بسم الله الرحمن الرحيم

Sudan University for Science and Technology
College of Graduate Studies

Thesis submitted in partial fulfillment of the requirements for the degree of M.Sc. of biomedical engineering

Breast Tissue Recognition On Mammogram Using Shock Filter And Linear Discriminate Analysis

تصنيف انسجة الثدي في صور اشعة الثدي باستخدام مرشح الصدمة والتحليل التصنيفي الخطي

By
Mihad Mahmoud Bakri

Supervisor
Dr. Megdi B. M. Amien

August 2014

Acknowledgment

First and foremost, I thank sincerely Allah for granting me the health, patience, insight and gave me the strength to complete this work. All praise and gratefulness is to Allah.

This thesis would not have been possible without the help and encouragement of several individuals over the course of the past year. First I would like to thank my research supervisor, Dr. Megdi B. M. Amien , for giving me the opportunity to work on this project. It has been a very challenging but rewarding experience. Thank you for your knowledge, encouragement, and patience throughout the process. I also wish to thank Dr. Mohamed Al fadel for He is help me Through his constructive comments and his suggestions throughout the experimental and thesis works have contributed to the success of this research .And thank him for making himself available to me at any time without ignoring . Finally, I would like to thank my family for sacrifice, support and care over the past year has uplifted me and made me a better person. I am extremely grateful to all my friends to encourage them to me.

Table of Contents

Page

List of Figures

Figures **Page**

List Of Tables

Tables **Page**

Abstract

Breast cancer is the most common cancer in women worldwide, in 2012; breast cancer caused “522 000" deaths, in women worldwide.The conventional visual-inspection method for early breast-cancer detection is impractical, non reproducible, and may bestows ambiguous results. Mammography is an effective technology that has demonstrated the ability to detect breast-cancer at early stages. Early detection of breast-cancer greatly improves the chance of survival. Therefore, a fully automated, accurate, and reliable computerized method it highly needed. This study introduces and proposes two methods for breast-tissue discrimination and early detection of abnormal region on mammogram. In the first one Shock filter has been designed and implemented to play key-role as a fully-automatic technique for enhancing image contrast and help for visual analysis the results showed that this Shock filter has validity as competitive results quality-wise; the second is a semi-automatic technique, which is consist of two phases; firstly based-on A novel Logical algorithm; texture-features were extracted from sub-image as Region-Of-Interest, to be as input to the classifier, to classify the selected breast-tissues into; Fate, Glandular, Dense, Begin, , or cancer. The classifier is based-on Linear Discriminant Analysis (LDA).

The proposed method was evaluated on 250 Sub-Images from mini-MIAS database, and the experimental results have shown that the proposed system achieves accuracy of 96.8%.

المستخلص

سرطان الثدي هو أكثر أنواع السرطان شيوعا لدى النساء في جميع أنحاء العالم ، في عام 2012 , تسبب سرطان الثدي "522000" حالة وفاة لدى النساء في جميع أنحاء العالم. طريقة الفحص البصري التقليدي عن الكشف المبكر لسرطان الثدي تعتبر طريقة غير عملية ، غير قابلة للتكرار وربما تمنح نتائج غامضة . التصوير الإشعاعي للثدي هو التقنية الفعالة التي أثبتت القدرة عن الكشف لسرطان الثدي في مراحل مبكرة. الكشف المبكر عن سرطان الثدي يحسن كثيرا من فرصة للبقاء على قيد الحياة. ولذلك ، فإن الطريقة الحوسبة (كاملة التلقائية) تكون دقيقة ويمكن الاعتماد عليها في الكشف عن سرطان الثدي.

تقدم هذه الدراسة وتقترح طريقتين لتمييز الأنسجة الثدي والكشف المبكر عن المنطقة غير الطبيعية على صور أشعة الثدي الرقمية . الأول هو مرشح الصدمة وقد تم تصميمه وتنفيذه كأسلوب تلقائي بالكامل وذلك من اجل زيادة تباين الصورة و التحليل البصري وأظهرت النتائج ان هذا المرشح لديه صلاحية ونتائج تنافسية ذات جودة حكيمة.والثاني هو الأسلوب شبه تلقائي والذي يتكون من مرحلتين أولا استخراج ميزات الملمس من الصورة الفرعية (المنطقة ذات الاهتمام) علي اساس خوارزمية منطقية جديدة وتكون هذه الميزات كدخل للمصنف , لتصنيفها الي انسجة دهنية , غددية , كثيفة, ورم حميد أو ورم خبيث. ويستند المصنف علي التحليل التصنيفي الخطي

تم تقييم الطريقة المقترحة على 250 صورة فرعية من قاعدة بيانات قياسية وأظهرت النتائج التجريبية ان النظام المقترح حقق دقة تساوي 96.8٪ .

.

Chapter One
Introduction

1.1 General View

Breast cancer ranks as one of the leading cancer types in number of new cases diagnosed and is second only to lung cancer as the most prevalent cause of cancer death in women. Early diagnosis of disease can lead to successful treatment. Early diagnosis needs a precise and reliable diagnostic procedure that allows physicians to distinguish between benign breast tumors and malignant ones [1].

Mammography is the process of using low-dose amplitude-X-rays to examine the human breast and is used as a diagnostic as well as a screening tool. Mammograms are considered the most reliable method in early detection of cancer[1].Computer-aided diagnosis (CAD) is system can help doctors in the diagnosis of breast cancer and serve as a useful "second opinion". CAD is a relatively young interdisciplinary technology combining elements of artificial intelligence and digital image processing with radiological image processing [2].

The Wavelet transform is powerful method for analysis of signals and images. Recently it has received interest as a method for contrast enhancement and extraction of feature in digitized mammography [3].

Learning Vector Quantization (LVQ) is an adaptive data classification method based on training data with desired class information. Although a supervised training method, LVQ employs unsupervised data clustering techniques to preprocess the data set and obtain cluster centers[4].

Linear discriminate analysis (LDA) is method used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events[5].

1.2 Problem Statement

Digital mammograms are among the most difficult medical images to be read due to their low contrast and differences in the types of tissues. Brest cancer is an abnormal tissue, these abnormal tissue are very subtle and varied in appearance, making diagnosis difficult and represent challenging even for specialists to detect abnormal tissue. Therefore The CAD system is used for further analysis. CAD system were developed by different methods, but still need more elaboration in order to increase the accuracy of the CAD system in order to detect abnormal areas in breast in real time which consider as second opinion for the radiologist.

1.3 Objectives

The objectives of this study as show below:

General objective:

The purpose of this study was to develop a complete procedure in CAD system for detecting abnormal region on mammogram using image processing technique.

Specific objectives are to:

1. Design and implement a suitable filter to remove noise and background and increase contrast.
2. Extract features of normal and abnormal mammogram .
3. Classify features of mammogram by different techniques.
4. Choose the best technique that as detect abnormal region from normal region in mammogram.

1.4 Significance Of The Study

This study will provide an expert system for mammogram analysis in real time which gives specialists an idea about the exact shape and size of any tumor present in the mammogram in especially early stages ,Thus increase the ratio of healing and reduce the misdiagnosis .It can be implemented in local healthcare facilities in Sudan.

1.5 Thesis Layouts

This study consists of five chapters, chapter one provides an introduction to the study. Where chapter two contains a summary of past and current efforts to solve the breast cancer detection problem and an explanation of the theories used to develop the algorithms that have been implemented to detect breast cancer. Chapter three describes the materials and methods which were used in this study . The results obtained and discussion of these results were shown in chapter four. Finally chapter five contains some concluding remarks and recommendations.

Chapter Two

Theoretical Background

This chapter reviews published research findings and theoretical developments related to CAD systems and various methodologies used.

2.1 Previous Studies

(H.S.Sheshadri et al. 2006) presented an approach for classification of breast tissue by statistic features (mean, standard deviation, smoothness, third moment, uniformity and entropy, The accuracy of this approach was 78%[6].

(S. Mohan 2012) presented an approach for the classification of microcalcification in digital mammogram based on Stochastic Neighbor Embedding (SNE) and K-Nearest Neighbor (KNN) classifier. SNE used to reducing high dimensionality data into relatively low dimensional data. Experimental results showed that the proposed system achieves 100% classification rate for normal and malignant cases and over 80% classification rate for benign and abnormal cases [7].

(Norhène et al.2012) in this paper, an efficient method for enhancement of breast cancer tumor on digital mammogram images has been proposed. It is based on the application of the shock filter. The shock filter has been applied to increase the contrast in mammogram images. Preliminary experimental results on several images show that the proposed method yields significantly superior visual quality and contrast compared to other well known methods, the result of the second-derivative measure of enhancement was improved from 14.8 to 26.8[8].

(Neha 2012) described the three techniques (Bayes learner, Decision Tree and Neural Network) and compared between this techniques in classification of breast cancer. The experiment concluded that Neural Network performance is better than the Decision Tree

classification and Naïve Bayes classification for early detection of breast cancer with better accuracy and precision[9].

(T.A.Sangeetha et al . 2012) proposed an efficient technique to enhance the mammogram image using various transforms. The various transforms are wavelet transform, Curvelet transform, contourlet transform and Nonsubsampled transform .The results were finding drawback of wavelet transform is the method in which problem of filling missing data will occur. In Curvelet method the disadvantage is poor directional specificity of the images. In contourlet transform the image enhancement cannot capture the geometric information of images and tend to amplify noises when they are applied to noisy images since they cannot distinguish noises from weak edges. This entire drawbacks is overcome by the Nonsubsampled Contourlet transform [10].

(Raja et al. 2012) demonstrated a novel approach for classifying mammograms by computer aided design using image processing and data mining techniques . This approach consists of four stages: Begin from preprocessing the breast image is standardized. Then suspicious regions of cancer are acquired from mammogram by K-means clustering technique. Features are extracted from these region and are given as input to the pretrained decision tree based classifier, which in turn classifies the mammogram into normal, benign and malignant. The results were as following out of 30 normal images 29 were correctly classified as normal and 1 was misclassified as benign. Among the 30 benign images fed to the system 26 were rightly classified as benign and 4 was wrongly classified to malignant. Among 30 malignant images 27 were correctly classified and 3 was misclassified to benign [1].

(Nizar et al.2013) proposed algorithm of detection of microcalcifications in mammogram by Discrete Wavelet Transform . This algorithm contain two stages: firstly were chosen the optimal level

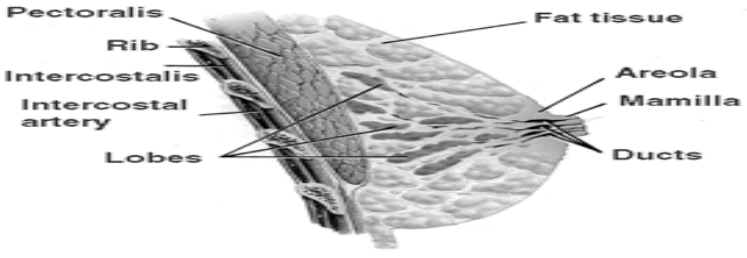
of decomposition and were chosen the wavelet which appears most suitable with mammogram. Then performed a multiresolution decomposition filter banks of the mammograms .The results were The binarisation of the reconstituted image as well as the local thresholding gave a clear reduction in the number of false positives [11].

(R.R.Janghel et al.2013) introduced a novel approach based on Hierarchical Learning Vector Quantization (HLVQ) to developed algorithm CAD system. The approach is divided into two phases. The first phase of the algorithm consists on multiresolution analysis based on 1-D discrete wavelet transform over profiles of microcalcifications extracted from mammographic images . The second phase were applid 2-D discrete wavelet transform in analysis and synthesis on screening mammograms in order to detect the microcalcifications. The HLVQ gave accuracy of 98.14% for diagnosis of breast cancer [4].

(R.Ramani et al.2013) presented an approach for detection of breast cancer based on Clustering Techniques. Clustering techniques may help to enhance the efficiency of the image recovery process. All the clustering techniques may obtain satisfaction results but not able to produce 100 % of accuracy[12].

2.2 Breast Anatomy

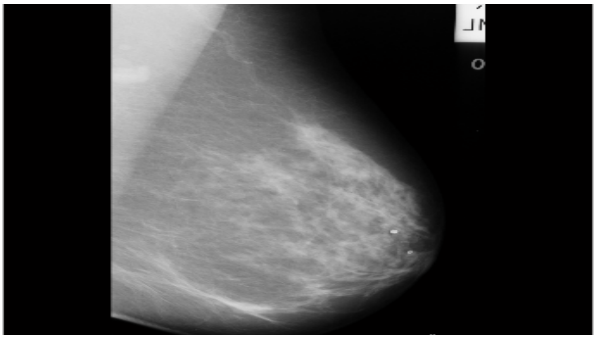
The breast is a complex organ consisting of many different types of tissue see figure(2.1)[3].



Fig(2.1): Structure Of Breast

It is primarily made up of fat and connective however tissue the breast also contains milk ducts, blood vessels, lymph nodes and structures lobes and lobules [13].There is a great variation in the structure of the breasts of different women , but usually the two breasts of one woman are very much alike [3].

When imaging the breast by mammography, the different forms of the tissue appear as different shades of grey depending on the level of absorbed radiation . Skin and fat tissue absorb very little radiation therefore usually not show, Glandular tissue is normally show in medium or light grey shades. The pectoral is muscle is located behind the glandular tissue, covering the ribs. Pectoralis is often visible as a whiter area in a normal mammogram in this area there are also many lymph glands and these are the reasons why breast cancer may easily spread to other parts of the body shown in figure(2.2)[3].



Fig(2.2): Mammography Image

2.3 Breast Cancer

Cancer starts when normal cells in the breast change and grow uncontrollably, Healthy cells reproduce themselves continuously throughout life, growing new tissue and replacing old or damaged tissue. This is a normal, controlled and orderly process. However , sometimes this process disturbed and cells begin to reproduce themselves in an abnormal way, building a tumor[3].

The word tumor means abnormal growth and may refer to both benign and malignant growth. Benign tumors tend to remain similar to the tissue of origin they remain differentiated. Generally they do not invade surrounding tissue or produce metastasis .The growth of a benign tumor is usually slow compared to the growth of a malignant tumor[3].

Malignant cells appear in many different forms. Some remain similar to the surrounding tissue and are referred to as well-differentiated, Some cells bearing very little similarity to surrounding tissue are referred to as undifferentiated or anaplastic . Undifferentiated or anaplastic malignancies are usually more aggressive in their growth and behavior than well-differentiated malignancies are [3].

The majority of breast cancers begin in either the lobules or the ducts. Breast cancer is classified as either invasive or non-invasive. Invasive, or infiltrating, breast cancer has the ability to spread to other parts of the body. Noninvasive, also referred to as in-situ, breast cancer does not spread into other parts of the body but may develop and become invasive and should therefore be removed[3].

2.4 Mammography

The actual absolute cure to breast cancer is the early detection before its symptoms. Mammography remains the most valuable and single successful technique for the early detection of breast cancer in breast imaging; a mammogram is a low dose an x-ray (radiography) picture or mammography exam of the breast. X-ray mammography is the only proven method capable of detecting nonpalpable cancers. Studies have shown that the mammography exam is a lot easier to women with fatty breasts than those with dense breasts (Breast Cancer). There are two types of mammograms: (a) Screening mammogram used to early detect breast cancer before its symptoms and (b) Diagnostic mammogram used

to evaluate patients with abnormal clinical findings and under treatments of breast cancer[14].

The quality of a mammogram image is based on various factors: the nature and accessories of the mammography unit, the use photographic film for image acquisition, storage and display; although films provide very high resolutions and good visibility of high contrast structures, the signal contrast is weak. This limits the exposure dynamic range of mammographic screen film systems to a factor of 25 – 50, which means that the image contrast in the fully glandular or fully adipose parts of the breast can be much lower than in the other areas. On the other hand, digital detectors have the significant advantage of a linear response over a wide of exposure conditions, giving constant contrast and a large dynamic range (Monnin), Signal to noise ratio limited by radiation , high Security , easy of copying and easy connecting with CAD [14].

2.4.1 Signs of Cancer in a Mammogram

There are some signs of cancer to look for in a mammogram. The primary signs are local distortion of soft tissue, distortion of glandular tissue and the existents of malignant microcalcifications. These signs may appear alone or together. Skin thickening, skin distortion, retraction of the nipple and oedema of the breast are considered as secondary signs. In this section some of these signs are described [3].

The attenuation of a tumor may vary depending on the type of tumor. Tumors such as lipoma consisting of fat, have low attenuation while scirrhus cancer with productive fibrosis have high attenuation. The shape also varies. Benign tumors such as fibro adenoma, lipomas and cysts are usually rounded and have distinct borders, while malignant tumors tend to be more irregularly in shape, often speculated, and have diffuse borders[3].

Microcalcifications are microscopic grains of calcium produced by the cells as the result of some benign or malignant process. Most calcifications are the result of some benign process. They may for instance be the rest products of broken down cells, a cyst or milk. The benign and malignant calcifications differ in shape, density and distribution. The calcifications produced by benign processes are generally scattered in distribution and have uniform shape and density. Benign microcalcifications appear in the ducts, the lobules or outside the glandular tissue. The calcifications produced by cancer cells are more irregular in shape, size and distribution. They are generally smaller than the benign calcifications, granular in shape and appear in clusters. 20% percent of all cluster calcifications are the result of malignant processes. Malignant microcalcifcations appear in either the ducts or the lobules. Calcifications outside glandular tissue are not a sign of breast cancer [3].

The calcifications have a much higher attenuation compared to the surrounding tissue, and absorb more radiation. Therefore, the calcifications are visible as bright spots in a mammogram. When suspicious microcalcifications are detected, a needlebiopsi can be performed and the calcifications can then be diagnosed pathologically[3].

2.5 Digital Image Processing

Digital image processing (DIP) is defined as a processing of digital image by a digital computer. In DIP systems, usually deals with array of numbers obtained from sampling spatial points of a physical image. After processing this array is produced other array, and these array are then used to reconstruct a continuous image for viewing, this processing is doing in several steps[15], [16]:

1. Data acquisition using sensor technology.
2. Pre processing is enhancement quality of image.

3. Segmentation is the process that subdivides a digital image into a number of uniformly homogeneous regions [17].
4. features Extracted which give measures of difference properties of image segments. Each segmented region in a scene may be described by a set of such features.
5. Classification to classified a set of meaningful classes.

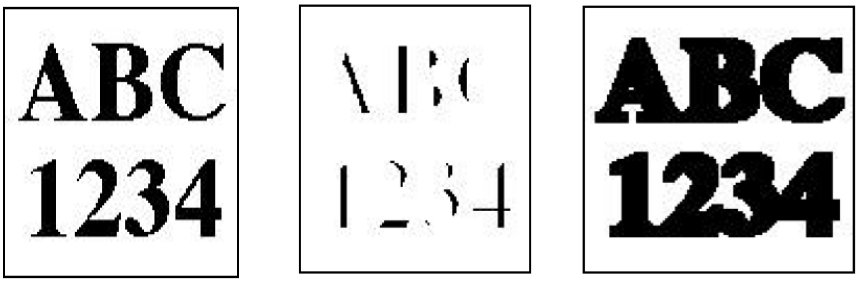
2.6 Computer-Aided Detection Systems (CADS)

In the computer-aided diagnosis systems, the image processing techniques are combined with artificial intelligence algorithms along with radiological image processing to form the Computer-Aided Diagnosis methods in which for instance tumors are detected. Radiologists use CAD to read and interpret radiographic images . the CAD system examines digitized film mammograms for evidence of suspicious masses or calcifications. Radiologists can display the findings and match with the results by scanning images[14] .

2.7 Mathematical Morphology

Mathematical Morphology is a powerful tool in the field of image processing and computer vision and is used for extracting, modifying and combining image components that are useful in the representation and description of region shapes. In morphology, the objects in an image are considered as set of points and operations are defined between two sets: the object and the structuring element (SE) .The shape and the size of SE is defined according to the purpose of the associated application[21]. Basic morphological operations are erosion and dilation. Dilation is a transformation that produces an image that is the same shape as the original, but is a different size. Dilation stretches or shrinks the original. And erosion is used to reduce objects in the image see figure (2.3)[21].

Other operations of morphology like opening and closing are sequential combination of erosion (dilation) and dilation (erosion)[21].



a. **b.** **c.**

Fig(2.3): Basic Morphological Operations (a. Original binary image b. Image after erosion c. Image after dilation)

2.8 Shock Filter (SF)

Shock filters belong to the class of morphological image enhancement methods. Most of the current shock filters are based on modifications of Osher and Rudin's formulation in terms of partial differential equations (PDEs) [25]. Shock filters offer a number of advantages: They create strong discontinuities at image edges, and within a region the filtered signal becomes flat. Thus, shock filters create segmentations. Since they do not increase the total variation of a signal, they also possess inherent stability properties. Moreover, they satisfy a maximum-minimum principle stating that the range of the filtered image remains within the range of the original image. Thus, in contrast to many Fourier- or wavelet-based strategies or linear methods in the spatial domain , over- and undershoots such as Gibbs phenomena cannot appear. This makes shock filters attractive for a number of applications where edge sharpening and a piecewise constant segmentation is desired [25]. The idea of shock filtering is explained as follows[25]:

Let be f a continuous image where $f: R^2 \rightarrow R$ the class of the filtered images $\{u(x, y, t) | t \geq 0\}$ may be produced as a result of developing the image f below the process[25]:

$$u_t = -sign\,(\Delta u)|\nabla u| \qquad \textbf{(2.1)}$$

$$u(x, y, 0) = f(x, y) \qquad (2.2)$$

Where subscripts denote partial derivatives, and $\nabla u = (u_x, u_y)^T$ is the (spatial)gradient of u. The initial condition (2.2) ensures that the process starts at time $t = 0$ with the original image $f(x, y)$. The image evolution proceeds in the following way: Assume that some pixel is in the influence zone of a maximum where its Laplacian $\Delta \boldsymbol{u} = \boldsymbol{u_{xx}} + \boldsymbol{u_{yy}}$ is negative. Then (3.1) becomes[25]:

$$u_t = |\nabla u| \qquad (2.3)$$

Evolution under this PDE is known to produce at time t a dilation process with a disk-shaped structuring element of radius t. At the influence zone of a minimum with $\Delta u < 0$, equation (2.2) can be reduced to an erosion equation with a disk-shaped structuring element [25]:

$$u_t = -|\nabla u| \qquad (2.4)$$

These considerations show that for increasing time, (1) increases the radius of the structuring element until it reaches a zero-crossing of Δu , where the influence zones of a maximum and a minimum meet. Thus, the zero-crossings of the Laplacian serve as an edge detector where a shock is produced that separates adjacent segments. The dilation or erosion process ensures that within one segment, the image becomes piecewise constant [25].

A number of modifications have been proposed in order to improve the performance of shock filters. For instance, that the second directional derivative $u_{\mu\mu}$ with $\mu||\nabla u$ can be a better edge detector than Δu . In order to make the filters more robust against small scale details, Alvarez and Mazorra [1] replaced the edge detector $u_{\mu\mu}$ by $v_{\mu\mu}$ with $v = K_\sigma * u$ In this notation, K_σ is a Gaussian with standard deviation σ and $*$ denotes convolution. Taking into account these modifications the shock filter becomes [25]:

$$u_t = -\text{sign}(v_{\mu\mu})|\nabla u| \qquad (2.5)$$

2.9 Basic Wavelet Theory

The wavelet transform reveals frequency information, called scale information, as well as information about the times at which different frequencies occur, called translation information[18].

The popular advantages Wavelet transform[11]:

1. Wavelet transform gives the both frequency and time information of the analyzed signal
2. Admissibility: wavelet transform can be used for first analyze and then reconstruct a signal without loss of information .
3. Regularity: Wavelet regularity is a key property to improve the detection of singularities .
4. Translation covariance: Shifting the original image produces a wavelet coefficient shifting without changing its structure.

2.9.1 Continuous And Discrete Wavelet Transforms

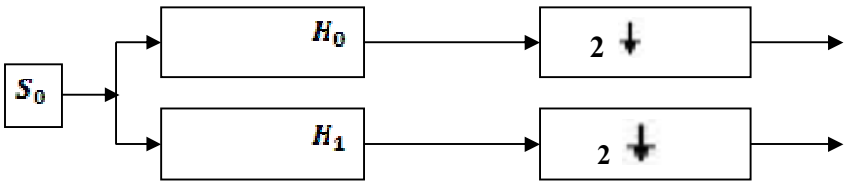
In a continuous wavelet transform the mother wavelet is continuously scaled and shifted along the data, potentially generating an infinite number of representations. This makes the continuous wavelet transform highly redundant and impractical to use. In practice, a discrete wavelet transform is used, allowing a predefined number of derivative datasets to be generated[18]. therefore is possible choose scales and positions based on powers of two called dyadic scales and positions then analysis will be much more efficient. were obtain such an analysis from the discrete wavelet transform (DWT). An efficient way to implement this scheme using filters was developed[19].

2.9.2 Filter Bank

Filter bank consisting of a high pass filter and a low pass filter which separates the signal into different frequency bands. In first section a short description of signal decomposition through filters is provided[3].

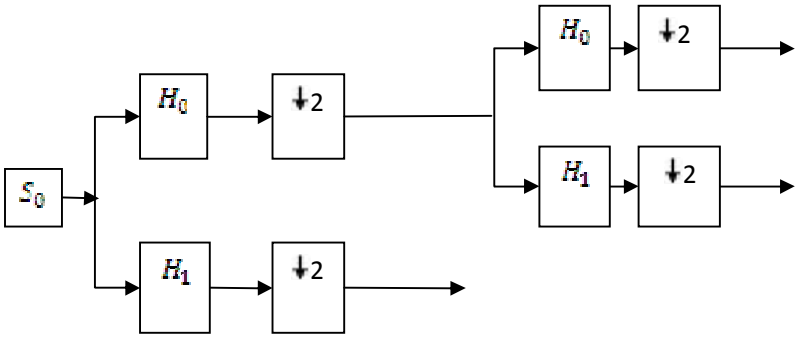
A filter is a time invariant linear operator that may be applied to a signal to remove or enhance certain frequency bands of the signal. By applying a low-pass filter to a signal S of length 2^n the high frequency bands of the signal are removed and a smooth version of the original signal is obtained. A high-pass filter removes the low frequency components of the subjected signal and the result is a signal containing the details (differences) of the original signal. By combining these two filters into a filter bank the original signal is divided into an average signal and a difference signal[3].

The problem with the used filters is that they each produce a signal of the same length as the original signal 2^n and the amount of data is thus doubled. To avoid this data growth each of the two signals are subjected to down-sampling , i.e. every other element is simply removed. The result of the down-sampling is the two signals $s_{n-1}$and d_{n-1} each of the length 2^{n-1} The decomposition process is see figure(2.4)[3].



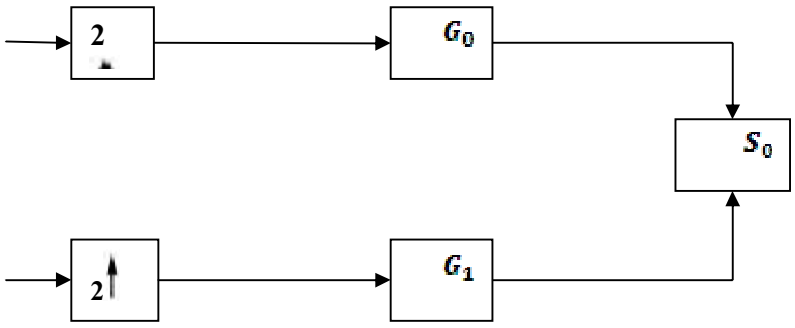
Fig(2.4): Decomposition Of A signal Using A filter Bank

A multi level decomposition of the original signal is obtained by repeating the decomposition process. The low-pass filtered output signal is used as input to the filter bank. A scheme for a multi level decomposition see figure(2.5)[3].



Fig(2.5): A multi Level Decomposition using a filter bank

The reconstruction of a decomposed signal may also be performed using a filter bank This reconstructive filter bank is a reversed version of the filter bank used for the decomposition. The mathematical manipulation that effects synthesis is called the inverse discrete wavelet transform (IDWT). The filters used in this filter bank are constructed to reverse the effect of the two filters used in the decomposition. To obtain a signal of the original length 2^n the sub-band signals d_{n-1} and d_{n-1} are subjected to upsampling prior to the filter operations. In the up-sampling process, a zero is inserted at every other position of the two sub-band signals. The two output signals obtained from the filters are then combined into one signal by an add operation. A filter bank for signal reconstruction see figure (2.6) Multi level reconstruction of a signal is done by repeating the reconstructive process in a manner similar to the multi level decomposition but reversed [3].



Fig(2.6): Reconstruction Using A filter Bank

Designing the filters in a wavelet filter bank can be quite challenging because the filters must meet a number of criteria for instance The criterion for aliasing cancellation is[20].

$$G_0\ (z) * H_0\ (z) + G_1\ (z) * H_1\ (z) = 0 \qquad (2.6)$$

Where $H_0\ (z)$ is the transfer function of the analysis lowpass filter, $H_1\ (z)$is the transfer function of the analysis highpass filter, $G_0\ (z)$ is the transfer function of the synthesis lowpass filter, and $G_1\ (z)$ is the transfer function of the synthesis high pass filter[21].Can be calculate the coefficients of $H_1\ (z)$

$$H_1\ (z) = [H_0\ (z)\ ,- H_0\ (z-1)\ , H_0\ (z-2), \dots] \qquad \textbf{(2.7)}$$

Can be calculate the coefficients of $G_0\ (z)$

$$G_0\ (z) = [H_0\ (z)\ , H_0\ (z-1)\ , H_0\ (z-2), \dots] \qquad \textbf{(2.8)}$$

Can be calculate the coefficients of $G_1\ (z)$

$$G_1\ (z) = [H_1\ (z)\ , H_1\ (z-1)\ , H_1\ (z-2), \dots] \qquad \textbf{(2.9)}$$

2.9.3 Application Of Wavelet Transform

The wavelet transform has been found to be very useful in many scientific areas, including mathematics ,statistics ,physics and econometrics .In signal and image processing areas, the methods of multi-resolution analysis have received a lot of interest[3].

2.10Texture Features

Texture is a very general notion that can be attributed to almost everything in nature. For a human, the texture relates mostly to a specific, spatially repetitive (micro)structure of surfaces formed by repeating a particular element or several elements in different relative spatial positions. Generally, the repetition involves local variations of scale, orientation, or other geometric and optical features of the elements[26].

Image textures are defined as images of natural textured surfaces and artificially created visual patterns, which approach, within certain limits, these natural objects. Image sensors yield additional geometric and optical transformations of the perceived surfaces, and these transformations should not affect a particular class of textures the surface belongs. It is almost impossible to describe textures in words, The term Texture generally refers to repetition of primitive texture elements called texels, the texels describe the spatial relations between them. Texture may be coarse, fine, smooth, granulated, regular or irregular[26].

Many statistical texture features are based on co-occurrence matrices representing second-order statistics of grey levels in pairs of pixels in an image. The matrices are sufficient statistics of a Markov/Gibbs random field with multiple pairwise pixel interactions[26].

A co-occurrence matrix is specified by relative occurrence frequencies of two pixels, separated by distance d, along the direction of angle , one with gray level i and the other with gray level j. A co-occurrence matrix is therefore a function of the distance d, the angle and gray levels. The co-occurrence matrix can be calculated for the whole image, but by calculating it in a small window which scans the image, the co-occurrence matrix can be associated with each pixel [27].

There are a set of textural features which can be extracted from each of the gray level spatial dependence matrix. The following equations represent definition some of these features [28].

Table (2.1): Textural Features

Texture feature	Equation
Energy (EG)	$EG = \sum_{i=0}^{n=1} \sum_{j=0}^{n=1} p^2(i,j)$
Entropy (EN)	$EN = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} p(i,j) \log_2 p(i,j)$
Inverse Difference Moment (IDM)	$IDM = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \frac{1}{1+(i-j)^2} p(i,j)$
Inertia (IN)	$IN = -\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i-j)^2 \, p(i,j)$

Where **n** : the number of grey level in the image. **P**: joint Probability. **i , j**:row & column.

2.11 Learning Vector Quantization (LVQ)

LVQ is an adaptive data classification method based on training data with desired class information. Although a supervised training method, LVQ employs unsupervised data clustering techniques (e.g., competitive learning) to preprocess the data set and obtain cluster centers[4].

The general concept of the LVQs comes from the Hebbian Learning. The Hebbian Learning is an understanding of the human brain and the learning associated with it. The Hebbian Learning is derived from the Hebb's postulate which states that "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth processes or metabolic changes takes place in one or both

cells such that A's efficiency, as one of the cells firing B, is increased. [4].

The approach is also termed as the winner-takes-all approach. This is because we first study the activities of all the neurons. Then we decide the winning neuron. The activity of this winning neuron is adjusted according to the answer. This may be visualized to be a very apt learning strategy for any classificatory problem where the winning neuron is the point in the input space and the final output is nothing but the class to which it belongs to. The continuous adjustment of the neuron activity and their effects on the neighboring neuronal activity in multiple epochs is done. This adjusts the various weights or parameters of the network in order to better adapt the network to the problem. This makes these networks perform well for the classificatory problems[4].

2.12 Linear Discriminate Analysis(LDA)

Linear discriminate analysis is statistical technique which was employed to find variables discriminate between several classes. The discriminant function is formulated by a linear combination of the feature variables as shown in equation (3.13) [22].

$$y = b + \sum_{i=1}^{n} a_i \; x_i \qquad (\mathbf{2.10})$$

Where n is the number of feature variables, $x_i$are the values of the feature variables,a_i are coefficients (or weights) estimated from the input data during training, and b is constant also estimated from the input data during training, so that the separation between the distributions[22].

2.13 Matlab And The Image Processing Toolbox

MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-

use environment where problems and solutions are expressed in familiar mathematical notation. Typical uses include the following[2]:

1. Math and computation
2. Algorithm development
3. Data acquisition
4. Modeling, simulation, and prototyping
5. Data analysis, exploration, and visualization
6. Scientific and engineering graphics
7. Application development, including graphical user interface building.

MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows formulating solutions to many technical computing problems, especially those involving matrix representations, in a fraction of the time it would take to write a program in a scalar non-interactive language such as C or Fortran[2].

2.14 Statistical Package For The Social Science (SPSS)

SPSS is a good first statistical package for people wanting to perform quantitative research in social science because it is easy to use and because it can be a good starting point to learn more advanced statistical packages. Many of the widely used social science data sets come with an easy method to translate them into SPSS; this significantly reduces the preliminary work needed to explore new data [23].

SPSS is used for

1. conducting statistical analyses
2. Manipulating data
3. Generating tables and graphs that summarize data

Advantages of SPSS [24]:

1. Easy-to-use pull-down menus, like Microsoft XP/Office

2. Users do not need to know complex statistical equations to use SPSS
3. SPSS is like a calculator: users enter the numbers and select the tasks, and the software does the rest
4. Results are clearly presented
5. Tables and graphs can be copied directly from SPSS into Word files .

2.15 Summary

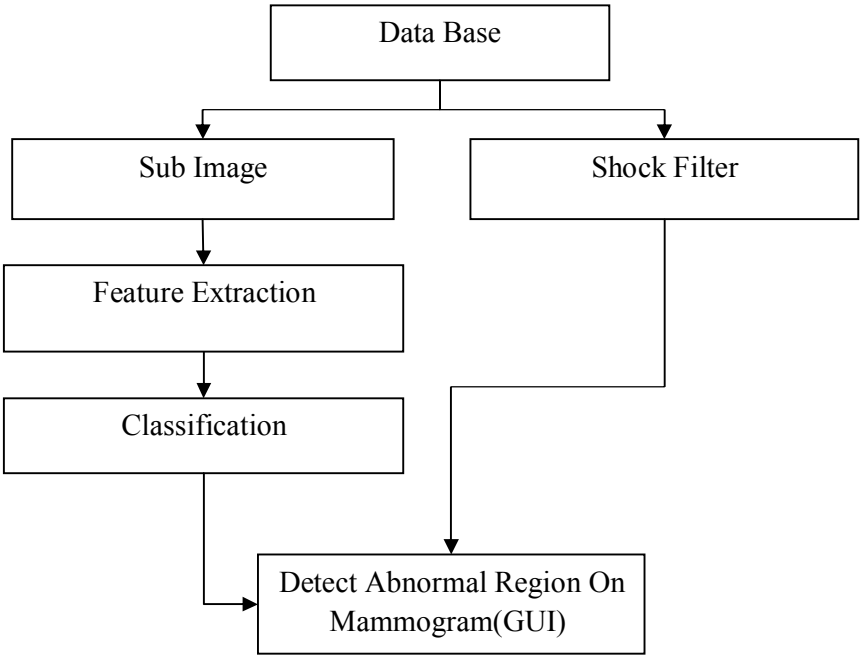
This chapter consist two sides , the first side the previous studies, are include last developments in CAD system, this developments either was in preprocessing stage or extracted feature stage or classification stage. The literature search demonstrates the need for further developments in CAD system because All the previous studies may obtain satisfaction results but not able to produce 100 % of accuracy. And second side explains the basic mathematical formulas for Shock filter , Inverse Discrete wavelet transform ,texture analysis , learning vector quantization and linear discriminate analysis .The next chapter gives a detailed explanation as to how these tools are implemented.

Chapter Three

Materials and Methods

3.1 Introduction

This chapter provides the actual methods used in this study to Breast Tissue Recognition On Mammogram. This is based on two methods to either use shock filter method or choose a sub image then extracted feature of this sub image and classification this feature ,shown in figure (3.1).



Fig(3.1): Block Diagram For Detection Abnormal Region In Digital Mammograms

3.2 Data Base

The nature of the data used for cancer research is of utmost importance in terms of being able to correctly evaluate the methods and results. Data for this study was obtained from Mammographic Image Analysis Society (Mias)[31].The database contains left and right breast images of 161 patients. Its quantity consists of 322 images, which belongs to three types such as Normal, benign and malignant. The database has been reduced to 200-micron pixel edge, so that all images

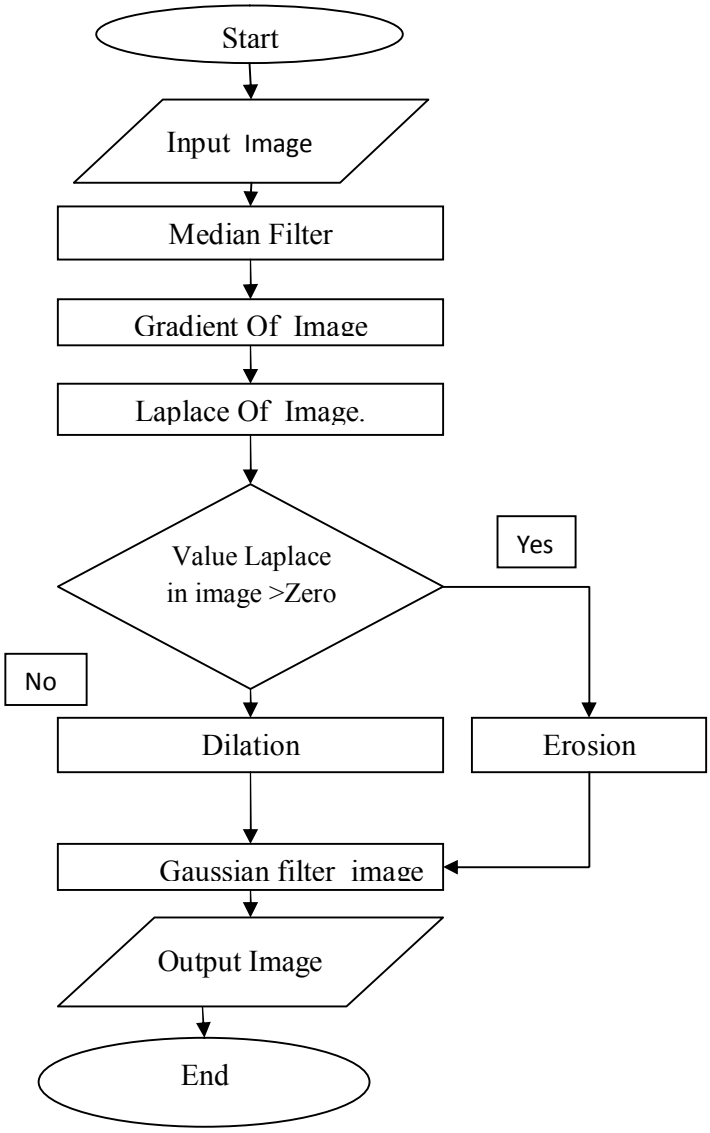
are 1024 x 1024. There are 208 normal, 63 benign and 51 malignant (abnormal) images[29]. In this study was chosen three sizes for sub image are (199×199 ,79×79 and 19×19) pixels. And was selected 150 sub-images as normal tissues (50 is fatty tissue , 50 is glandular tissue and 50 is dense tissue), and 100 sub-images as abnormal tissue (50 is benign tissue and 50 is malignant tissue).

3.3 Shock Filter

The mammogram images contain low contrast and most this images contain background and Existing artifacts like written labels. In this study was used shock filter in order to remove background ,artifacts, pectoral muscles ,edges detection and increase contrast in abnormal region in mammogram. This was done by flow chart, shown in figure (3.2) and programming this flow chart in mat lab (see Appendix).

In first stage is applied median filter on mammogram to reduce "salt and pepper" noise then calculated the gradient of image in second stage in order to calculate the Laplace in the third stage see equation (2.1), then is tested the Laplace if positive value then the pixel is considered to be in the influence zone of a minimum(erosion) and If the Laplace is negative value then the pixel is considered to be in the influence zone of a maximum (dilation),Last stage is applied Gaussian filter on image to Remove Gaussian noise.

The performance of the shock filters is evaluated by calculate of contrast on image , This was done by programming mat lab code (see Appendix).



Fig(3.2): Flow Chart For Shock Filter

3.4 Feature Extraction

Transforming the input data into the set of features is called features extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. The traditional goal of feature extraction is to characterize an object to be recognized by measurements whose values very similar for objects in same category and very different

for object in different category[1]. Features extraction phase from consists of two components: A novel Logical algorithm and texture feature.

3.4.1 A novel Logical algorithm:

Convolution is an important concept in this algorithm, Convolution can be used to define a general input–output relationship in the time domain analogous to the Transfer Function in the frequency domain [31].The basic Logical algorithm is the following (program in matlab , see Appendix) :

1. Input image*(X)* (size of *X*=*Ls* × *Ls* and *X* in time domain).
2. Input coefficients of low pas filter $H_0(z)$ (from tool box of mat lab ,that match to coif let family from order one(coif1) by experience this is the best).
3. Calculate coefficients of high pass filter from $H_0(z)$ (eq 2.7).
4. Input the vector (*S*) from image (*X*) (length of *S* is *Ls* in one dimension (horizontal) by experience found that it the best).
5. Divide this vector to two segment $x_0\, x_1$ (x_0 from first element in *S* to *Ls/2* element in *S* (length of x_0 is *Ls/2*)this segment conceder low frequencies and (x_1 from ((*Ls/2*)+1)element in *S* to *Ls* element in *S* (length of x_1 is *Ls/2*)this segment conceder high frequencies.
6. Up sampling for $x_0\, x_1$ (become lengths of $x_0\, x_1$ are *LS*).
7. $y_0 = x_0\, Convolution\ with\ H_0(z)$(length $of\ y_0\ (Ly_0\ )$=length $(x_0\ )$ + $length\ (H_0\ ) - 1)$.
8. $y_{1=}x_1\, Convolution\ \ with\ H_1(z)$*(length* $of\ y_1\ (Ly_1\ )$=*length(*$x_1\ )$ + $length\ (H_1\ ) - 1)$.
9. $Y = y_0 + y_1$ (3.1)

(according to addition in matrix)

10. Remove extra points in Y from end(according to length of vector *S* because length of Y must be Ls).
11. Output *Y* (vector in frequency domain).

3.4.2Texture Feature:

It which play very important role in detecting abnormalities of mammograms because of the nature of mammograms. In this study were used entropy ,energy, inverse different moment and inertia were applied on the coefficients were extracted using A novel Logical algorithm. These Features were calculated from the Spatial Gray Level Dependency (SGLD) matrix, SGLD matrix of the sub image which displays the gray level spatial-dependency along horizontal angular ($\theta = 0^o$) and a distance d of 1 pixel on an image, to calculate textural measures. A co-occurrence matrix is specified by relative occurrence frequencies $p(i, j, d, \theta)$ of two pixels, separated by distance (d), along the direction of angle (θ), one with gray level (i) and the other with gray level (j). A co-occurrence matrix is therefore a function of the distance, the angle and gray levels. in this study it calculated with a window 20x20 pixels which scan the image. Inside each window the joint probability was calculated.. This was done by programming mat lab (see Appendix). The feature extraction stage is evaluated by SPSS program.

3.5 Classification

The overall objective of image classification is to automatically categorize all pixels in an image into land cover classes or themes. In this study was used two classifier to detect abnormal region in digital mammogram , Were used each of them individually, the first classifier is Learning Vector Quantization . This classifier was applied on the features which were extracted in order to classified each target classes. Firstly The data is divided into training set 60 regions of interest and testing set

20 ROIs. Then the network is trained with training set of all samples individually. Then is tested with test set the accuracy is calculated, This was done by programming mat lab (see Appendix).

The second classifier is The linear discriminate analysis, the aim of the LDA is to find the principal axis that provides the maximum separation between the distributions of the discriminate scores for the five classes (fatty, glandular , dense , benign and malignant).The linear discriminant analysis can be performed in a two-stage process using the statistical package SPSS Version 20.0. First, a stepwise procedure is performed to identify from all available input features the suitable feature variables for the formulation of the discriminate function. Second, the selected features of the input cases are used to determine the coefficients of each feature variable in the discriminate function to achieve maximum separation, this was done by programming matlab (see Appendix).

The results are validated using Sensitivity, Specificity, Accuracy and positive and negative predictive values see table (3.1) .

Table (3.1): Validation Measure

	Validation measures	**Formulas**
1	Sensitivity(SE)	$SE = \frac{TP}{TP + FN}$
2	Specificity(SP)	$SP = \frac{TN}{TN + FP}$
3	Accuracy(AC)	$AC = \frac{TP + TN}{TN + TP + FN + FP}$
4	Positive predictive values(PPV)	$PPV = \frac{TP}{TP + FP}$
5	Negative predictive values(NPV)	$NPV = \frac{TN}{TN + FN}$
6	Matthews Correlation Coefficient(MCC)	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$

TP: True Positive case TN: True Negative FN: False Negative FP: False Positive

3.6 Graphical User Interface For Real Time Analysis:

A Graphical User Interface is the most common type of user interface seen today. One of the main advantages of creating a standalone GUI is to be able to make it available to a user with no Matlab knowledge [30].Compilation of all previous programs in one **GUI**, this was done by programming mat lab (see Appendix).

3.7 Summary:

This chapter involves number of different steps which were used in this study for detection of abnormal region in mammogram , Actual results generated by this steps will be presented in the following chapter.

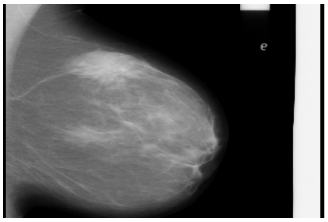
Chapter Four

Results and Discussion

In this chapter were displayed results of shock filter, then were displayed results of feature extraction by SPSS program , then were displayed results of classification of abnormal region from normal region based on learning vector quantization or linear discernment analysis. and a discussion on what has been done and the observations that were made throughout. has been verified with the ground truth given in the database (mini-MIAS database & DDSM).

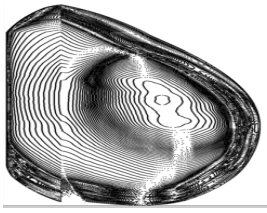
4.1 Shock Filter

Shock filter was applied on mammogram, shown in figure (4.1). was found a lot of Gaussian noise on figure (4.1.b) for this was used Gaussian filter to remove it, shown in figure (4.1.c).



a.

b.

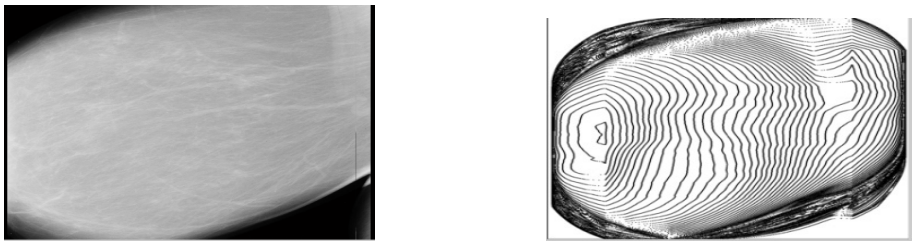


c.

Fig(4.1) : Shock Filter On Abnormal Mammogram (**a.** Original Mammogram **b.** Mammogram After Shock Filter **c.** Mammogram After Gaussian Filter)

Shock filter was applied on normal mammogram, shown in figure (4.2),the value of contrast in normal mammogram is 0.23 and in Shock Filter of mammogram is 0.56.

When Shock filter was applied on abnormal mammogram ,shown in figure(4.1.c), the value of contrast in abnormal mammogram is 0.47 and in Shock Filter of mammogram is 2.25.



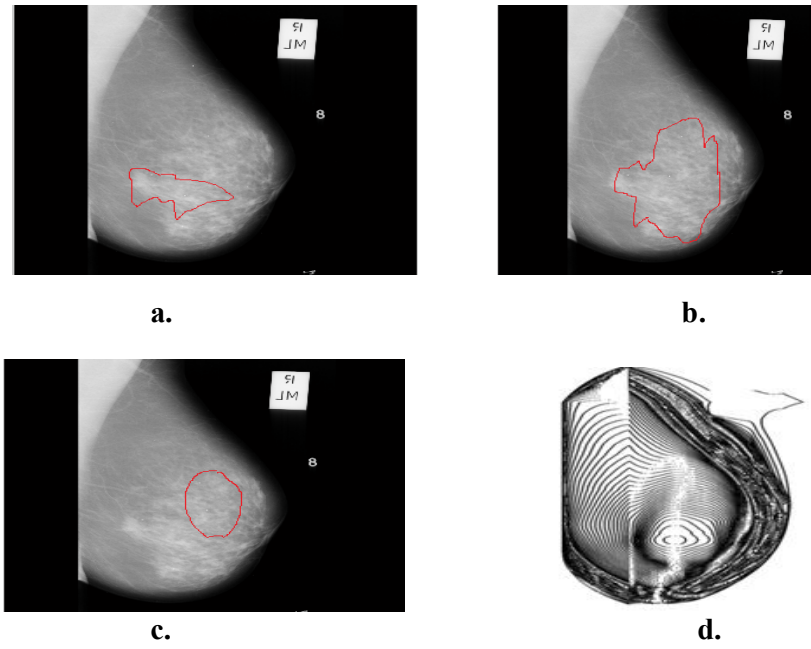
a. **b.**

Fig(4.2): Shock Filter On Normal Mammogram (**a.** Normal Mammogram **b.** Mammogram After Shock Filter).

The value of contrast on abnormal mammogram is greater than normal mammogram , making it easier to detect abnormal region on mammogram.

When compare shock filter with previous studies , found that has high ability to remove background from mammogram , increasing contrast (from 1.5 in previous studies to 2.25 in shock filter) and detect edges of mammogram, shown in figure(4.1.c).

The following figure (4.3.a.b.c) shows the variation between specialists to detect abnormal region in each mammogram. also detection of abnormal region on mammogram using Shock Filter, shown in figure (5.3.d). were found that the diagnosis by shock filter is better than diagnosis by specialists , has been verified with the ground truth given in the database (mini-MIAS database),shown in figure(4.3).



a. **b.**

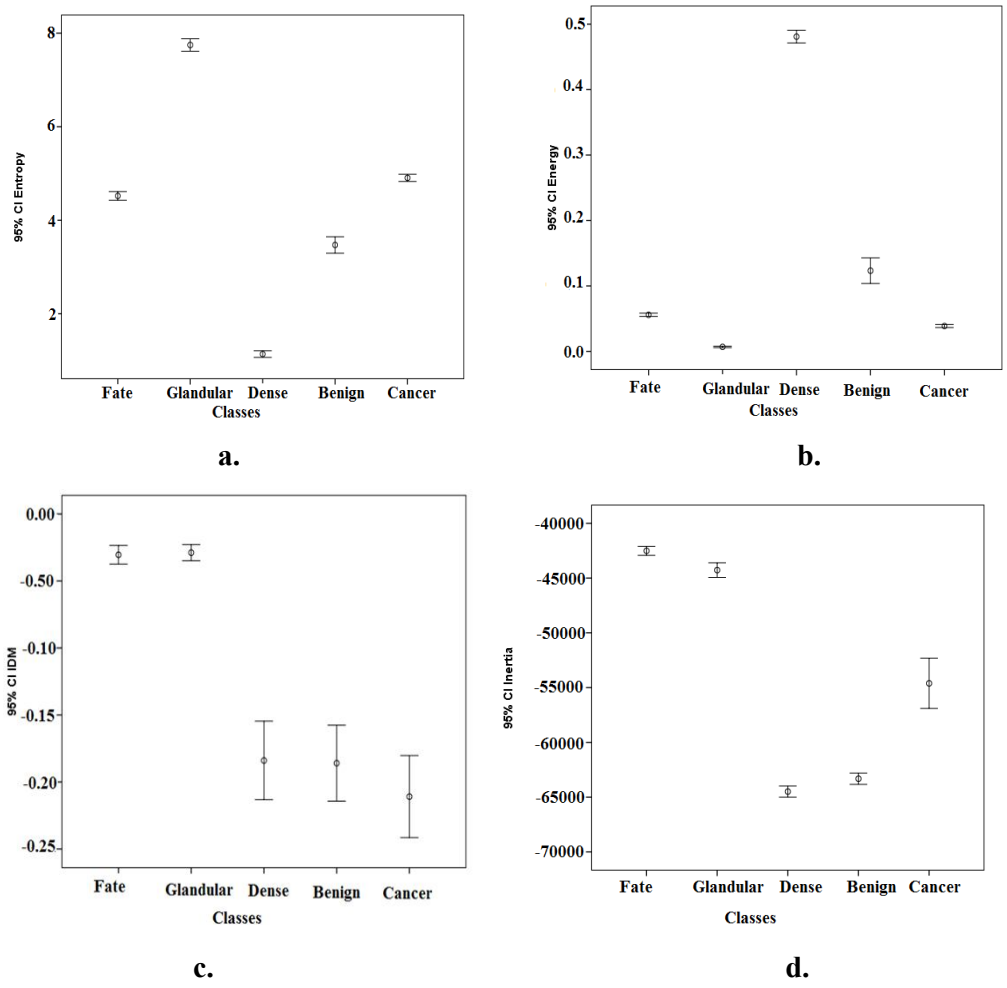
c. **d.**

Fig(4.3): Mammogram As Diagnosed By Specialists (**a.** Specialist A **b.** Specialist B **c.** Specialist C **d.** Mammogram after Shock Filter)

4.2 Feature Extraction

When was applied A novel Logical algorithm and Texture Features on sub-image and was extracted the features, then was collected this features in SPSS was used error bars to evaluate this features , shown in figure (4.4) was found that entropy is high in glandular and it is low in dense and is median in fate and cancer benign which measures ,because it is a measure randomness between tissues, shown in Figure (4.4.a). then energy is a measure of local homogeneity and therefore it represents the opposite of the Entropy. It is high in dense , it is low in glandular and median in fate ,cancer and benign , shown in figure(4.4.b). Inverse difference moment is the measure of local homogeneity, observed that it has large deviation in some classes and it has small deviation in other classes ,also it has very small values so it cannot discriminate between classes shown in figure (4.4.c) wherefore it was excluded. Inertia be high in smooth surfaces and it be low in rough surfaces, was found that it is high in dense and it is low in fate because fate consist of more details

.inertia can be discriminate between cancer and other classes, shown in figure(4.4.d)



Fig(4.4): Evaluation of Features Extracted By SPSS(**a.** Entropy **b.** Energy **c.** Inverse Different Moment **d.** Inertia)

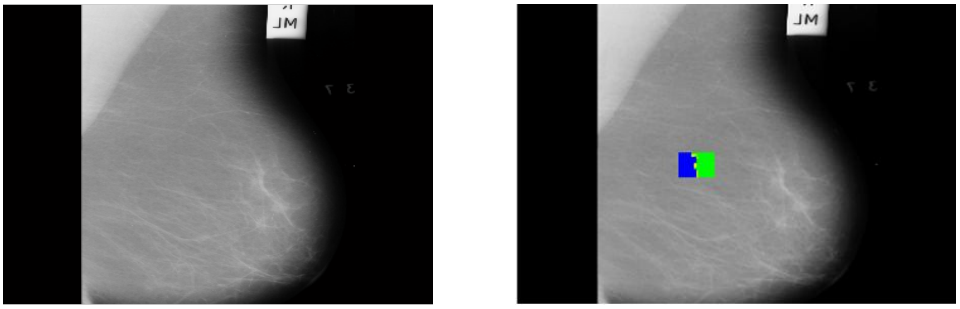
4.3 Classification

In this study was used two methods for classification abnormal region on mammogram Learning Vector Quantization (LVQ) and Learn discriminate analysis(LDA).

LVQ was applied on normal feature (fate, glandular and dense) achieved accuracy for training data 35% and accuracy for test data 22%.

In figure(4.5) show normal tissue on mammogram and was chosen sub image (79 ×79) is fate tissue (blue =dense ,green =glandular and yellow =fate). show yellow region very small but it is expected to be

large because this region is fate tissue(from data base) this confirms the accuracy which was obtained. may be cause is number data which were used in training stage, because if were used large data may be access to better result. and may be the cause of this low accuracy are features which selected in this study.



a. **b.**

Fig(4.5): Recognition Normal Tissues On Mammogram by LVQ (**a.** Original Image **b.** Mammogram After LVQ).

When LVQ was applied on abnormal feature was achieved accuracy for training data 25% and accuracy for test data 12%. In figure(4.6) show abnormal region on mammogram and was chosen sub image (79×79) is abnormal region (same color of normal tissue +red =abnormal tissue) show red region very small but it was expected to be large because this region is abnormal tissue(from data base)this confirms the accuracy which was obtained. and observed that it has accuracy is lower than normal features because normal tissues are different, but abnormal tissue are median (contain different Characters and Characters of normal tissues).

a. **b.**

Fig(4.6): Recognition Abnormal Tissues On MammogrambyLVQ (**a.** Original Image **b.** Mammogram After LVQ).

LDA was applied on the features that are collected in SPSS to classify normal Sub Image .This method achieved 100% of original grouped cases correctly classified, shown in table(4.1).

Table(4.1): Classification Of Normal Features By LDA

Classes	Predicted Group Membership			Total
	Fate	Glandular	Dense	
Fate	50	0	0	50
Glandular	0	50	0	50
Dense	0	0	50	50
Fate	100.0	.0	.0	100.0
Glandular	.0	100.0	.0	100.0
Dense	.0	.0	100.0	100.0

Then applied LDA on normal features and abnormal features , this method achieved accuracy 96.8% of original grouped cases correctly classified, shown in table (5.3). And also are obtained of Classification function coefficients, shown in table (4.2).

Table(4.2): Classification Function Coefficients

Features	Classes				
	Fate	Glandular	Dense	Benign	cancer
Entropy	82.403	123.766	91.999	80.092	88.434
Energy	725.835	1044.574	1147.068	741.947	752.128
Inertia	-.004	-.005	-.005	-.005	-.005
Constant	-290.413	-587.196	-491.030	-348.017	-362.900

Table(4.3): Classification Of Normal And Abnormal Features

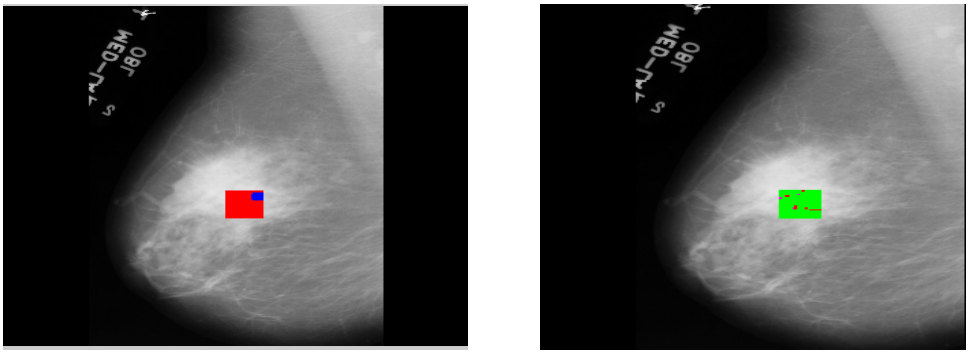
Classes	Predicted Group Membership					Total
	Fate	Glandular	Dense	Begin	cancer	
Fate	50	0	0	0	0	50
Glandular	0	49	0	0	1	50
Dense	0	0	50	0	0	50
Begin	0	0	0	45	5	50
cancer	1	0	0	1	48	50
Fate	100.0	.0	.0	.0	.0	100.0
Glandular	.0	98.0	.0	.0	2.0	100.0
Dense	.0	.0	100.0	.0	.0	100.0
Begin	.0	.0	.0	90.0	10.0	100.0
cancer	2.0	.0	.0	2.0	96.0	100.0

The following figure (4.7) Shows the variation between specialist and LDA to detect abnormal region on mammogram. diagnosed of the specialist is normal mammogram (dense) and diagnosis of LDA is abnormal mammogram. has been verified with the ground truth given in the database (mini-MIAS database) was found this mammogram is abnormal. This confirms diagnosis LDA.

a. **b.**

Fig(4.7) The Variation Between Specialist And LDA (**a.** Diagnosed by Specialist **b.** Diagnosed by LDA).

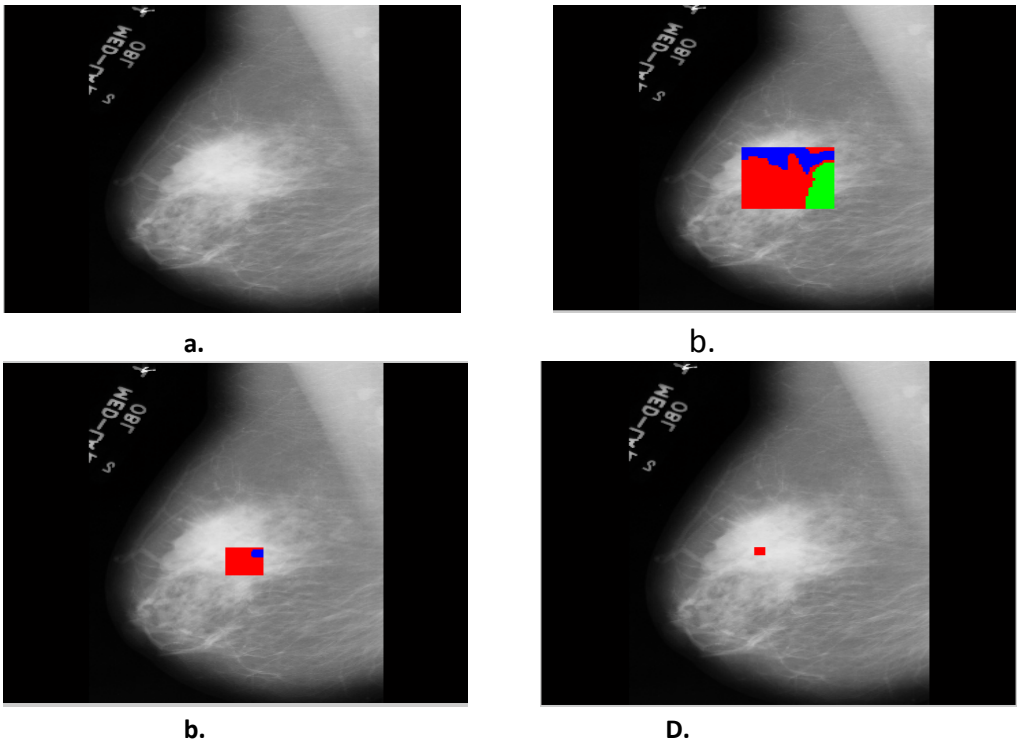
When comparing LDA with LVQ, was found that LDA higher accuracy 96.8% than LVQ, shown in figure (4.9), was found red region (abnormal region) is appeared clearly in LDA .



a. **b.**

Fig(4.9): Result of comparison between(LDA&LVQ)(**a.** LDA for Sub image **b.** LVQ for Sub image).

LDA was applied on different sizes of sub image of mammogram are (199×199 ,79×79 and 19×19),shown in figure(4.8) (The same colors previous).was found that (19×19) was given the best result to detect abnormal tissue on mammogram , because the tissues of breast interfere with each other. if it the bigger, can acquire overlap and if it smaller ,can acquire sub image is not clear .



Fig(4.8): Different Sizes Of Sub image Of Mammogram(**a.** Original Image **b.** Sub Image (199× 199) **c.** Sub Image (79× 79) **d.** Sub Image(19 × 19)).

4.4 Data analysis

The approach was tested on a set of 250 sub images . Table (4.4) gives the different parameter values TP (true positive), TN (true negative) ,FP (false positive) and FN (false negative) .

Table(4.4): Confusion Matrix

	Normal	**Abnormal**	**Total**
Normal	149 (TP)	1 (FP)	150
Abnormal	7 (FN)	93 (TN)	100

The classification results were validated using Matthews correlation coefficient(.933) , Sensitivity , Specificity , Accuracy and Positive and Negative Predictive value see table(4.5).

Table(4.5): Validation Of Classification Results

	Validation measures	**Percentage**
1	Sensitivity	95.5
2	Specificity	98.9
3	Accuracy	96.8
4	Positive Predictive Value	99.3
5	Negative Predictive Value	93

When compared between LDA with previous studies in Positive Predictive Value(PPV) and Specificity(SP) , was found that LDA are (PPV=99.3% &Sp=98.9%) either previous studies are(PPV=96 % & Sp =55.6%).

4.5 Graphical User Interface

After confirming the results of all the system, was created graphical user interface (GUI) which allows the users to take full advantage of the multitasking capabilities, shown in figure(4.10)

a. **b.**

Fig(4.10): GUI To Detect Abnormal Region On Mammogram**(a.** Input Image To GUI **b.** Image After LDA On GUI).

4.6 Summary

Summary of this chapter is to presented all results of this study, began from results of applying shock filter on normal and abnormal region on mammogram, compare between diagnosis of doctor with diagnosis of shock filter, then was presented results of texture feature when were applied on different classes, then was presented results of Linear Discermint Analysis and learning vector quantization and compare them, In the end , was created graphical user interface.

Chapter Five
Conclusion and Recommendation

5.1 Conclusion

Breast cancer has become a common health problem in developed and developing countries during the last decades and also the leading cause of mortality in women each year. early detection of the disease greatly improves the chance of survival. The overarching goal of this study was to develop a complete procedure in CAD system for detecting abnormal region in mammogram using image processing technique. Two methods have been use to achieve of this objective and was the first method is full automatic system using shock filter where it has ability to remove background , noise and the pectoral muscle ,increase contrast on abnormal region in mammogram . The second method is semi automatic system, it consist three phases: choose sub image then extract the features and classify this feature by Linear Discriminate Analysis(LDA) or Learning Vector Quantization . when compare LDA and LVQ, was found LDA better than LVQ for detection abnormal region on mammogram and where LDA achieved high accuracy 96.8%. .

5.2 Recommendation

- For further studies, it seems important to decrease the time of program execution of Shock Filter.
- The Proposed system can be implemented in local healthcare facilities in Sudan.
- The two applied methods in future can be combined to detect abnormal region on mammogram where shock filter can be used to determine region of interest (abnormal region), then features can be

extracted from the allocated region and lastly classify this features into cancer or benign .

- Other classifiers can be applied for higher classification accuracy .

References

[1] Rajashekar K.R. “Detection and Classification of tumors in Digital Mammogram”,International Journal of Computer Applications (0975 – 8887), 2012.

[2] Maha .D.A, Asmaa K.H, Nada A.M, " Computer-Aided Detection of Benign Tumors of the Female Breast ",thesis, University of Medical Sciences & Technology, Khartoum, 2008.

[3] Anna Soderstrom," Using Wavelet Methods to Extract Microcalcifications in Mammograms",thesis, Department of Computing Science of Umeoa University, Sweden,1997

[4] R.R.Janghel, Ritu Tiwari and Anupam Shukla.”Breast Cancer Diagnostic System using Hierarchical Learning Vector Quantization”, International Journal of Computer Applications (0975 – 8887),Access 2013.

[5] Linear Discriminat Analysis, http://www.investopedia.com , Access 2013.

[6] H. S. Sheshadri and A Kandaswamy, " Breast Tissue classification using statistic feature extraction of mammography", J. Medical Imaging and Information Sciences , Vol.23, No.3, pp.105-107, 2006.

[7] S. Mohan Kumar, "Classification of Microcalcification in Digital Mammogram using Stochastic Neighbor Embedding and KNN Classifier", International Journal of Computer Applications ,2012

[8] Norhène GARGOURI BEN AYED, Alima DAMAK MASMOUDI and Dorra SELLAMI MASMOUDI.“Robust Contrast Enhancement for Digital Mammography", International Journal of Computer Applications (0975 – 8887) Volume 56– No.9,2012.

[9] Sandeep Chaurasia and Prasun Chakrabarti "An Application of Classification Techniques on Breast Cancer Prognosis",International

Journal of Computer Applications (0975 – 8887) Volume 59– No.3, 2012.

[10] T.A.Sangeetha and A.Saradha, " An Efficient Way to Enhance Mammogram Image in Transformation Domain ",International Journal of Computer Applications (0975 – 8887) Volume 60– No.2 ,2012.

[11] Nizar Ben Hamad, Khaled Taouil and Med Salim Bouhlel." Mammographic Microcalcifications Detection using Discrete Wavelet Transform",International Journal of Computer Applications (0975 – 8887) Volume 64– No.21, 2013.

[12] R.Ramani and S. Valarmathy, " Breast Cancer Detection in Mammograms based on Clustering Techniques- A Survey ",International Journal of Computer Applications (0975 – 8887) Volume 62– No.11, January 2013.

[13] Female breast anatomy, www.myoclinic.com, Access 2013

[14] Alphonsine Imaniraguha," Implementation Of Medical Imaging With Telemedicine For The Early Detection Of Diagnosis Of Breast Cancer To Women In Remote Areas", thesis, College of Applied Science & Technology (CAST) ,The Rochester Institute of Technology, United States,2008.

[15] Bernd Jahne, "Digital image processing concepts algorithms, and scientific applications", Springer-Verlag Berlin, Hidelberg 1991.

[16] William K. Pratt, "Digital Image Processing _PIKS Inside", third edition, john Wiley and Sons, Canada, 2001.

[17] Tinku Acharya, Ajoy K. Ray, "Image Processing Principles and Applications". John Wiley and Sons, Inc. 2005.

[18] Carissa Erickson," Automated Detection of Breast Cancer Using SAXS Data and Wavelet Features", thesis, Division of Biomedical Engineering, University of Saskatchewan, Saskatoon,2005.

[19] Michel M. ,Yves M. ,Georges O. and Jean M." Wavelet Tool Box For Use With Matlab", second edition , The Math Works, Inc. , 2001.

[20] JOHN L. SEMMLOW," Biosignal and Biomedical Image Processing MATLA B-Based Applications", Marcel Dekker, Inc. , New York,2004.

[21] K.Sreedhar and B.Panlal, " ENHANCEMENT OF IMAGES USING MORPHOLOGICAL TRANSFORMATIONS ", International Journal of Computer Science & Information Technology Vol 4, No 1, 2012.

[22] Heang P.C, Dalton Wei, Mark A.H et al, "Computer aided classification of mammographic masses and normal tissue: Linear discriminate in texture feature space",thesis,Department of Radiology, University of Michigan, USA, 1995.

[23] SPSS Tutorial, www.hmdc.edu/projectslspsstut.shtml,2013.

[24] SPSS and surveys, www. folioasq.pbworks.com,2013

[25] coherence shock filter, www.mia.uni-saarland.de,2013

[26] Texture Features, http//: www.cs.auckland.ac.nz ,2014

[27] Leonardo O.M, Geraldo B.J, Arist O.C, Anselmo C.P, and Marcelo Gattass, " Detection of Masses in Digital Mammograms using K-means and Support Vector Machine ", Electronic Letters on Computer Vision and Image Analysis, Vol 8,No2, 2009

[28] Haralick, M. Robert, "Textural Features for Image Classification", IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-3, No. 6,1973.

[29] Mini MIAS, http://peipa.essex.ac.uk/ipa/pix/mias , Access 2012

[30] Pragati Kapoor,' Image Segmentation and Asymmetry Analysis of Breast Thermograms for Tumor Detection', International Journal of Computer Applications, Volume 50 – No.9, 2012.

[31] John L. Semmlow, "Biosignal and Biomedical Image Processing MATLA B-Based Applications",Marcel Dekker, Inc, U.S.A. ,2004

Appendix A. Shock Filter

```
%%shock Filter
% read the input image
x=input('input image ','s');
a=imread(x);
figure, imshow(a)
%=========================================
% gradient of image
[gradIX,gradIY]=gradient(double(a));
% second order gradient of image
[gradIX,gradIY]=gradient(gradient(double(a)));
figure,imshow(absGradI2);title('laplace Image');
It=-sign(absGradI2);
[m1,n1]=size(It);
%=========================================
[m2,n2]=size(s);
for i=1:m1
        for j=1:n1
                if It(i,j)>=0;
                    if ((i+k==w)&&(j+l==v))
                            P(w,v)=1;
                     End
%=========================================
% Gaussian filter parameter
% Gaussian filter
G=fspecial('gaussian',15,sigma);
figure,imshow(P);title('gaussian Image');
```

Appendix B. Feature extraction

```
% input matrix
x=input('input image ','s');
I=imread(x);
figure, imshow(I)
%================================%Imcroup for image
hold on
 [x1,x2]=ginput(1);
x=imcrop(I,[x1,x2,49,49])
 [m,n]=size(x);
 %======Applied A novel Logical algorithm :
[LD,HD,LR,HR] = WFILTERS('coif1')  %family of wavelet coiflet
h0=LR;
L=1;
p=synthesizeIM(x,h0,L);           %level one
%============================ windowing and classification
for i=d:5:m
   for j=d:5:n
      c=c+1;
      w=p(i-d+1:i,j-d+1:j);
      p1=jointp(w);
      en(c)=entropy01(p1);
      eg(c)=energy(p1);
      idm(c)=invdiff(p1);
      in(c)=inertia(p1);
       end
end
```

Appendix C. Classification

C.1 Training of Learning Vector Quantization(LVQ)

```
%%%Training program
x1=[input feature of fate tissue  ];
 x2=[feature of glandular tissue];
 x3=[ feature of dense tissue];
 x4=[input feature of cancer tissue  ];
 x5[input feature of benign tissue ];
%=========================================
 x=[x1,x2,x3,x4,x5];
 TC=[1 2 3 4 5 ];  % target tissue
%=========================================
net = newlvq(x,4,[.2 .2 .2 .2 .2 ], 0.01);
 net.LW{2,1};
net.trainParam.epochs =50;
net = train(net,x,T);
Y = sim(net,x);
%=========================================
```

c.2 Test Program OF LVQ

```
%%% test program
function Yc1=P2LVQ(P)
P1LVQ;
Yc1 = vec2ind(Y);
```

C.3 Program LVQ to detect abnormal region

```
 %%%% LVQ to detect abnormal region
clc
close all
% input matrix
```

```
x=input('input image ','s');
I=imread(x);
figure, imshow(x)
%=========================================
h0=LR;
L=1;
p=synthesizeIM(x,h0,L);
%=========================================
 for i=d:5:m
        for j=d:5:n
            c=c+1;
            w=x(i-d+1:i,j-d+1:j);
            p1=jointp(w);
            en(c)=entropy01(p1);
            eg(c)=energy(p1);
             in(c)=inertia(p1);
             z=[en(c),eg(c),in(c)];
            Yc = P2LVQ(z);
%=========================================
if Yc  == 1
          plot ('y')
        elseif Yc == 2
                plot ('g')
        elseif Yc == 3
               plot ('b')
        end
```

C.4 ***Linear Discriminant Analysis***

*%%%%***Linear discriminant analysis***

```
x=input('input image ','s');
```

```
I=imread(x);
figure, imshow(I)
 hold on
%==================================%Appliedinverse wavelet
h0=LR;
L=1;                                          %level one
p=synthesizeIM(x,h0,L);                       %Applied synthesize
%=========================================
for i=d:5:m
          for j=d:5:n
                   w=p(i-d+1:i,j-d+1:j);
                    p1=jointp(w);             %Joint probability
                    en(c)=entropy01(p1) ;  %calculate entropy
                    eg(c)=energy(p1) ;  %calculate energe
                    in(c)=inertia(p1) ;   %calculate inertia
      Y(1)=-.004 * in(c) + 725.835 * eg(c)  + 82.403 * en(c) - 290.413;
      Y(2)=-.005* in(c) + 1044.574* eg(c)  +  123.766* en(c) - 587.196;
      Y(3)=-.005 * in(c) + 1147.068 * eg(c)  + 91.999 * en(c) - 491.030;
                 if id == 1
                           plot ('y')
                  elseif id == 2
                           plot (k,v,'g')
                  elseif id == 3
                            plot ('b')

                       end
```

Appendix D. Graphical User Interface

```
%%%Graphical User Interface
%%GUI
F_MainFigure = figure('Color',[1 0.2 0.6], ...
'Name', 'Detection of abnormal region on mammogram ', ...
'NumberTitle', 'off');
%==================================Frame in low design
h_f_background = uicontrol('Parent', F_MainFigure, ...
'BackgroundColor',[1 .3 .8], ...
'Position',[2 2 500 50], ...
'Style','frame');
%=========================================
b = uicontrol('Parent', F_MainFigure,...
'Callback','w266', ...
'Interruptible', 'off', ...
'Position',[20 10 50 30], ...
'String','SHF');
%=========================================
b = uicontrol('Parent', F_MainFigure,...
'Callback','LDA1', ...
'Interruptible', 'off', ...
'Position',[70 10 50 30], ...
'String','IDWT ');
```