lcc_small	1	1	0.967517	0.938208	0.956537	0.94126
case11Lcc_small	2	1	0.458266	0.866025	-0.88153	-0.12726
case11Lcc_small	3	1	0.766257	-0.40876	-0.30768	-0.24771
case11Lcc_small	5	1	0.666458	-0.5873	0.301625	-0.17916
case11Lcc_small	7	1	1	0	1	0
case11Lcc_small	9	1	0.75671	0.426176	0.442774	0.240866
case11Lcc_small	12	1	-1	0	0	0
case11Lcc_small	13	1	0.911465	1	0.999569	1
case11Lcc_small	15	1	0.965616	0.986241	0.939024	0.986241
case11Lmlo_small	1	1	0.930281	0.886262	0.912072	0.892854
case11Lmlo_small	2	1	-0.44899	1	0.466321	-1
case11Lmlo_small	3	1	-0.48987	0.046407	0.560386	-0.66494
case11Lmlo_small	4	1	0	0	-1	0
case11Rcc_small	1	1	0.699288	0.28408	0.709979	0.634428
case11Rcc_small	3	1	1	0	1	0
case11Rcc_small	5	1	0.888624	0.64626	0.69346	0.297575
case11Rcc_small	6	1	-1	0	1	0
case11Rcc_small	7	1	0.995471	0.364845	0.455257	0.479607
case11Rcc_small	8	1	0	0	-0.52728	0
case11Rmlo_small	1	1	0.76688	0.659409	0.743096	0.618956
case15Lcc_small	1	1	-0.35906	-0.12593	0.858629	-0.73166
case15Lcc_small	2	1	0.688623	0.546538	0.736921	0.675728
case15Lcc_small	3	1	0.922443	0.861091	0.898463	0.84486
case15Lcc_small	4	1	0.760571	0.756221	0.710897	0.409264
case15Lcc_small	5	1	0.131807	-0.38383	0.308051	0.74851
case15Lcc_small	6	1	0.644718	2.75E-17	0.539142	0.763399
case15Lcc_small	8	1	0.448733	-1	-0.45367	1
case15Lcc_small	9	1	0.80393	0.336712	0.349948	0.214444
case15Lcc_small	13	1	0.556524	0.208149	0.420203	0.257556
case15Lcc_small	14	1	0.240353	0.447994	0.033019	-0.0452
case15Lcc_small	16	1	0.225385	-0.12094	0.717322	0.105706
case15Lmlo_small	1	1	0.818948	0.495022	0.89089	0.851039
case15Lmlo_small	2	1	0.497613	0.348	0.694681	0.54994
case15Lmlo_small	3	1	0	0	0.26029	0
case15Lmlo_small	4	1	0.844853	0.761595	0.882027	0.777052

Table : 6.5 GLCM based texture feature- Energy for four directio
0 deg, 45 deg, 90 deg, 135 deg

Name of file	Label no.	Distance	0 deg	45 deg	90 deg	135 deg
case11Lcc_small	1	1	0.000537	0.000395	0.000494	0.000428
case11Lcc_small	2	1	0.166667	0.333333	0.25	0.333333
case11Lcc_small	3	1	0.039063	0.045139	0.039781	0.045139
case11Lcc_small	5	1	0.166667	0.25	0.166667	0.25
case11Lcc_small	7	1	0.5	1	0.5	1
case11Lcc_small	9	1	0.013999	0.016529	0.014275	0.015611
case11Lcc_small	12	1	0.5	0	0	0
case11Lcc_small	13	1	0.25	0.5	0.333333	0.5
case11Lcc_small	15	1	0.25	0.333333	0.222222	0.333333
case11Lmlo_small	1	1	0.000507	0.000417	0.000469	0.00043
case11Lmlo_small	2	1	0.25	0.5	0.333333	0.5
case11Lmlo_small	3	1	0.083333	0.1	0.066667	0.1
case11Lmlo_small	4	1	0	0	0.5	0
case11Rcc_small	1	1	0.019133	0.020408	0.017857	0.021241
case11Rcc_small	3	1	0.5	1	0.5	1
case11Rcc_small	5	1	0.125	0.166667	0.111111	0.166667
case11Rcc_small	6	1	0.5	1	0.5	1
case11Rcc_small	7	1	0.2	0.25	0.125	0.25
case11Rcc_small	8	1	0	0	0.2	0
case11Rmlo_small	1	1	0.001513	0.001494	0.001473	0.001592
case15Lcc_small	1	1	0.125	0.142857	0.071429	0.142857
case15Lcc_small	2	1	0.007219	0.007234	0.006719	0.007716
case15Lcc_small	3	1	0.000685	0.000709	0.000753	0.000667
case15Lcc_small	4	1	0.011901	0.012142	0.011317	0.011529
case15Lcc_small	5	1	0.111111	0.166667	0.125	0.166667
case15Lcc_small	6	1	0.083333	0.111111	0.083333	0.111111
case15Lcc_small	8	1	0.333333	0.5	0.25	0.5
case15Lcc_small	9	1	0.012813	0.014286	0.013324	0.015102
case15Lcc_small	13	1	0.008611	0.010068	0.009088	0.010068
case15Lcc_small	14	1	0.083333	0.111111	0.083333	0.111111
case15Lcc_small	16	1	0.055556	0.0625	0.045139	0.0625
case15Lmlo_small	1	1	0.024943	0.030204	0.02625	0.030204
case15Lmlo_small	2	1	0.015385	0.017778	0.014275	0.016667
case15Lmlo_small	3	1	0	0	0.25	0

Figure 6.6 GLCM based texture feature- Homogeneity for four directions 0 deg, 45 deg, 90 deg, 135 deg

Name of file	Label no.	Distance	0 deg	45 deg	90 deg	135 deg
case11Lcc_small	1	1	0.293799	0.237241	0.282259	0.245173
case11Lcc_small	2	1	0.274074	0.240741	0.172727	0.138889
case11Lcc_small	3	1	0.336485	0.177686	0.189358	0.167067
case11Lcc_small	5	1	0.291667	0.254167	0.405754	0.202778
case11Lcc_small	7	1	0.5	0.25	0.333333	0.5
case11Lcc_small	9	1	0.244669	0.211326	0.148653	0.156946
case11Lcc_small	12	1	0.25	0	0	0
case11Lcc_small	13	1	0.233631	0.6	0.305556	0.15
case11Lcc_small	15	1	0.354167	0.833333	0.319444	0.188889
case11Lmlo_small	1	1	0.24155	0.183642	0.220123	0.202096
case11Lmlo_small	2	1	0.188492	0.25	0.25	0.295455
case11Lmlo_small	3	1	0.189462	0.171213	0.2495	0.210428
case11Lmlo_small	4	1	0	0	0.25	0
case11Rcc_small	1	1	0.19121	0.096563	0.161243	0.172229
case11Rcc_small	3	1	0.2	0.333333	0.333333	0.142857
case11Rcc_small	5	1	0.388095	0.179226	0.227513	0.251263
case11Rcc_small	6	1	0.196429	0.166667	0.416667	0.2
case11Rcc_small	7	1	0.164286	0.110885	0.155531	0.101408
case11Rcc_small	8	1	0	0	0.307051	0
case11Rmlo_small	1	1	0.179652	0.146471	0.170035	0.152589
case15Lcc_small	1	1	0.075375	0.117893	0.204039	0.060354
case15Lcc_small	2	1	0.249579	0.148785	0.2118	0.174109
case15Lcc_small	3	1	0.182545	0.174311	0.20539	0.15408
case15Lcc_small	4	1	0.29515	0.293303	0.303162	0.25558
case15Lcc_small	5	1	0.117339	0.123611	0.165405	0.332407
case15Lcc_small	6	1	0.488889	0.418519	0.436111	0.444444
case15Lcc_small	8	1	0.091414	0.2	0.108226	0.28125
case15Lcc_small	9	1	0.187153	0.132333	0.110044	0.130602
case15Lcc_small	13	1	0.196062	0.148689	0.171272	0.139836
case15Lcc_small	14	1	0.205455	0.12143	0.100533	0.12242
case15Lcc_small	16	1	0.246901	0.095092	0.20235	0.293332
case15Lmlo_small	1	1	0.212793	0.182977	0.269415	0.194092
case15Lmlo_small	2	1	0.181832	0.155445	0.197515	0.270263
case15Lmlo_small	3	1	0	0	0.435897	0
case15Lmlo_small	4	1	0.22299	0.17814	0.21532	0.183725

Chapter 7

Classification of Microcalcifications

7.1 Introduction

The problem of mammogram interpretation can be decomposed into two sub-problems. The first deals with the detection and localization of regions of interest (ROIs) containing microcalcifications. The second, and more difficult problem, is the characterization of the suspected microcalcification as malignant and benign.

Detection of microcalcification is already done and discussed in chapter 5. now our next approach is to implement a classifier which gives a better accuracy to distinguish between malignant and benign.

In this chapter, Support Vector Machine based classification algorithm is discussed.

7.2 Two Class Support Vector Machine

SVM is a constructive learning procedure rooted in statistical learning theory. It is based on the principle of structural risk minimization, which aims at minimizing the bound on the generalization error (i.e., error made by the learning machine on data unseen during training) rather than minimizing the mean square error over the data set. As a result, an SVM tends to perform well when applied to data outside the training set. SVM has also been found to outperform competing methods in several real-world applications [23][24].

Support vector machine are classified into two types - hard margin support vector machine and soft margin support vector machine. In hard-margin support vector machines training data are linearly separable in the input space. Soft-margin support vector machines is used when training data are linearly inseparable in the input space.

In our model for SVM, the soft margin technique is used to maximize the margin in the feature space. To simplify the computation, the feature space is taken to be the non-linear projection of the input data to a higher dimensional space by a kernel function.

7.3 Soft Margin Support Vector Machine

In hard-margin support vector machines, we assumed that the training data are linearly separable. When the data are linearly inseparable, there is no feasible solution, and the hard-margin support vector machine is unsolvable[24][25].

The decision function for classification when the training data are linearly separable is given by

$$D(x) = \dot{w^T}x + b$$

where x_i is m-dimensional training inputs (i = 1,..... , M) belong to Class 1 or 2 and the associated labels be $y_i = 1$ for Class 1 and -1 for Class 2, and w is an m-dimensional vector, b is a bias term.

The SVM classification function for the linearly separable training data follows the following inequality

$$w^T x + b = \begin{cases} \geq 1 & for\ y_{i=1} \\ \leq -1 & for\ y_{i=-1} \end{cases}$$

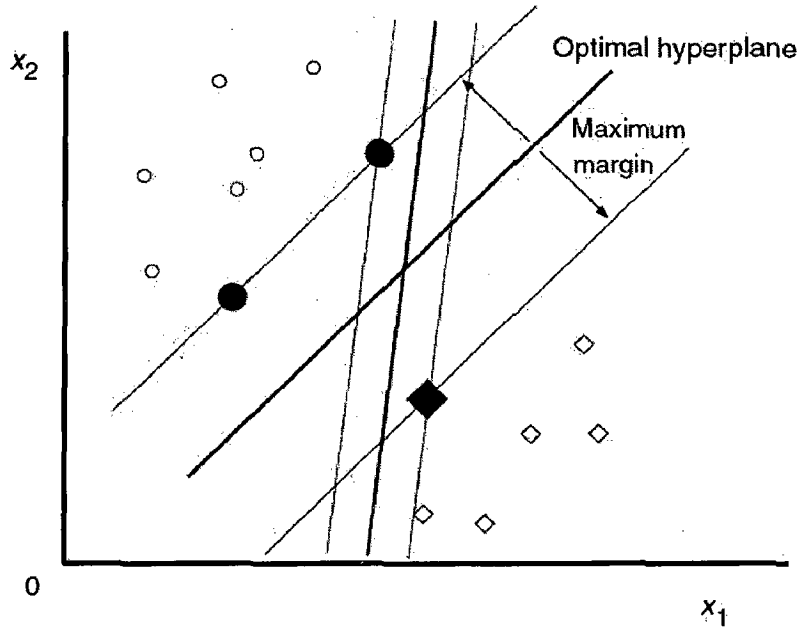


Figure 7.1: Optimal separating hyperplane in a two-dimensional space for linearly seperable case [24].

Now we extend the support vector machine so that it is applicable to an inseparable case.

To allow inseparability, we introduce the nonnegative slack variables ξ_i ($\geq$0) and now the inequality becomes

$$y_{i(}w^T x + b) \geq 1 - \xi_i \quad for \;\; i = 1, ..., M$$

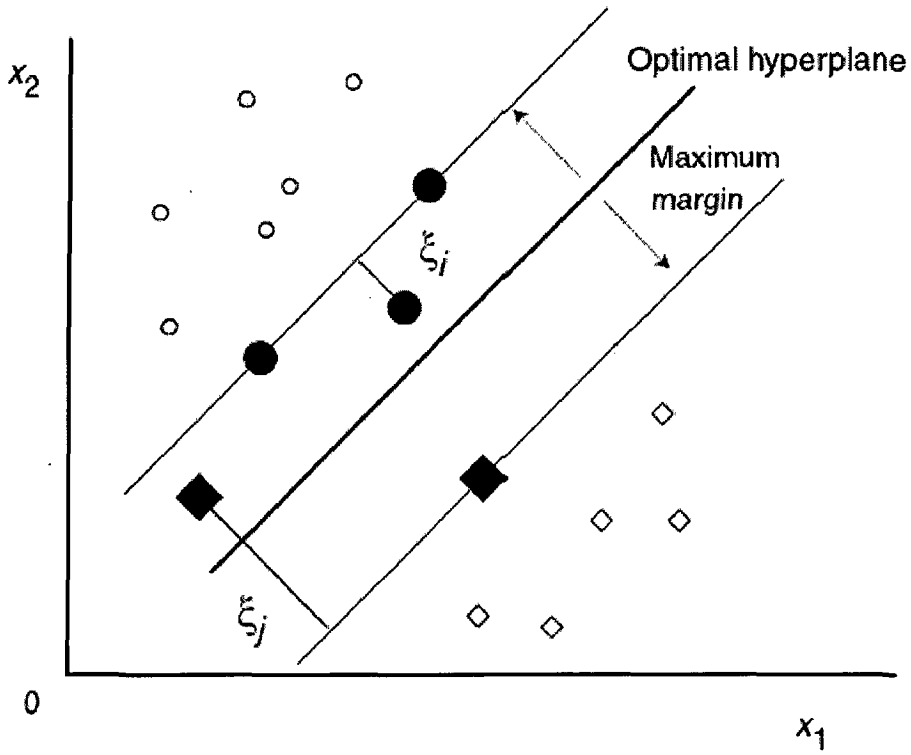


Figure 7.2: Optimal separating hyperplane in a two-dimensional space for inseperable case [24].

By the use of slack variables ξ_i, feasible solutions always exist. For the training data x_i, if $0 < \xi_i < 1$ (ξ_i in Fig. 7.2), the data do not have the maximum margin but are still correctly classified. But if $\xi_i \geq 1$ (ξ_i in Fig. 7.2) the data are misclassified by the optimal hyperplane. To obtain the optimal hyperplane in which the number of training data that do not have the maximum margin is minimum, we need to minimize te following so called *structural risk* function [23][24].

$$Q(w, b, \xi) = \frac{1}{2} \parallel w \parallel^2 + C \sum \xi_i^p \quad for \;\; i = 1....M$$

subject to the constraints

$$y_{i(}w^T x + b) \geq 1 - \xi_i \quad for \;\; i = 1, ..., M$$

where C is a user-specified, positive parameter, ξ_i are slack variables. In particular, when the two classes are separable, minimizing the structural risk amounts to maximizing the separating margin between the two classes.

The cost function constitutes a balance between the empirical risk (i.e., the training

where γ is a positive parameter for controlling the radius. Rewriting above equation

$$K(x,y) = exp(-\gamma \parallel x \parallel^2)exp(-\gamma \parallel y \parallel^2)exp(2\gamma x^T y)$$

Thus K(x,y) is a product of kernels. As $exp(2\gamma x^T y)$ is an infinite summation of polynomials. Thus it is a kernel. In addition $exp(-\gamma \parallel x \parallel^2)$and $exp(-\gamma \parallel y \parallel^2)$ are proved to be kernels and the product of kernels is also a kernel. The resulting decision function is given by

$$D(x) = \sum \alpha_i y_i exp(-\gamma \parallel x - y \parallel^2) + b$$

In the work we used Radial basis function kernel for mapping training set into higher dimensions.

7.5 Result and Discussion

7.5.1 Input data

Standard mammogram images from the Mcgill University database with known region of interest ROI are used. After Enhancement and segmentation, total number of 669 samples are obtained from mammograms containing microcalcifications. out of which 474 samples are from benign class and 175 samples are from malignant class.

These samples are divided into two groups. - training data set and testing data set. Training data set contains 274 samples of benign class and 100 samples of malignant class. Total number of samples used for training SVM network are 374. similarly testing data set contains 200 samples of benign class and 75 samples of malignant class. Total number of samples used for training SVM network are 275.

For each sample total number of 28 features are calculated as discussed in chapter 6.

7.5.2 Result

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 function, its associated parameter, and the regularization parameter C must be decided.

In this work, kernel function used is radial basis function. The varies parameters for SVM like regularization parameter C, sigma of RBF etc. are tuned to give maximum classification accuracy. The parameter C is varied form 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, =2 it gives **86 %** accuracy of classification.

$$y_{i(}w^T x + b) \geq 1 - \xi_i \quad for \;\; i = 1, ..., M$$

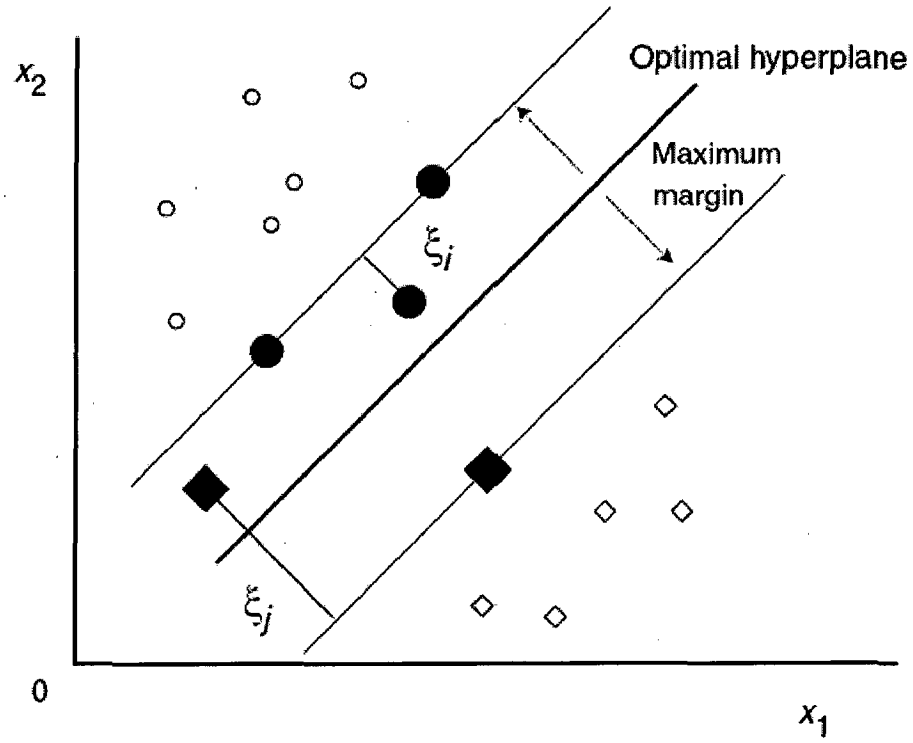


Figure 7.2: Optimal separating hyperplane in a two-dimensional space for inseperable case [24].

By the use of slack variables ξ_i, feasible solutions always exist. For the training data x_i, if $0 < \xi_i < 1$ (ξ_i in Fig. 7.2), the data do not have the maximum margin but are still correctly classified. But if $\xi_i \geq 1$ (ξ_i in Fig. 7.2) the data are misclassified by the optimal hyperplane. To obtain the optimal hyperplane in which the number of training data that do not have the maximum margin is minimum, we need to minimize te following so called *structural risk* function [23][24].

$$Q(w, b, \xi) = \frac{1}{2} \parallel w \parallel^2 + C \sum \xi_i^p \quad for \;\; i = 1....M$$

subject to the constraints

$$y_{i(}w^T x + b) \geq 1 - \xi_i \quad for \;\; i = 1, ..., M$$

where C is a user-specified, positive parameter, ξ_i are slack variables. In particular, when the two classes are separable, minimizing the structural risk amounts to maximizing the separating margin between the two classes.

The cost function constitutes a balance between the empirical risk (i.e., the training

errors reflected by the second term) and model complexity (the first term). The parameter C controls this trade-off. The purpose of using model complexity to constrain the optimization of empirical risk is to avoid overfitting, a situation in which the decision boundary too precisely corresponds to the training data, and thereby fails to perform well on data outside the training set.

7.4 Kernel Functions

In a support vector machine the optimal hyperplane is determined to maximize the generalization ability. But if the training data are not linearly separable, the obtained classifier may not have high generalization ability although the hyperplanes are determined optimally. Thus to enhance linear separability, the original input space is mapped into a high-dimensional dot-product space called the feature space [24].

Now using the nonlinear vector function g(x) = (g_1(x), . . . , g_l(x))T that maps the m-dimensional input vector x into the l-dimensional feature space, the linear decision function in the feature space is given by

$$D(x) = w^T g(x) + b$$

where w is an l-dimensional vector and b is a bias term.

Some of the kernels that are used in support vector machines are as follow

1. ***Linear Kernels :*** If a classification problem is linearly separable in the input space, we need not map the input space into a high-dimensional space. In such a situation we use linear kernels:

$$K(x, y) = x^T y$$

2. ***Polynomial Kernels :*** The polynomial kernel with degree d, where d is a natural number, is given by

$$K(x, y) = (x^T y + 1)^d$$

Here, 1 is added so that cross terms with degrees equal to or less than d are all included.

When d = 1, the kernel is the linear kernel plus 1. Thus, by adjusting b in the decision function, it is equivalent to the linear kernel.

3. ***Radial Basis Function Kernels :*** The radial basis function (RBF) kernel is given by

$$K(x, y) = exp(-\gamma \parallel x - y \parallel^2)$$

where γ is a positive parameter for controlling the radius. Rewriting above equation

$$K(x, y) = exp(-\gamma \parallel x \parallel^2) exp(-\gamma \parallel y \parallel^2) exp(2\gamma x^T y)$$

Thus K(x,y) is a product of kernels. As $exp(2\gamma x^T y)$ is an infinite summation of polynomials. Thus it is a kernel. In addition $exp(-\gamma \parallel x \parallel^2)$and $exp(-\gamma \parallel y \parallel^2)$ are proved to be kernels and the product of kernels is also a kernel. The resulting decision function is given by

$$D(x) = \sum \alpha_i y_i exp(-\gamma \parallel x - y \parallel^2) + b$$

In the work we used Radial basis function kernel for mapping training set into higher dimensions.

7.5 Result and Discussion

7.5.1 Input data

Standard mammogram images from the Mcgill University database with known region of interest ROI are used. After Enhancement and segmentation, total number of 669 samples are obtained from mammograms containing microcalcifications. out of which 474 samples are from benign class and 175 samples are from malignant class.

These samples are divided into two groups. - training data set and testing data set. Training data set contains 274 samples of benign class and 100 samples of malignant class. Total number of samples used for training SVM network are 374. similarly testing data set contains 200 samples of benign class and 75 samples of malignant class. Total number of samples used for training SVM network are 275.

For each sample total number of 28 features are calculated as discussed in chapter 6.

7.5.2 Result

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 function, its associated parameter, and the regularization parameter C must be decided.

In this work, kernel function used is radial basis function. The varies parameters for SVM like regularization parameter C, sigma of RBF etc. are tuned to give maximum classification accuracy. The parameter C is varied form 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, =2 it gives **86 %** accuracy of classification.

Table 7.1: Result of Classification of Microcalcification

	Benign	Malignant	Total
No. of features	28	28	
No. of cases for training	274	100	374
No.of cases for testing	100	75	175
No. of cases classify	85	65	150
Accuracy	**85%**	**86.66%**	**86%**

From the above table, accuracy obtain for bening cases is 85% and for malignant cases is 86.66%. overall accuracy of 86% percent is achieved.

Chapter 8

Conclusion and Scope for Future Work

8.1 Conclusion

This work proposed a computer aided diagnostic system for the detection and classification of microcalcification in mammograms. The whole system is divided into four major parts - Enhancement, Segmentation, Feature Extraction and selection and classification.

This thesis work proposed Top Hat enhancement algorithm which is based on morphological operations. In this technique, the major problem is the choice of structuring element and size of structuring element. we tuned the size and shape of structuring element for the better performance and output. Evaluation indices is also calculated for the analysis of enhancement algorithm.

The next step is segmentation of microcalcification. For this an algorithm based on edge detection and morphological operation is implemented.This algorithm is first tested on synthetic image which contains known number of microcalcification and known shape. Then the algoithm is applied on the Mcgill University database of mammograms.

Shape and texture based features are extracted from the segmented microcalcification. The feature set obtained is used for the classification purpose. In the last part, SVM based classifier is modeled based on Radial basis function kernel with C=10, α=2. 86% accuracy of the classification is obtained.

8.2 Scope for future work

The performance of enhancement algorithm depends on the imaging properties of the database. The imaging property may affect the performance of proposed algorithm. Therefore, the algorithm must be tuned by testing algorithm on different database. In our work 86% of classification accuracy is achieved with RBF kernel. There is a scope for classification using different kernel function and optimizing the SVM Parameters to achieve better accuracy.

Bibliography

1. Interactive Mammography Analysis Web Tutorial- http://sprojects.mmi.mcgill.ca/mammography.

2. Understanding cancer - www.breastcancer.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. 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.

5. 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.

6. Sihem Bouyahia, Jerome Mbainaibeye, Noureddine Ellouze, "Computer aided diagnosis of mammographic images", First International Symposium on Control, Communications and Signal Processing, IEEE, 2004, pp-259-262.

7. 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.

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. Laine AF, Schuler S, Fan J and Huda W, "Mammographic feature enhancement by multiscale analysis", IEEE Trans. Med. Imag., 1994, pp-725-740.

10. Moti Melloul, Leo Joskowicz , "Segmentation of microcalcification in X-ray mammograms using entropy thresholding", CARS 2002.

11. C.H.Li, P.K.S. Tam, "An iterative algorithm for minimum cross entropy thresholding", Pattern recognition Letters 19, 1998, pp-771-776.

12. 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 classication for mammographic microcalcifications" , Elsevier: Computerized Medical Imaging and Graphics 29, 2005, pp- 419-429.

13. S.Singh, Vinod Kumar, H.K. Verma, and Dilbag Singh, "SVM based system for classification of microcalcification in digital mammograms", Proceedings of the 17th IEEE EMBS Annual International Conference, August 2006.

14. J.K.Kim, H.W.Park, "Statistical textural features of microcalcications in digitized mammograms", IEEE Trans. Med. Imaging, 1999, 18(3), pp-231-238.

15. "Digital Image Processing "- Second Edition-Rafel C. Gonzalez, Richard E. Woods.

16. H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, H.N. Du, "Approaches for automated detection and classification of masses in mammograms", Elsevier: Pattern Recognition 39, 2006, pp- 646 - 668.

17. A.Vanzo, G.Ramponi, G.L.Siearanza , "An Image enhancement technique using polynomial filters", International Conference on Image processing , IEEE, 1994, pp.477-481.

18. Edge detection tutorial- http://www.pages.drexel.edu/~weg22/edge.html.

19. S. Halkiotisa, T. Botsisa, M. Rangoussi, "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks ", Elsevier - Signal Processing 87, 2007, pp- 1559-1568.

20. 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.

21. 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 .

22. Matlab help - http://www.mathworks.com/access/help/techdoc/matlab.shtml.

23. Liyang Wei, Yongyi Yang, Robert M. Nishikawa, and Yulei Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications", IEEE Transactions on Med. Imag., Vol. 24, No. 3, March 2005.

24. "Support vector machines for pattern classification" - The book series , Advances in pattern recognition , Springer - 2005.

25. "Support vector machines: Theory and Applications" - The book series, Springer - Studies in fuzziness and soft computing, Volume 177.