IBCIS: INTELLIGENT BREAST CANCER IDENTIFICATION SYSTEM

A THESIS SUBMITTED TO THE
GRADUATE SCHOOL OF APPLIED SCIENCES
OF
NEAR EAST UNIVERSITY

By

ABED EL KADER HELWAN

In Partial Fulfillment of the Requirements for the Degree
of Master of Science

in

Biomedical Engineering

NICOSIA, 2015

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

Name, last name: Abed el kader Helwan

Signature:

Date:

ACKNOWLEDGMENT

ABSTRACT

This thesis aims to develop an intelligent breast cancer identification (ICBIS) system based on image processing techniques and neural network classifier. Recently, many researchers have developed image recognition systems for classifying breast cancer tumors using different image processing and classification techniques. The challenge is the extraction of the real features that distinguish the benign and malignant tumor. The classifications of breast cancer images have been performed using the shape and texture characteristics of the images. The asymmetry, roundness, intensity levels and more are the exact shape and texture features that distinguish the two types of breast tumors. Image processing techniques are used in order to detect tumor and extract the region of interest from the mammogram. The following data processing operations have been done for detection of images: thresholding, filtering and adjustments, canny edge detection, and some morphological operations. Shape and texture features are then extracted using GLCM (Gray-Level Co-Occurrence Matrix) algorithm in order to accurately classify the mammograms into normal, benign, and malignant tumors. GLCM is a factual system for inspecting composition offers that consider the spatial relationship of pixels. It describes the composition of an image by figuring how frequently combines of pixel with particular qualities and in a tagged spatial relationship happen in a image. The images used are obtained from a public database available on the internet that contains mammography images (DDSM). Based on this data the system was simulated using Matlab software and the experimental results show a good identification rate of 97%.

Keywords: Breast cancer, malignant tumor, canny edge detection, morphological operations, GLCM

ÖZET

Bu çalışma görüntü işleme teknikleri ve sinir ağı sınıflandırıcı dayalı akıllı meme kanseri tanımlama (ICBIS) sistemini geliştirmeyi amaçlamaktadır. Son zamanlarda, birçok araştırmacı farklı görüntü işleme ve sınıflandırma teknikleri kullanılarak meme kanseri tümörlerini sınıflandırmak için görüntü tanıma sistemlerini geliştirdi. Burdaki zorluk iyi ve kötü huylu tümörleri ayırt etmektir. Meme kanseri görüntülerin sınıflandırmalar görüntülerinin şekli ve doku özellikleri kullanılarak yapılmıştır. asimetri, yuvarlaklık, yoğunluk düzeyleri ve daha meme tümörlerinin iki tür ayırt tam şekil ve doku özellikleri vardır. Görüntü işleme teknikleri tümör tespit ve mamografi ilgi bölgeyi çıkarmak için kullanılır. eşikleme, filtreleme ve ayarlamalar, tedbirli kenar algılama, ve bazı morfolojik operasyonlar: Aşağıdaki veri işleme işlemleri görüntülerin tespiti için yapılmıştır. Şekil ve doku özellikleri daha sonra doğru normal huylu ve kötü huylu tümörlerin içine mamogram sınıflandırmak amacıyla GLCM (Gri-Seviye Ortak Oluşumu Matrix) algoritması kullanılarak ayıklanır. GLCM piksel mekansal ilişki düşünün doku özelliklerini inceleyeren istatistiksel yöntemdir. Bir görüntüde ne sıklıkla belirli değerleri ve belirli bir mekansal ilişki içinde piksel çiftleri hesaplayarak bir görüntünün dokusunu karakterize eder. kullanılan resimler mamografi görüntüleri içeren internet üzerinde bir kamu veritabanı (E.coli) elde edilmiştir. Bu verilere dayanarak sistem Matlab yazılımı kullanılarak simüle edilmiş ve deneysel sonuçlar % 97 iyi bir tanımlama oranı göstemiştir.

Anahtar Kelimeler: Meme kanseri, malign tümör, tedbirli kenar algılama, morfolojik işlemler, GLCM

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
INTRODUCTION

1.1 Overview

This chapter is an introduction of the thesis. It explains the basic concept used in the thesis, provides some medical information about the breast cancer, by providing some numerical data about the number of deaths due to breast cancer. In addition, it discusses some previous related research works. At the end, this chapter lists the basic aims of the thesis and discusses the structure as well.

1.2 Introduction

The breast cancer is about the most common types of cancer among women worldwide. The diseases are widely distributed among women in South Africa, according to the Cancer Association of South Africa (Cansa, 2012) and according to World Health Organization. Breast cancer is also the top cancer in women in both the developed and the developing world. According to the World Health Organization (WHO) breast cancer was the reason of 502,000 deaths in 2005 alone. Furthermore, in western countries breast cancer represents about 25% to 30% of the total incidence of cancers in women and is the main cause of 15% to 18 % of mortality (WHO, 2008). Worldwide, 1.301.867 of new cases of breast cancer is registered and deaths were 464.454 followed by other types of cancer (WHO, 2008). Breast cancer is a dangerous medical condition needs to be diagnosed and early detected in order to prevent its growth and reduce the percentage of deaths caused by it. Breast cancer screening can be done using different imaging techniques. The most common screening technique is the mammography. This kind of imaging technique is a specific form of radiography that uses radiation lower than those of conventional radiography such as routine x-ray. The mammography produces breast images called mammograms in order to diagnose and detect the presence of intruders or abnormal structures in the breast (Zaman, 2005).

1.3 Literature Review

Many and different methods have been used for the detection of breast cancer using image processing techniques. In order to come out with a new and unique one, there must be a review of the previous work related to our topic. Faizon and Sun (Faizon et al., 2000) proposed a method for breast cancer detection using thresholding and tracking to identify the breast border, but no discussion the accuracy of the results was presented. Their paper described some preliminary works in the analysis of asymmetries in digitized mammograms. They proposed a method for enhancing the asymmetries. The method is to first register, and then bilaterally subtract two mammograms of the left and right breast side in the medio-lateral view. Then, these asymmetries are analyzed in order to provide a tool for computer aided diagnosis (CAD). A system was proposed by (Padayachee et al., 2007) for identification of the barest edge using areas enclosed by the ISO - intensity contours. They used a different image processing techniques in order to identify the breast cancer in a mammogram. Such techniques are first Thresholding which involves selecting a single gray-level from an analysis of the gray-level histogram, and then segment the mammogram into the background and breast tissue in order to extract the region of interest. Another technique used by J. Padayachee is Iso-intensity breast edge detection. This method is rarely used by researchers and itsencounters, drawing of borders by radiologists to quantitatively evaluate the results of the automated borders. ISO-intensity contouring was also used by the authors in orderto select the optimal gray-level threshold for the breast border.

Another author called Anna Rejani (Rejani, 2009) developed an algorithm to assist in identifying breast cancer in its earlier stages to not grow or develop. In that algorithm, there was a combination between several image processing techniques like imagenegative, thresholding and segmentation techniques for detectionof tumor in mammograms. A system for the a system for classification of mammograms by the breast composition was proposed by Silva and Menotti (Silva et al., 2012). In their paper, a breast classification was based on the density of mammograms. The density method of calculation in this paper was different than others, since the texture descriptors were extracted and analyzedfor the representation of breast tissue density on mammograms. The developed system was experimentally tested and the obtained accuracy was 77.18%.

An Automated Approach for qualitative assessment of breast density and lesion feature extraction was suggested by Spandana Paramkusham (Paramkusham, 2013). This study provided a new method for measuring breast density, in which the density is the number of dense tissue pixels divided by the number of pixels in the Breast. According to their study, they classified the breast cancer into three classes: normal, benign, and malignant. Also, they provided percentage interval for each class. The authors also used other techniques in order to segment the breast cancer and extract the meaningful features such as extraction of breast borders, segmentation of mass, and morphological thinning and dilation.

1.4 The Aim of the Thesis

The aim of the thesis is the designing of breast cancer identification system using texture and shape characteristics of the breast images. The thesis is a part of the ongoing researches for detecting and diagnosing breast cancer that aim to reduce the rate of occurrence of that disease and detect it in its earlier stages in order to treat it prior to its growth and development. However, the thesis aims to use different and additional methods and techniques to reach the desired purpose: detecting breast cancer and classifying it into three main classes: Normal (no cancer), Benign, and Malignant. My thesis is based on the combination of image processing techniques and artificial neural networks, in order to come out with an effective detection system that is capable of detecting breast cancer in its early stages, and classifying it into three classes. Different image processing techniques such as RGB to grayscale conversion, image filtering using different types of filters (median and Gaussian filters), image adjustment, image thresholding, edge detection using canny operators, morphological techniques (erosion, image opening and closing). The image processing techniques are embedded into the system in order to:

- Enhance the image quality

- Filter the image and suppress the noise

- Enhance and detect edges

- Detect and segment the tumor region

- Use morphological techniques

- Extract the region of interest (ROI).

- Extract shape and texture features

- Classify the tumor

1.5 Thesis Structure

The rest of this report is divided into 4 chapters and organized as follows:

Chapter 1 introduces clinical background of breast cancer in several aspects including the anatomy of the breast, the disease nature and definition, as well as the causes and symptoms of the disease. Moreover, the classification of the breast cancer is described and illustrated in this chapter, as well as the diagnosis methods.

Chapter 2 provides a detailed explanation about the first phase of IBCIS which is the image processing phase. In this chapter, we explain the image processing phase using graphs, figures, and flowcharts in order to explain our new developed algorithm for the identification of breast cancer. Each image processing technique is explained in details with examples and figures from the database images.

Chapter 3 discusses the second phase of the designed system called classification phase that is based on an artificial neural network. In this chapter, we define neural network (NN) and explain its concepts, in addition the learning algorithm of NN that is used in IBCIS which is the backpropagation algorithm. Moreover, the chapter 3 provides the training results of the system using tables, figures and curves such as the learning curves. The performance of designed system and the experimental results are also discussed in this chapter.

Chapter 4 is a results of comparison of IBCIS with previously designed breast cancer identification systems for the same purpose, which is the intelligent identification of breast cancer using a neural classifier.

1.6 Summary

In this chapter, we introduce the research work by defining its aims and motivations. We discussed some previous work related breast cancer identification systems. Moreover, we discussed the medical background of the breast cancer, then formulated the aim of the thesis and define its structure in details.

CHAPTER TWO
CLINICAL BACKGROUND

2.1 Overview

This chapter provides a medical background of the breast cancer. It discusses the anatomy of the breast, as well as defines the breast cancer, including its causes, symptoms and current diagnosis methods. Moreover, in this chapter, we discuss the different classes of the breast cancer: their features, types, and symptoms. The detection and diagnosing techniques are also presented.

2.2 Breast Anatomy

Breasts are comprised of fat and breast tissue, and additionally nerves, corridors, veins, and connective tissue to hold everything set up. The principal midsection muscle (the pectoralis muscle) is found between the bosom and the ribs in the midsection divider. Figure 1.1 demonstrates the anatomical structure of the breast (Jatoi et al., 2010).
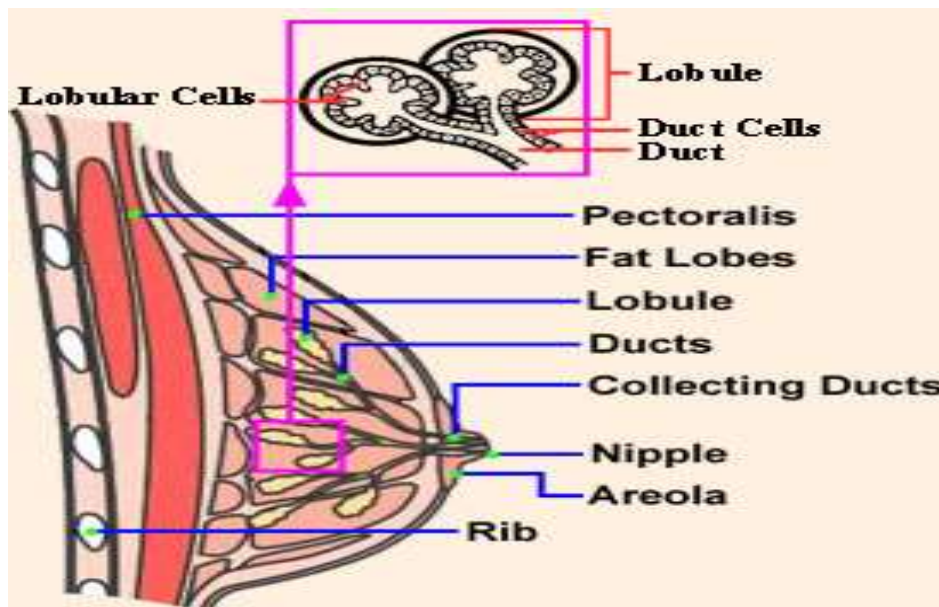
Figure 1: Breast structure (Jatoi et al., 2010)

Breast tissue is a complex network of lobules (small round sacs that produce milk) and ducts (canals that carry milk from the lobules to the nipple openings during breastfeeding) in a pattern that looks like bunches of grapes. These "bunches" are called lobes.
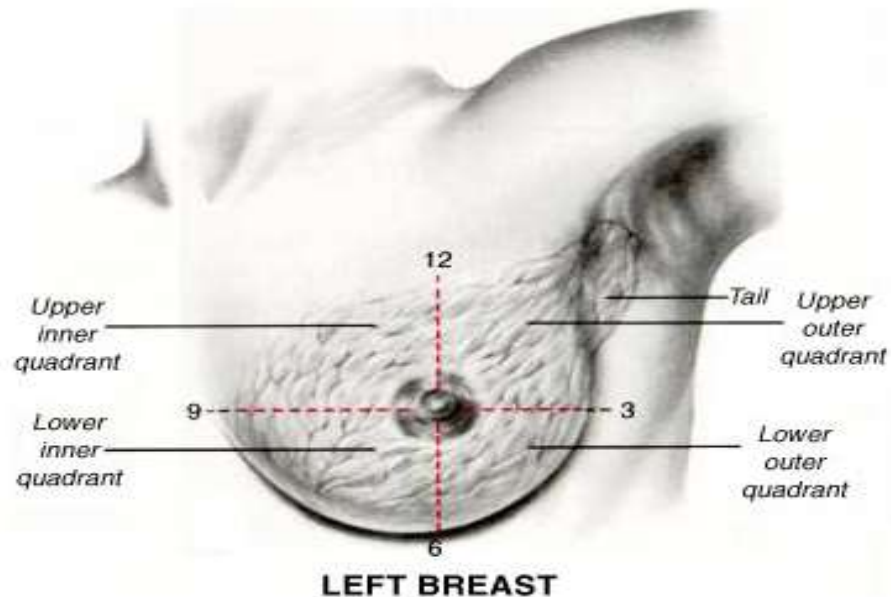
Figure 2: Breast Quadrants: UOupper outer, Lo lower outer, LI lower inner, UI upper inner (Anthony, 2010)

The breast is separated into quadrants: upper external, lower external, lower inward, upper internal. A vertical line and an even one converge at the areola. The most noteworthy amassing of the glandular tissue can be found in the upper external quadrant. The focal altered share incorporates the areola and areola (Figure 2). The quadrants positions are situated by numbers focused around the clock face (Jatoi et al., 2010).

Breast tissue is overlying the chest (pectoral) muscles. The breass of women are made of particular tissues that create milk (glandular tissue), and additionally fat tissue. The measure of fat decides the span of the bosom.

The association is a piece of creating milk from the bosom to 15 to 20 areas, called flaps. Inside every projection is a littler structure, called lobules, where milk is created. Milk goes through a

system of little tubes called channels. Channels of correspondence, and meet up in huge channels, which turn out at last skin in the areola and approached the dull zone of skin around the areola and areola.
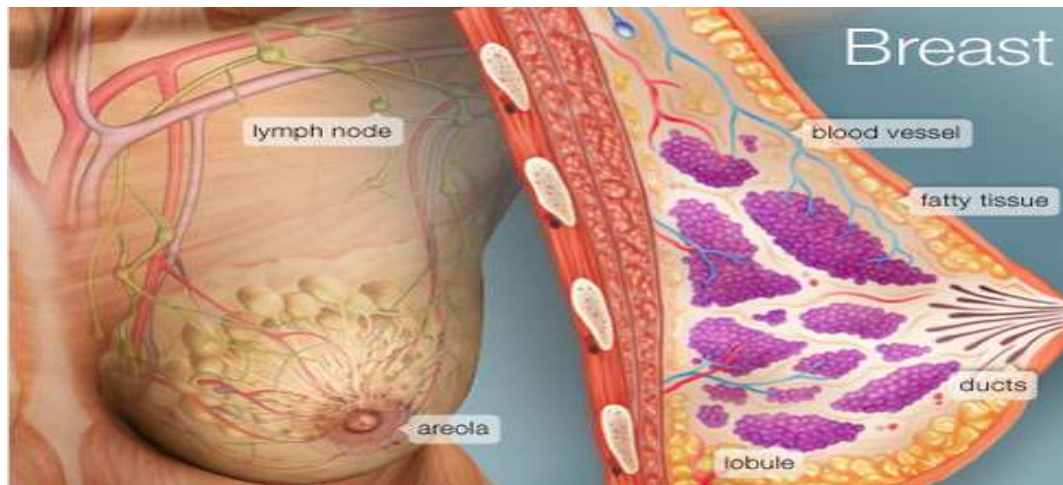
Figure 3: Breast internal structure (Anthony, 2010)

Furthermore, the connective tissue and ligaments of the breast provide support to the breast and give it its shape. Nerves provide sensation to the breast. The breast, as all body organs and parts, contains blood vessels, lymph vessels, and lymph nodes.

2.3 Breast Cancer

Tumor, as a rule, is a gathering of ailments that cause cells in the body to anomalous change and become out. Most sorts of malignancy cells in the long run structure a protuberance or mass called a tumor. These masses are generally benevolent; that is, they are not harmful, do not become wildly or spread, and are not life-debilitating and do not result in death. Growths are the locales where the tumor starts and spreads. Bosom malignancy is a dangerous cell development in the breast. The tumor may spread to different regions of the body if left untreated (Anthony, 2010). Breast growth starts in bosom tissue, which is comprised of:

• Glands for milk generation, called lobules,

• Ducts that unite the lobules to the areola, and

• Fatty, connective, and lymphatic
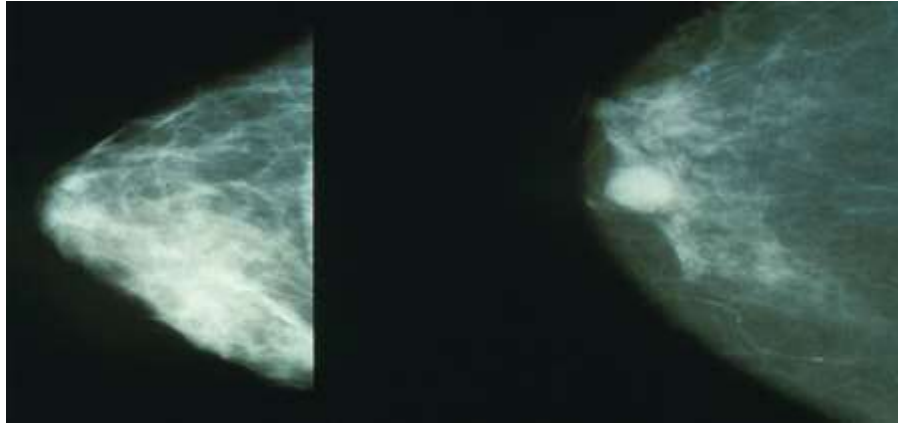
Figure 4: A mammogram that shows a normal breast (left), and cancerous breast (right) (Anthony, 2010)

2.3.1 Causes of breast cancer

Breast malignancy is the second driving reason for death among ladies. Still, nobody can know the precise reasons for a bosom growth. Specialists and scientists at times know why one lady creates a breast tumor and an alternate doesn't, and most ladies who have bosom disease will never know the careful reason. What we do know is that breast growth is constantly brought on by harm to a cell's DNA as all different sorts of diseases (National Breast Cancer Foundation, 2012).

In other words, The breasts of women are made of particular tissues that create milk (glandular tissue), and additionally fat tissue. The measure of fat decides the span of the bosom.

The association is a piece of creating milk from the bosom to 15 to 20 areas, called flaps. Inside every projection is a littler structure, called lobules, where milk is created. Milk goes through a system of little tubes called channels. Channels of correspondence, and meet up in huge

9

channels, which turn out at last skin in the areola and approached the dull zone of skin around the areola.

2.3.2 Types of Breast Cancer (Anthony, 2010)

There are many and different types of breast cancer. They can be divided into two categories: invasive and in situ. These separate types of breast cancer are based on the way the cancer cells look under the microscope. The breast cancer may be a combination of these types of the invasive and in situ cancers, then it forms a breast tumor or the cancer cells may not form a tumor at all.

- **Carcinoma**

This is defined as a cancer that begins in the lining layer (epithelial cells) of organs like the breast. Mostly all breast cancers can be called as carcinomas, thus they are categorized as either ductal carcinomas or lobular carcinomas.

- **Adenocarcinoma**

This is a kind of carcinoma that starts in the glandular (tissue secretes a substance). The conduits and lobules of the breast are glandular tissues since they make bosom drain, so the tumors that begin in these two ranges are frequently called adenocarcinoma.

- **Carcinoma in situ**

This kind of breast cancer is characterized as an early phase of tumor, when it is restricted to the layer of cells where it started. In bosom disease, in situ implies that the tumor cells stay kept to channels (ductal carcinoma in situ). The cells have not developed into (attacked) deeper tissues in the breast or spread to different organs in the body. The ductal carcinoma of the breast is now and then alluded to as non-obtrusive or preinvasive bosom disease in light of the fact that it may form into an intrusive breast malignancy if left untreated.

- **Invasive (infiltrating) carcinoma**

This is one of the perilous and hazardous bosom diseases since it is the particular case that has officially become past the layer of cells where it began (instead of carcinoma in situ). Most

breast growths are intrusive carcinomas—either obtrusive ductal carcinoma or obtrusive lobular carcinoma

- **Sarcoma**

This type of cancer is the one that that start in connective tissues such as muscle tissue, fat tissue, or blood vessels. Sarcomas of the breast are rare.

2.3.3 Symptoms of breast cancer (Anthony, 2010)

1. A change in how the breast or nipple feels

- A nipple delicacy or a protuberance or thickening in or close to the bosom or underarm region

- A change in the skin composition or an amplification of pores in the skin of the bosom (some depict this as like an orange peel's surface)

- A protuberance in the bosom (It's critical to recollect that all bumps ought to be researched by a medicinal services proficient, however not all knots are carcin(discussed next section) (Figure 4).

Figure 5: A change in nipple feeling (Jatoi et al., 2010)

2. A change in the breast or nipple appearance

- Any unexplained change in either the shape or size of the breast

11

- Dimpling anywhere on the breast

- Unexplained swelling of the breast (especially if on one side only) (Figure 5).

Figure 6: A change in the nipple appearance (Jatoi et al., 2010)

3. Any areola release especially clear release or bleeding release:

- It is likewise vital to note that a smooth release that is available when a lady is not breastfeeding ought to be checked by her specialist, despite the fact that it is not connected with bosom growth (Figure 6)

Figure 7: A nipple discharge (Jatoi et al., 2010)

2.3.4 Breast cancer screening/ diagnosing techniques

The most useful and effective breast screening and diagnosing machine is the mammography. It is used to aid in the early detection of breast diseases in women (Society of North America, 2013). This screening tool uses X-rays to produce images of the breast. These produced images

are called mammograms (MAM-o-grams), and they are used to find early signs of breast cancer such as clusters of calcium (microcalcifications) or dense masses(susangkoman, 2014). Figure 1 and 2 shows two mammograms of normal and cancerous breast cancer.
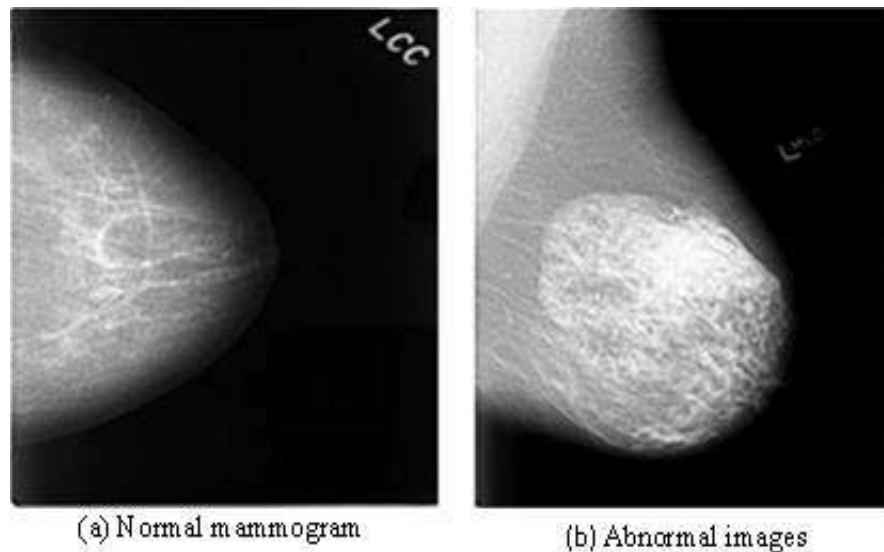
Figure 8: DDSM breast images

The mammography system recently has been advanced and it includes digital mammography, computer-aided detection and breast tomosynthesis.

- **Digital mammography**

This kind of mammography, just the x-beam film is supplanted by strong state locators that change over x-beams into electrical signs. These finders are much the same as the ones found in computerized cams. At that point, these electrical signs are changed over to pictures of the bosom that can be seen on machine Computer-aided detection (CAD):

These are systems that search for abnormal areas of density, mass, calcifications that indicate breast cancer in the images obtained from either conventional mammography or digital mammography.

- **Tomosynthesis**

This is a three-dimensional (3-D) breast imaging in which the x-beam tube in a mammography moves in a bend over the bosom amid the presentation. At the point when moving, a progression of flimsy cuts through the bosom are made, which considers enhanced identification of disease.

2.3.5 Breast Cancer Classification

Breast cancer can be classified into two classed: benign and malignant cancer. Most of the breast cancers start in the cells that line the ducts (ductal cancers). Some others start in the cells that line the lobules (lobular cancers), while a small number start in other tissues.

- **Benign breast cancer (lumps)**

A breast lump, purported benign breast canceris characterized a swelling or thickening in the breast. It is basic and nine out of 10 are generous (not carcinogenic). This sort of tumors won't transform into breast disease (Figure 9).
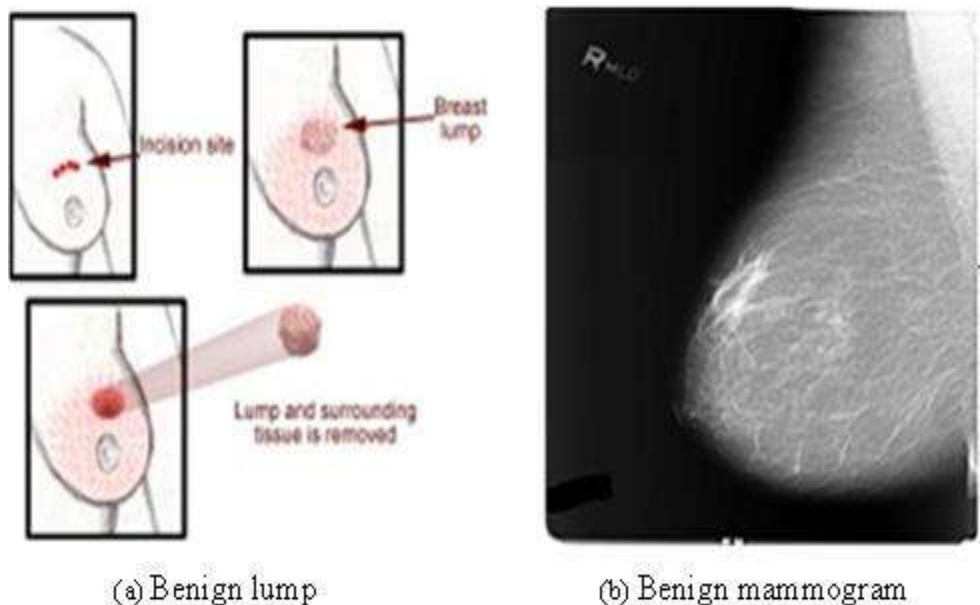
Figure 9: Benign tumor (Medscape, 2013)

- **Malignant breast cancer**

A dangerous breast malignancy or tumor is a gathering of disease cells that may develop or attack into encompassing tissues or spread too far off territories of the body. As talked about awhile ago, threatening bosom tumor is numerous shorts, for example, carcinoma, sarcoma, and ductal carcinoma. This sort of tumor needs to be evaluated and diagnosed by biopsy to affirm that it is a malignant tumor.
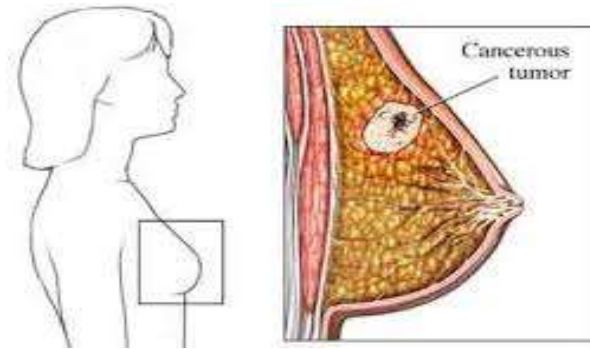
Figure 10: Benign tumor (Medscape, 2013)
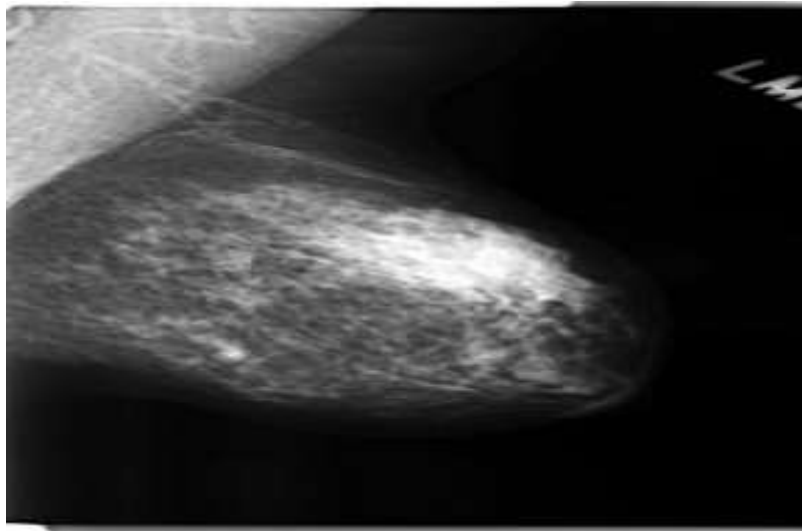
Figure 11: Malignant mammogram (Heath et al., 2001)

- **Features based comparison between normal, benign, and malignant breast tumors:**

Many features differ the benign and malignant cancer, such as color, size, symmetry etc…figure below shows the different geometric and visual features of both benign and malignant cancers in order to classify them accurately.
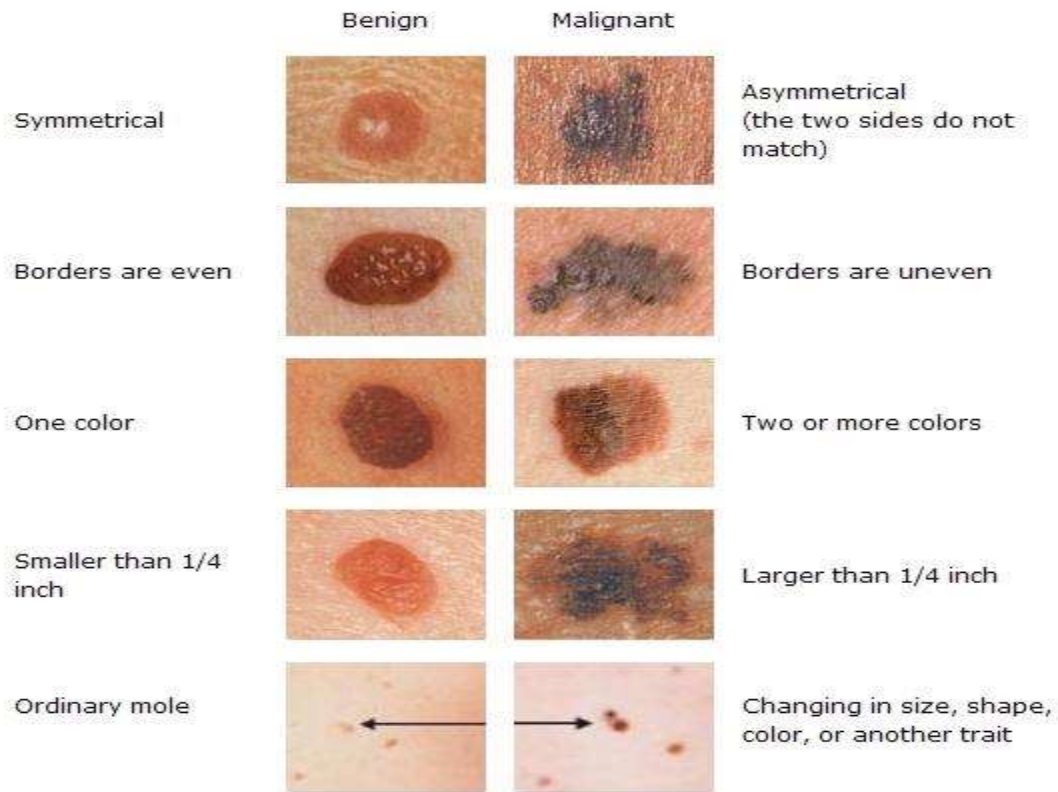
Figure 12: Features based comparison of benign and malignant breast tumor (Medscape, 2013)

2.4 Proposed Identification System

In our designed system for the identification of breast cancer, we first process the images using some image processing techniques in order to extract the region of interest (tumor). Then, we move to the next stage, which is the feature extraction. The GLCM (Gray level co-occurrence matrix) is used to extract some features which reflect the shape such as roundness, asymmetry… and intensity levels of the tumor, such as mean, standard deviation, energy, etc... These features

16

are then fed to a neural classifier, which classifies the images into 3 output classes: normal, benign, and malignant tumor.

2.5 Summary

In this chapter, we went through the anatomy of the breast in order to explain the breast cancer in details. We explained the breast cancer and its causes. Then, we went through the signs and symptoms of breast cancer and the available screening techniques currently used for diagnosing it. We discussed also the features of breast cancer that are used in our thesis for classification the cancer into normal, benign, and, malignant.

CHAPTER THREE
INTELLIGENT BREAST CANCER IDENTIFICATION SYSTEM: IMAGE PROCESSING PHASE

3.1 Overview

This chapter discusses the second phase of the designed system which is called the image processing phase. The chapter explains the image processing principles and basics, also explains the image processing technique in details. The description of the techniques used in data processing stage such as filtering, segmenting, morphological operations, and features extracting are given in details. Some flowcharts and figures used in the data processing stage of IBCIS are presented.

3.2 Image Processing Principles

A computerized picture f[m,n] can be portrayed in a 2D discrete space is gotten from a simple picture f(x,y) in likewise a 2Dcontinuous space utilizing a methodology of inspecting that is habitually alluded to as digitization (Gonzalez and Woods, 2002).
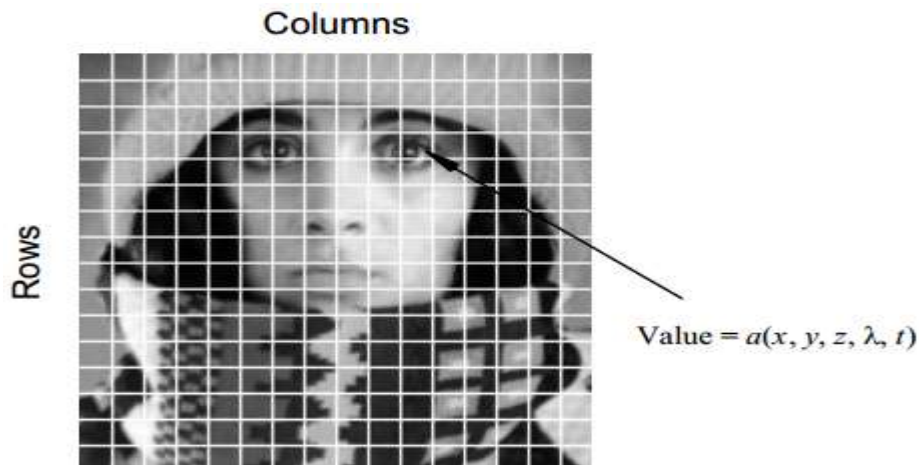
Figure 13: Digitization of analog image.

Image processing is defined as graphics techniques that can be operated only on images in order to process, analyze, and visualize images. It is some operations that take images as input, and produce images as output in return (Gonzalez and Woods, 2002).

3.3 Intelligent Breast Cancer Identification System Methodology

In this thesis, an intelligent breast cancer identification system is developed. The system is implemented using Matlab programing language (Matlab 2013 software tools). IBCIS is based on different image processing techniques used in order to identify the breast tumor. The breast images are obtained from DDSM (Heath, 2001); a public database available on the internet. The images are of size 221*358 pixels. They are processed and then rescaled to 250*250 pixels for the purpose of fast computing. Then, texture features are extracted using GLCM technique. Tn addition, the shape features are alse extracted from images in order to be fed into the next phase which the neural classification.

The images are first converted to grayscale using the luminosity method. Then they are adjusted to increase the pixel intensity in an image, so that the tumor can be clearer and brighter. The images undergo threshold computing for the purpose of segmenting the region of interest (tumor) located in the breast. The next step is to extract that tumor in order then to detect its edges using canny operators. We use also some morphological techniques such as dilation and opening in order then to extract the region of interest which is the tumor in the benign malignant cases. 7 features are then extracted from the region of interest image using Gray Level Co-occurrence Matrix; an algorithm used mostly in medical applications for extracting gray level features and some breast cancer related features such as standard deviation, mean, asymmetry, roundness, uniformity etc.. Figure 16 shows the flowchart of the developed identification system. It lists the methods used in the system. Figure 17 represents an abnormal breast image that undergoes all the processes discussed. Figure 18 illustrates a normal mammogram that undergoes all discussed processes till the extraction of its region of interest.

In the Figure 17, we used two types of filters in order to smooth the image and remove the noise. The use of two filters has been just to compare between them and find the best in removing noise

and smoothing image. Visually we can notice that the median filter is better for removing noise and enhancing the image quality.
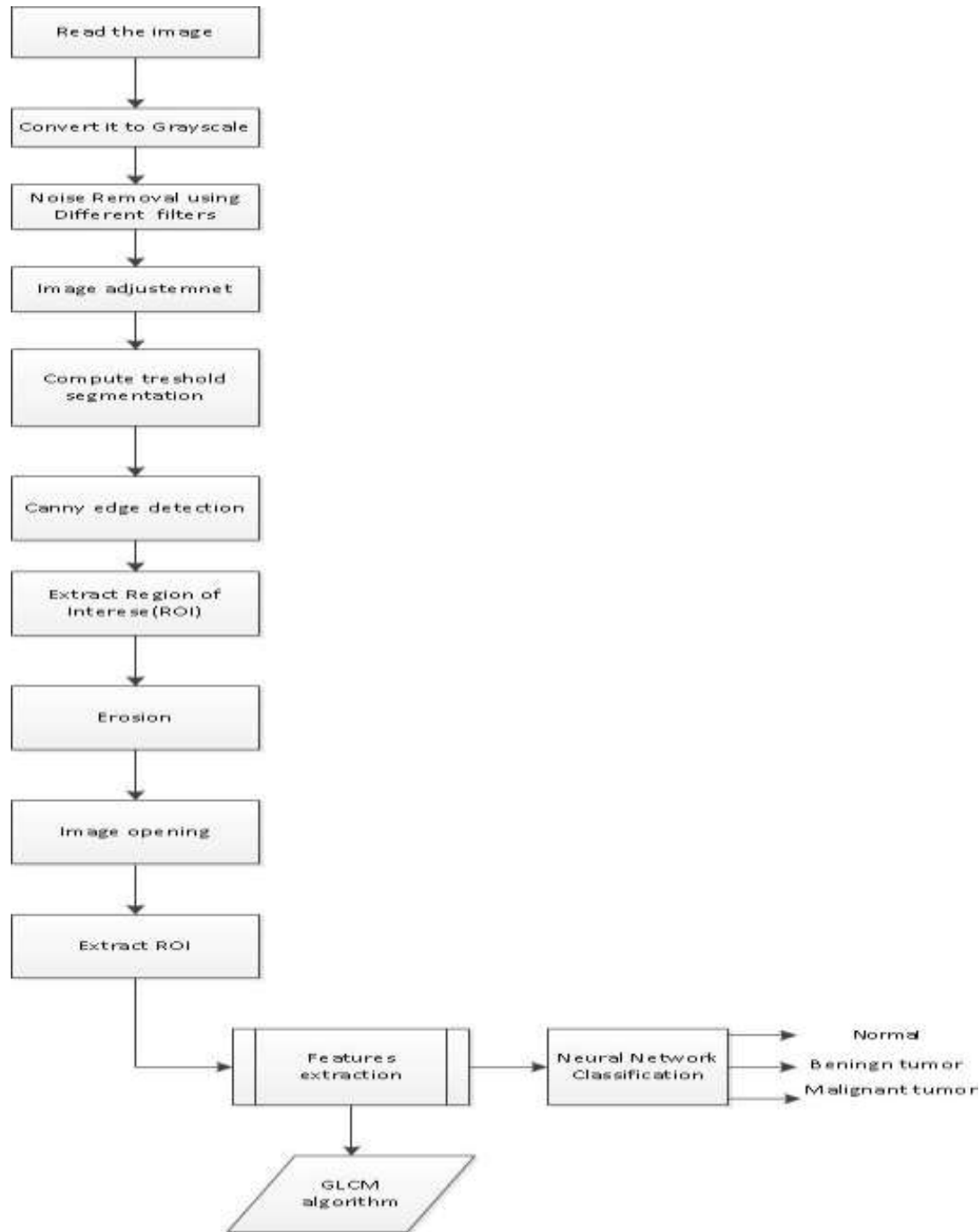
Figure 14: Flowchart of the proposed system

(a) RGB image   (b) Graysale image   (c) Average filtered image   (d) Median filtered image

(e) Adjusted image   (f) Tresholded image   (g) Segmented image   (h) Eroded image

(i) Opened image   (K) Extracted tumor
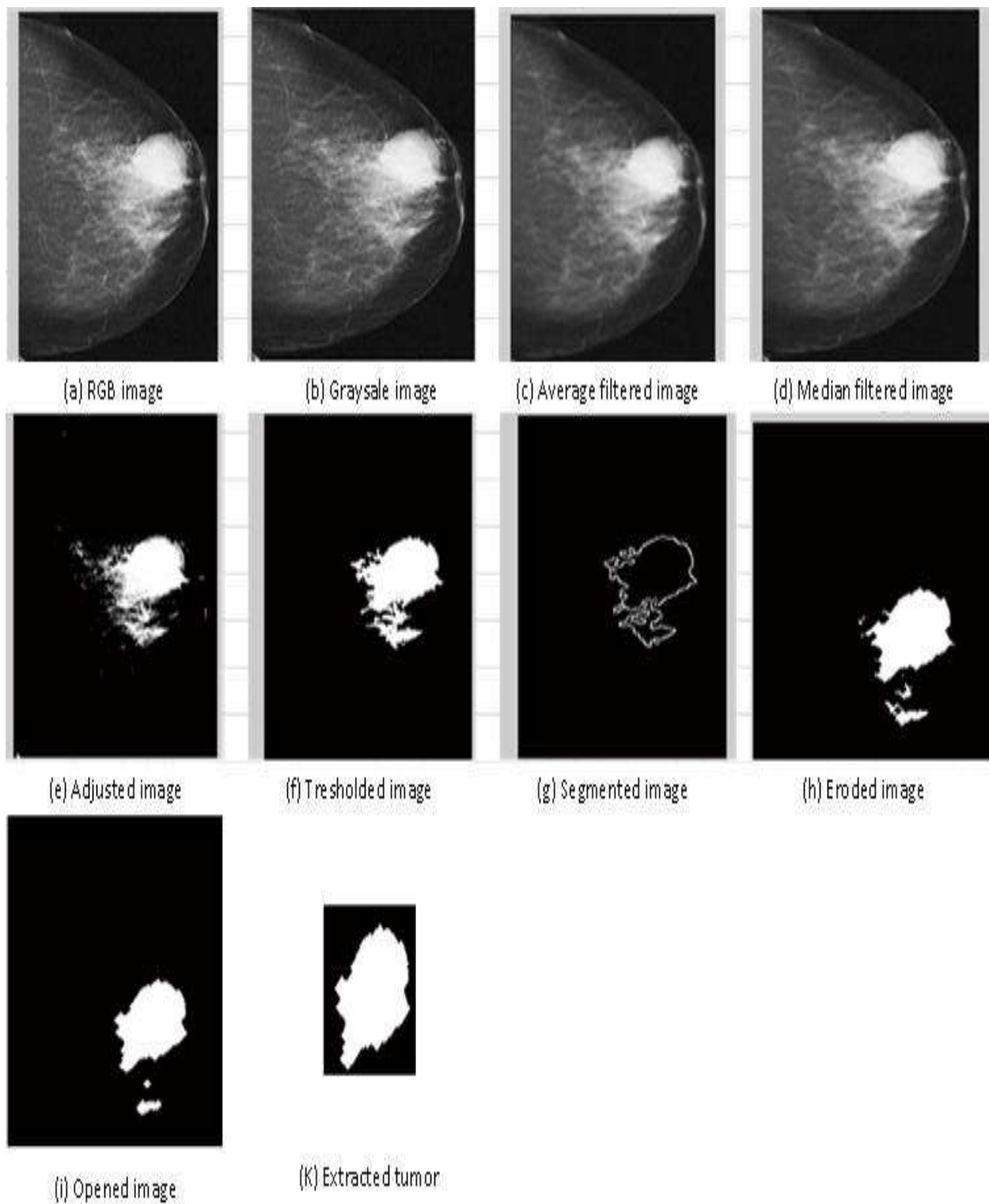
Figure 15: Breast cancer image undergoes all methods of the proposed system

(a) RGB image   (b) Grayscale image   (C) Average filtered image   (d) Median filtered image

(e) Adjusted image   (f) Tresholded image   (g) Segmented image   (h) Eroded image

(i) Opened image
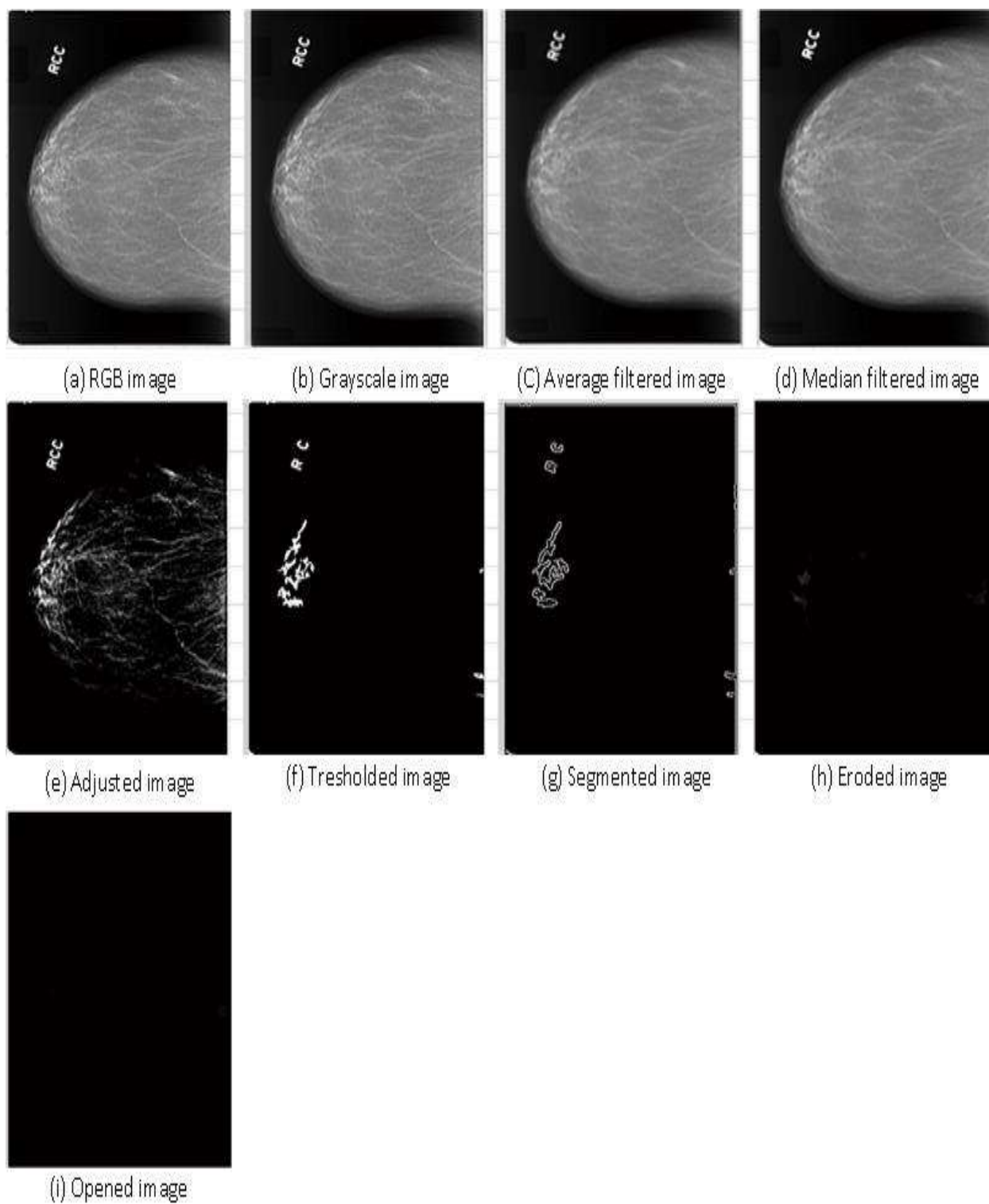
Figure 16: Normal breast image undergoes all methods of the proposed system

3.4 Grayscale Conversion

The first step is to convert the RGB image to grayscale. This conversion is done using the luminosity method which relies on the contribution of each color of the three RGB colors. Using this method, the grayscale image is brighter since the colors are weighted according to their contribution in the RGB image not averagely (Church et al., 2008).

(a Original image)   (b) Grayscale image
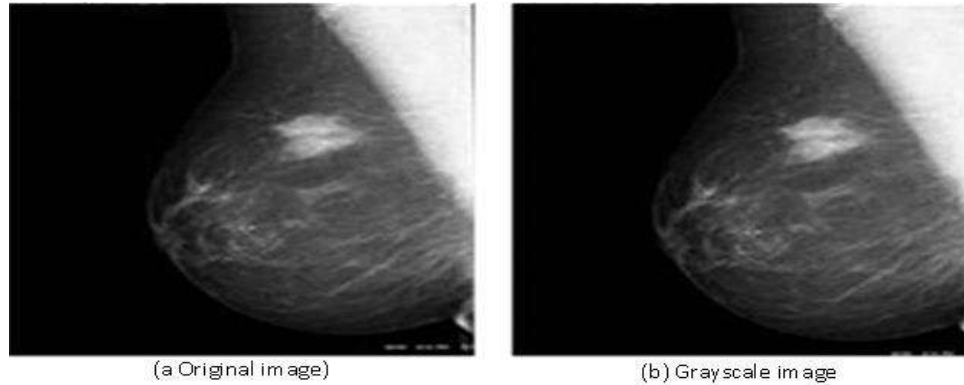
Figure 17: Grayscale conversion

3.5 Image Filtering

A smoothing filter is mainly used to reduce noise in an image. In a specific pixel, it takes into consideration the neighbor pixels to that studied pixel. The pixels are filtered out and the noise is reduced by taking the neighboring pixels into account. In our system we use two types of filters for smoothing images. One of the most useful filters used for smoothing is the median filter. This type of filter is used to reduce impulsive noise or the salt-and pepper in an image with preserving the useful features and image edges (Church et al., 2008). Median filtering is a linear process in which the output of the being processed pixel is found by calculating the median of a window of pixels that surrounds that studied pixel (Church et al., 2008).

Average filter or so called mean filter is another type of smoothing filters and it is used to reduce noise in images as well.The concept of such filtering method is simply to replace each pixel value in an image with the mean or "average" value of its neighbors, including itself. Thus, this helps in eliminating unrepresentative pixels of their surroundings. Mean filter can be thought as a

convolution filter; since it is based on a kernel,which represents the shape and size of the neighborhood to be sampled when calculating the mean (Gonzalez and Woods, 2002). In the proposed system we use a 10*10 kernel. Figures 17 and 18 show the effects of median and average filters on a cancerous breast image.

(a Original image)   (b) Median filtered image
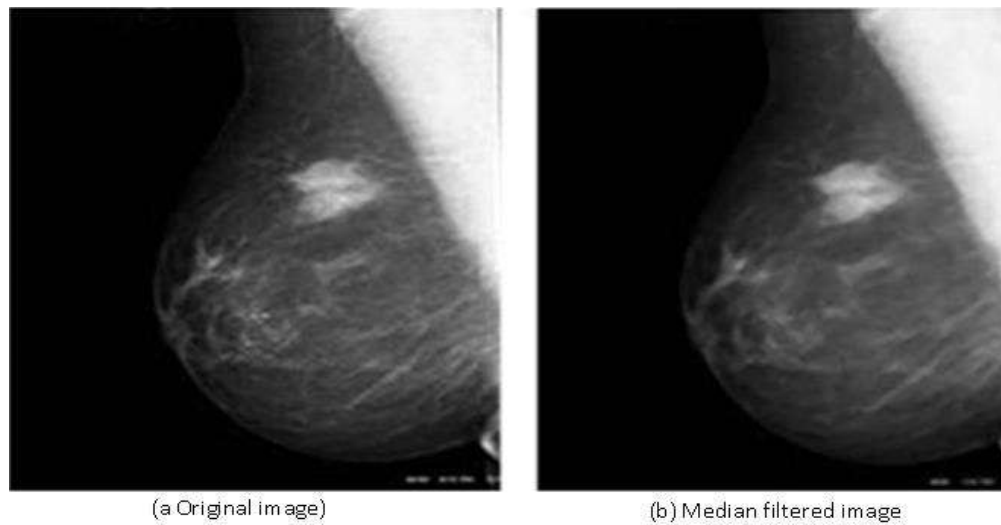
Figure 18: Image filtering

(a) Original image   (b) Average filtered image
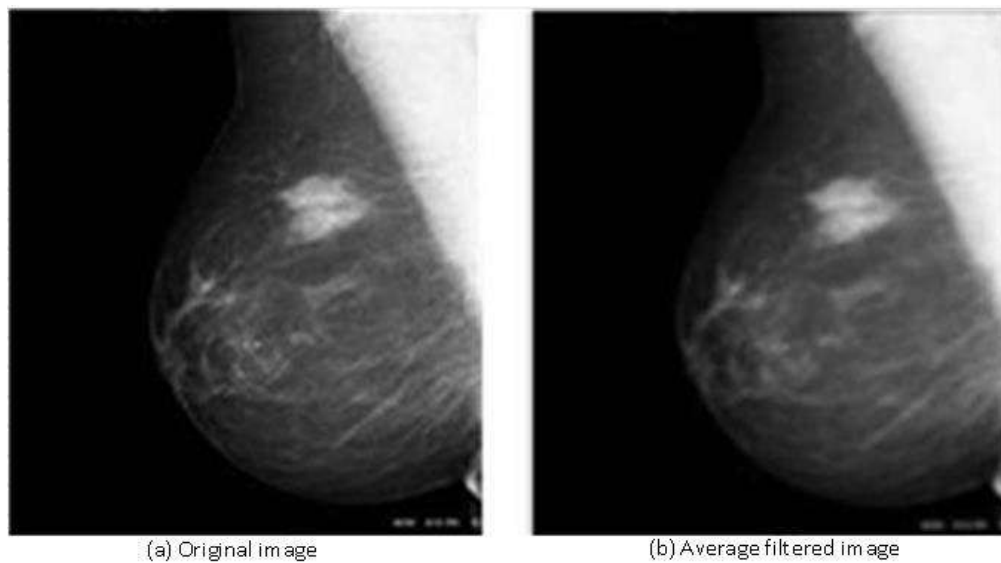
Figure 19: Image smoothing

3.6 Image Adjusting

The images then undergo intensity adjustment in which the input image's intensities are mapped to a new range of in the output image. This can be done by setting the low and high input intensity values that should be mapped and the scale over which they should be mapped (Figure 19) (Gonzalez and Woods, 2002).

(a) Median filtered image   (b) Adjusted image of a
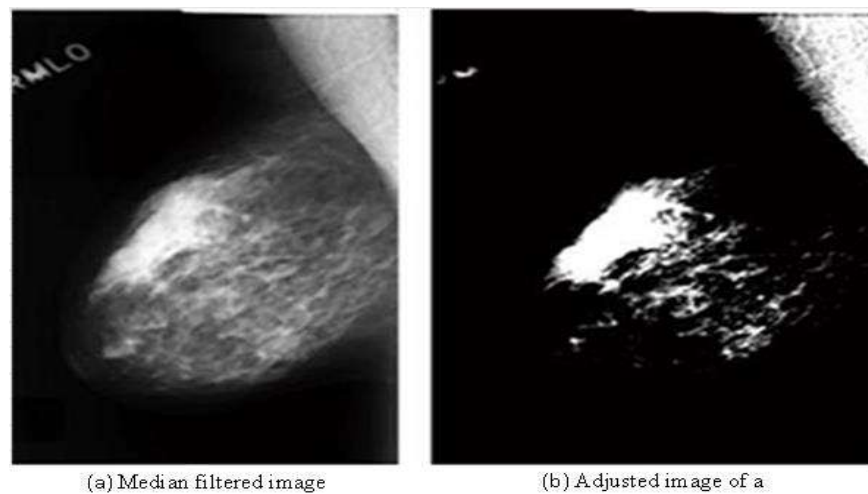
Figure 20: Image intensity adjustment

3.7 Thresholding

Thresholding is the separation of region of images into two regions. One region corresponds to the foreground region, in which it contains the objects that we are interested in. The other region is the background, corresponds to the unneeded objects. This provides segmentation of the image based on the image different intensities and intensity discontinuities in the foreground and background regions. The input of this method is usually a grayscale or color image, while the output is a binary image representing the segmentation. The black pixels refer to background and white pixels refer to foreground. The segmentation is achieved by a single parameter known as the intensity threshold. This is set by analyzing the histogram of the image which represents the intensity distributions of the image. During Thresholding, each pixel is compared to that threshold value. If the pixel value is greater than that threshold, then this pixel is considered as

foreground pixel (white). If the pixel value is lower than that threshold value, then the pixel is considered as background pixel (black)(Gonzalez and Woods, 2002).

Figure 20 illustrates a breast cancer image that undergoes Thresholding of 0.42 as threshold value.

(a) Original image   (b) Average filtered image
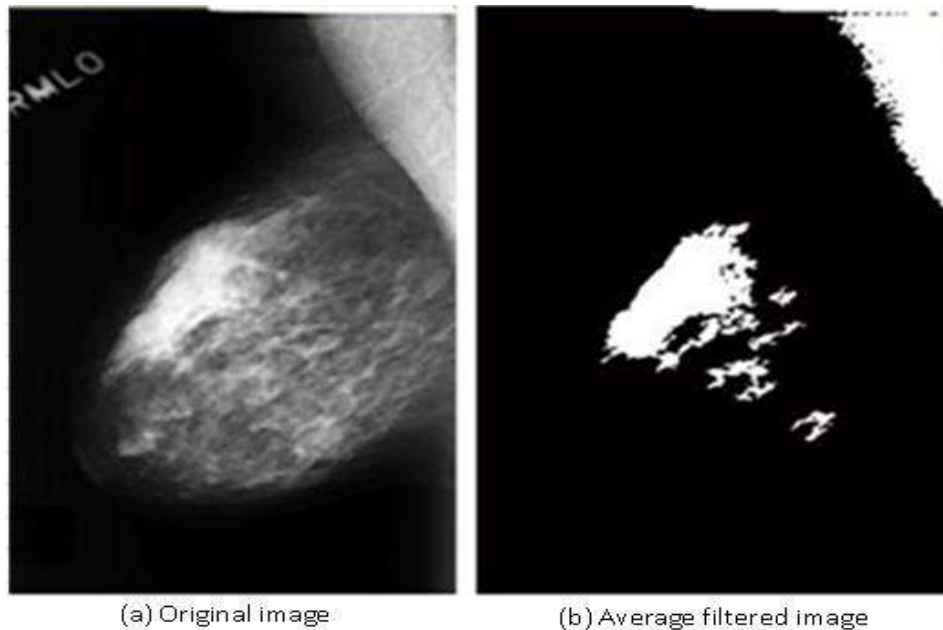
Figure 21: Thresholding

3.8 Segmentation Using Canny Edge Detection

Segmentation is the process of partitioning the image into different and many regions (Abiyev, 2013) This process can be done using different methods. The most common method for segmentation is the edge detection using the canny operator. The latter is an algorithm used for detecting a range of edges in an image. It detects the intensity discontinuities and finds boundaries of objects in an image by classifying pixels into edges. A pixel is classified as pixel if the gradient magnitude of it is greater than those of its both neighbors on the left and right sides (Figure 1.e) (Helwan, 2014).

3.9 Morphological Techniques

Morphology can be defined as a set of image processing operations that process images based on shapes. These operations can be done by applying a structuring element in an input image, resulting in an output image of the same size. The structure element is a matrix consists of 0's and 1's, where the 1's are called the neighbors. The value of each pixel in the output image is set according to a comparison of the corresponding pixel in the input image with its neighbors. Structure element has many shapes according to its application. Here, the "disk" structure element with a "radius" of 2 is used to extract the background. The most common morphological operations are erosion and dilation. The latter is used to respectively remove or add a pixel at object boundaries (Helwan, 2014).

(a) Original image   (b) Eroded image
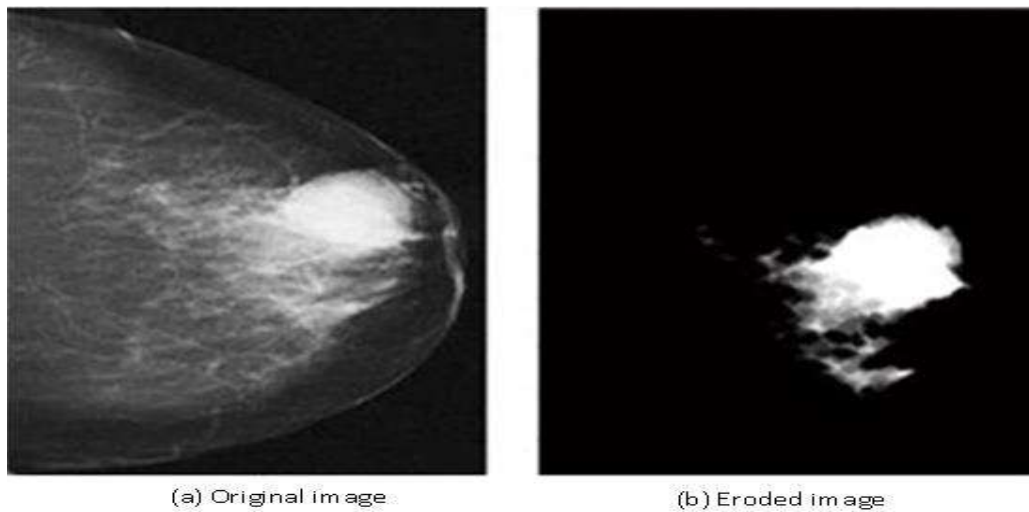
Figure 22: Image Erosion

3.10 Feature Extraction Using GLCM (Gray level co-occurrence matrix)

Feature extraction is a technique of extracting some visual content of images for indexing, retrieval, and classifying. It describes the relevant needed information in a pattern which facilitates the classification task by a certain procedure (Kumar and Bhatia, 2014).

The feature extraction technique basically aims to obtain the most relevant information from the original image and then represent that information in a lower dimensional space. Thus, when the input data to an algorithm is too large for the processing then it is suspected to be redundant, which means it has much data, but not much information, then the input data should be transformed into a reduced representation set of features which is called features vector that contain the extracted features (Mohanaiah et al., 2013).
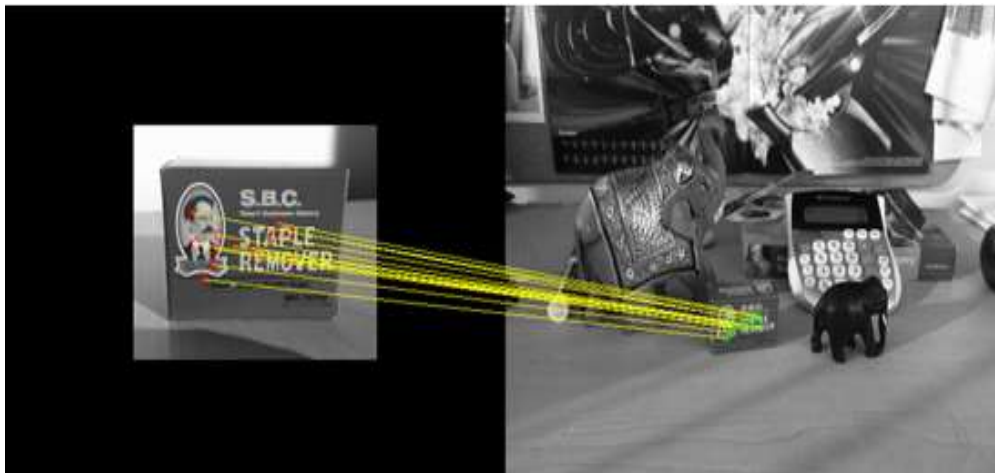
Figure 23: Features extraction example

One of the most common feature extractor is the Gray-Level Co-occurrence Matrix (GLCM). This method is such a technique of extracting second order statistical texture features from images or data.A GLCM can be considered as a matrix in which the number of rows and columns is equal to the number of gray levels, G, in the image. The matrix element P (i, j | Δx, Δy) is the relative frequency with which two pixels, separated by a pixel distance (Δx, Δy), occur within a given neighborhood, one with intensity 'i' and the other with intensity 'j'. The matrix element P (i, j | d, ɵ) contains the second order statistical probability values for changes between gray levels 'i' and 'j' at a particular displacement distance d and at a particular angle (ɵ). Using a large number of intensity levels G implies storing a lot of temporary data, i.e. a G × G matrix for each combination of (Δx, Δy) or (d, ɵ).

3.10.1 Extraction of texture and shape features of an image using GLCM

Different ways have been found for the purpose of the analysis of textures such as the Wavelet approach or even the Fourier approach. However, the simplest method for analysis is the one related to the way the human visual system perceives texture, which is the very first approach to texture analysis defined by Haralik (Gebejes and Huertas, 2013).

Haralick defines fourteen textural features measured from the probability matrix in order to extract the characteristics or features of texture statistics of remote sensing images. In this designed system we use GLCM to extract seven important texture features which are: Energy, Contrast, Homogeneity, Correlation, Entropy,and standard deviation, mean.

Table 1: Extracted shape and texture features

Features	Feature number
Roundness	1
Uniformity	2
Asymmetry	3
Compactness	4
Entropy	5
Standard deviation	6
Mean	7

The definitions below describe all the features used in our system for the classification operation and the meaning of each one in the actual texture analysis case is explained.

- Roundness: it is the gray level variation in a gray level co-occurrence matrix (Subashini et al., 2010).

$$R = \frac{4A\pi}{P^2} \tag{1}$$

Where A is the area of the segmented region of interest and P is its perimeter. If the Roundness is greater than 0.90 then, the object is circular in shape.

- Uniformity: it is denoted as U and it is a texture measure that is based on the histogram of the segmented ROI (Clausi and Zhao, 2002).

$$U = \sum_{k=0}^{L-1} pr_k^2$$

(2)

Where Pr_k is the probability of the k-th grey level. Because the Pr_k has values in the range of (0 to 1) and their sum equals 1, U is maximum when the numbers of pixels in all grey levels are equal, resulting in all the gray levels to be equal probable and their distribution to be uniform, and decreases otherwise.

- Asymmetry: it is to assess or evaluate whether the intensity levels tend to the dark side or light around the mean (Silva and Menotti, 2012).

$$\sum_{i=0}^{L-1} (z_i - m)^3 P(z_i)$$

(3)

- Entropy: it can be defined as a disorder, where in the case of texture analysis is a measure of its spatial disorder (Subashini et al., 2010).

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

(4)

Where Pr_k is the probability of the k-th grey level, which can be calculated as Z_k, M×N, Z_k is the total number of pixels with the k-th grey level and L is the total number of the available grey levels in an ROI of size M×N.

- Shape or Compactness: Since the shape of the segmented ROI is one of the important features that distinguish the benign and malignant tumors, shape features are extracted from each ROI prior to classification (Clausi and Zhao, 2002).

$$C = \frac{p^2}{4\pi A}$$

(5)

Where P is the perimeter, A is the area of the segmented ROI in pixels. The 4π factor is added to the denominator such that the compactness of a complete circle is 1.

- Mean (average): it is the average intensity of the image. Concerning mammograms, the denser tissue is, the higher the average intensity (Clausi and Zhao, 2002).

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

(6)

Where p(i,j) represents the pixel value at point (i,j), in an ROI of size M×N.

- Standard deviation: it can be defined as a measure of the contrast intensity grows, according to the irregularity of the texture (Silva and Menotti, 2012).

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

(7)

Table 2 shows some values of the extracted seven features of some processed cancerous images before normalization.

Table 2: Features values of some cancerous images before normalization

Features	Image #1	Image #2	Image #3	Image #4	Image #5
1	0.23888	0.261011	0.34932	0.801204	0.291032
2	0.828	0.85138	0.8270	0.8367	0.8665
3	0.79479	0.47971	0.8091	0.86087	0.83596
4	0.98473	0.93786	0.8521	0.78917	0.98921
5	0.3465	0.3821	0.4530	0.46461	0.47120
6	0.43665	0.3952	0.2830	0.23461	0.41120
7	6.162	8.952	7.639	6.867	10.336

These normalized features data are then fed into the neural classifier which has the capability of classifying these features into three classes through its experience adopted during the training phase.
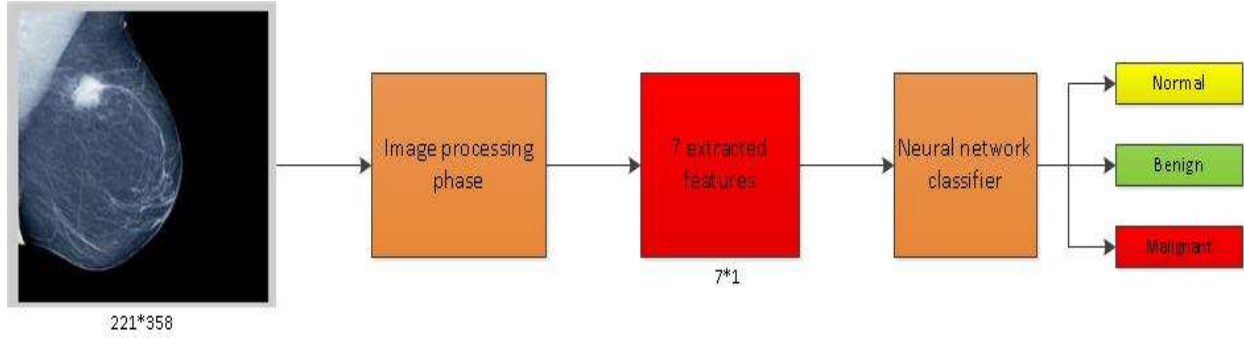
Figure 24: The different phases of the proposed system

3.11 Summary

In this chapter, we have discussed the second phase of the intelligent breast classification system, which is named the image processing phase. We presented algorithm of the system through a flowchart that lists the image processing techniques used in order to extract the patterns of interests (tumor). Moreover, we have explained each image processing technique in details with the help of figures and tables. In addition, we defined the shape and texture features of the images in the system and used their values for classification purposes.

CHAPTER FOUR

IBCIS: NEURAL NETWORK CLASSIFICATION PHASE

4.1 Overview

This chapter discusses the classification phase of IBCIS. The chapter explains the neural network concepts and mechanisms, and the backpropagation learning algorithm used in the designed system. We describe the training stage of the system through the update of the neural network parameters. We give an analysis of the results; describe the recognition rate, and learning curve. The last part of this chapter is the experimental results which show performance results of our system through tables and graphs.

4.2 Intelligent Breast Cancer Identification System: Classification phase

In this thesis, we propose a new approach for the classification of breast cancer using an intelligent system based on some extracted features using image processing techniques and a classifier. IBCIS comprises of two phases: image processing and classification phase. As discussed in the previous chapter, the images are processed using filters and some edge detectors, then some features are extracted from images in the first phase, in order to be then fed into the classifier which in our system, a backpropagation neural network.

4.2.1 Backpropagation neural network

An artificial neural network is a remodeling of the human brain information processing system. It is a multilayer system in which each layer is composed of multiple nodes which represent the neurons. Each node is connected to the others by means of edges through the weights which the information are transmitted (Abiyeh, 2012). ANN mainly consists of multilayers: input layer, hidden layers, and also output layer. The input from the previous layer is multiplied by the adjusted weights. At each node or neuron the weighted inputs are added and then the combined inputs pass through a non-linear transfer function in order to produce the desired output (Jain et al., 1996). ANN is basically developed to solve data mining applications. It is an adaptive

learning technique in which it has a different and specific learning methodology; the learning by examples. Therefore; some complex tasks can be handled using neural networks such as prediction, recognition, and classification (Rojas, 1996). Various learning and training algorithms can be used to train the network. One of the most public used algorithms is the backpropagation algorithm. In order to produce the desired output, the input weights should be adjusted and the correction-error should be reduced. The most popular used learning algorithm for updating the weights and correcting the learning error is the backpropagation algorithm. Backpropagation is a learning technique for the feedforward multilayer neural networks. It has two passes through the different layers; the forward pass and the backward pass. In the forward pass the weights are summed and then combined in the output layer. In the backward pass the weights are corrected. The actual output is subtracted from the desired one in order to produce the error. The error is then propagated back to all previous layers in order to update the weights and get the desired output (Al-Milli, 2013).
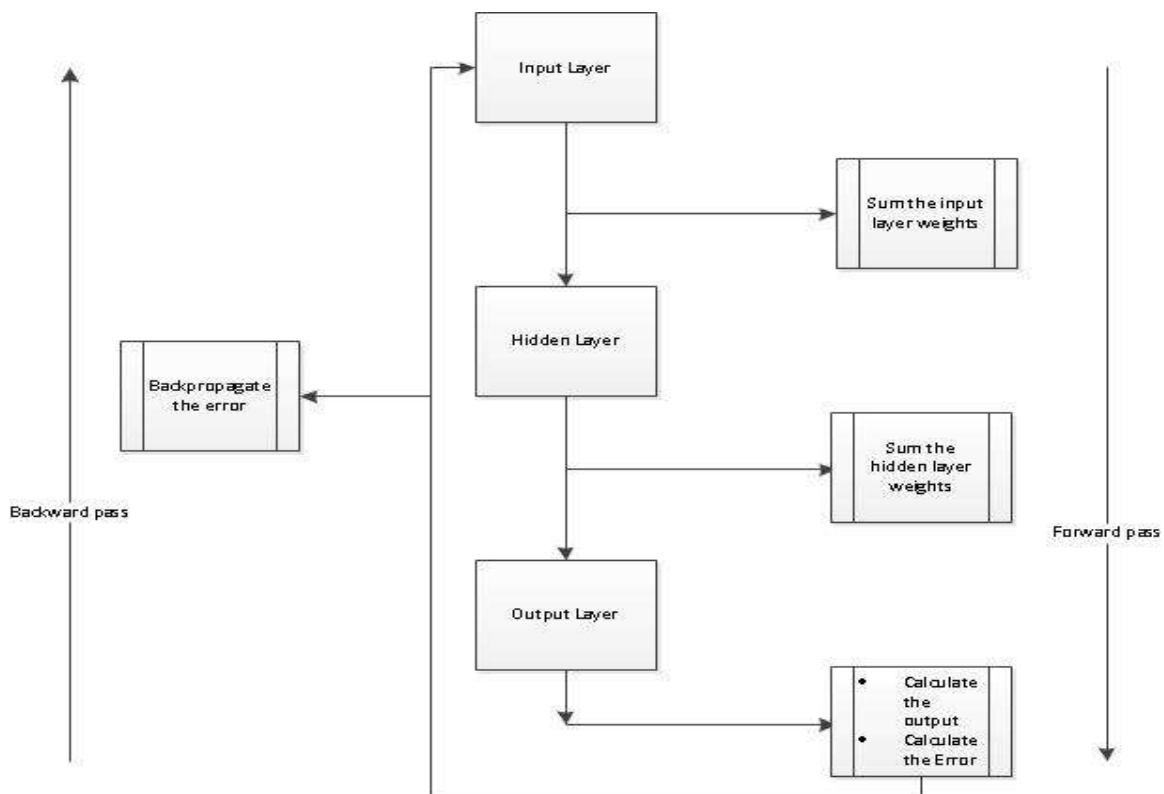
Figure 25: Backpropagation algorithm

34

4.2.2 System database

Since the proposed system is an image processing system, there should be available amounts of images in order to train and test the system. The images used for the system are obtained from an available public database called The Digital Database for Screening Mammography (DDSM) (Heath et al., 2013). The images obtained from this database are of size 221*358 true color images. Directly, our system converts the images to grayscale and reduces their sizes to 250*250 for fast processing purposes.

The figure below illustrates some images obtained from the DDSM database.

(a) Normal breast image   (b) Benign tumor   (c) Malignant cancer breast
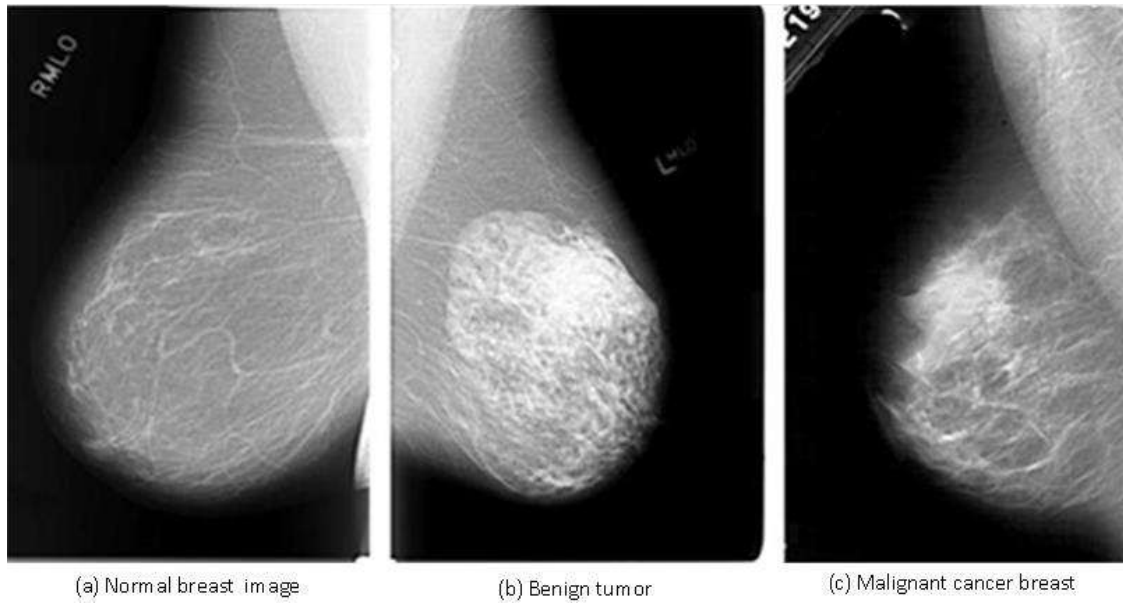
Figure 26: DDSM breast images

Table 3: Output classes and coding

Output Classes	Output coding
Normal (no tumor found)	1 0 0
Benign tumor	0 1 0
Malignant tumor (cancer)	0 0 1

4.2.3 ANN topology

The network was created on the Matlab software using the back propagation algorithm. The first step was to create a basic network and train it for simple operation such as 'AND" or 'OR' in order to reduce the mean sum error value to 0.001. the backpropagation learning algorithm was used with the adaptive learning rate and momentum rate for training network; with the function 'traingdx' and with the transfer function 'logsig'. The network was fed with the normalized datasets for the three sets and their output targets respectively. Figure 26 illustrates a multilayer neural network with 7 neurons in the inputlayer, 20 neurons in the hidden layer, and 3 neurons in the output layer. We ran the experiments for 5000 iterations.
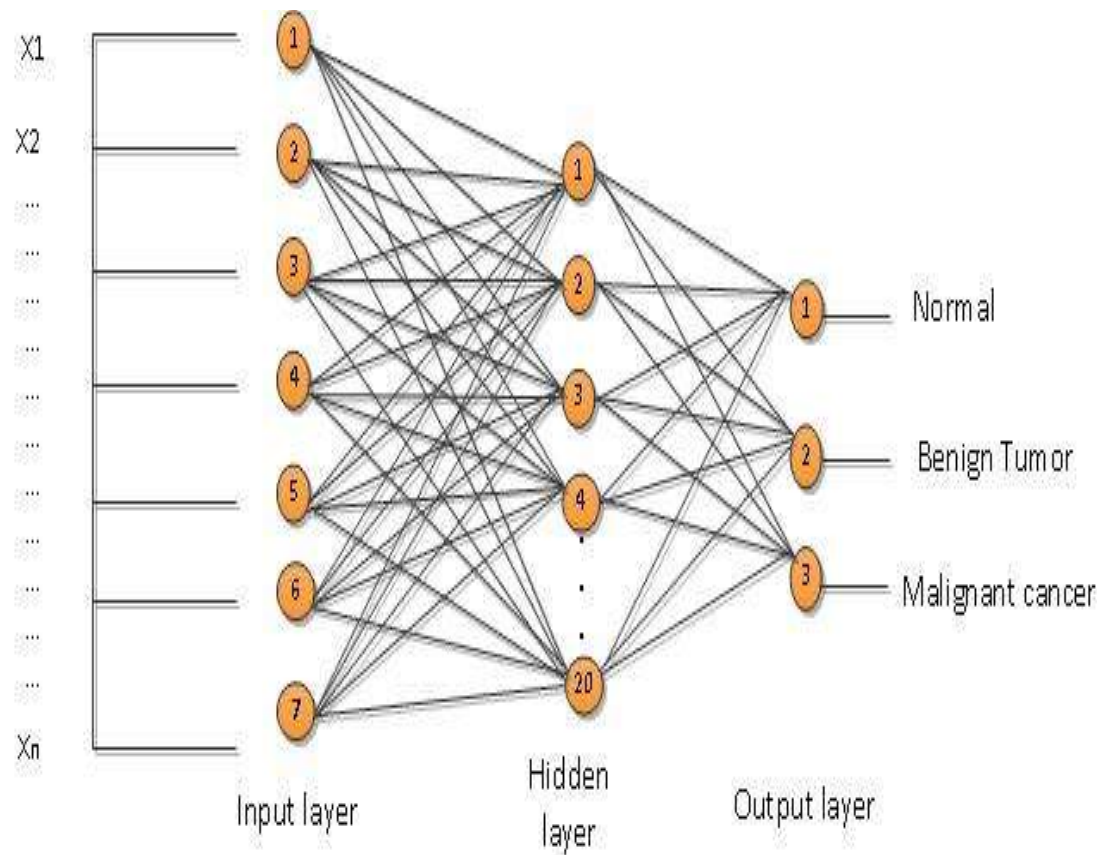
Figure 27: ANN architecture

X1…X_n represent inputs of the network. The connections between the neurons called the weights. Each neuron in the input layer is connected to the succeeded neurons in the hidden layer. Moreover, each neuron in the hidden layer is connected to the three output neurons. Sigmoid function is used as a transfer function for the network.

Table 4: ANN input parameters

Parameters	Value
Number of neurons in input layer	7
Number of neurons in output layer	3
Number of neurons in hidden layer	20
Iteration number	5000
Learning rate	0.001
Momentum rate	0.5
Error	0.001
Training time (sec)	300
Activation Function	Sigmoid

Table 5 shows all the parameters used when training the network. The network ran for 5000 iterations with a learning rate of 0.001, a momentum rate of 0.5 and a minimum error of 0.001 since it is a medical application.

4.2.4 The system training

To create an efficient network that can be capable of such classification task, it is best to train the network on simple tasks first. To do this, the network is first trained on ideal vectors until it has a low sum squared error. In order to decrease the error value to 0.001, we have to start training the neural network on simple and mathematical operations such as 'AND' or 'OR' operations.

The training is done using backpropagation learning algorithm with both adaptive learning rate and momentum, with the function *'traingd'*.After making sure that the error is minimized; we started feeding the neural network with the input images and their targets respectively.

The network was trained on 300 breast images: 100 images normal, 100 images are benign tumor, and 100 are malignant tumor (cancer). The table 5 represents the training set of images which consists of 3 types of images: normal, benign, and malignant. It also shows the total number of database breast images used for training and testing phase.

Figure 27 illustrates the learning curve of the developed system that is based on a backpropoagation neural network. The network was trained on a large number of images obtained from the DDSM database for mammography breast images. Thus, the result is perfect and the network was well trained since the mean square error was diminishing as much as the number of iterations was increasing.

Table 5: Training and testing number of images

	Normal images	Benign tumor images	Malignant tumor images	Total number of images
Training	100	100	100	300
Testing	100	100	100	300
Total	200	200	200	600
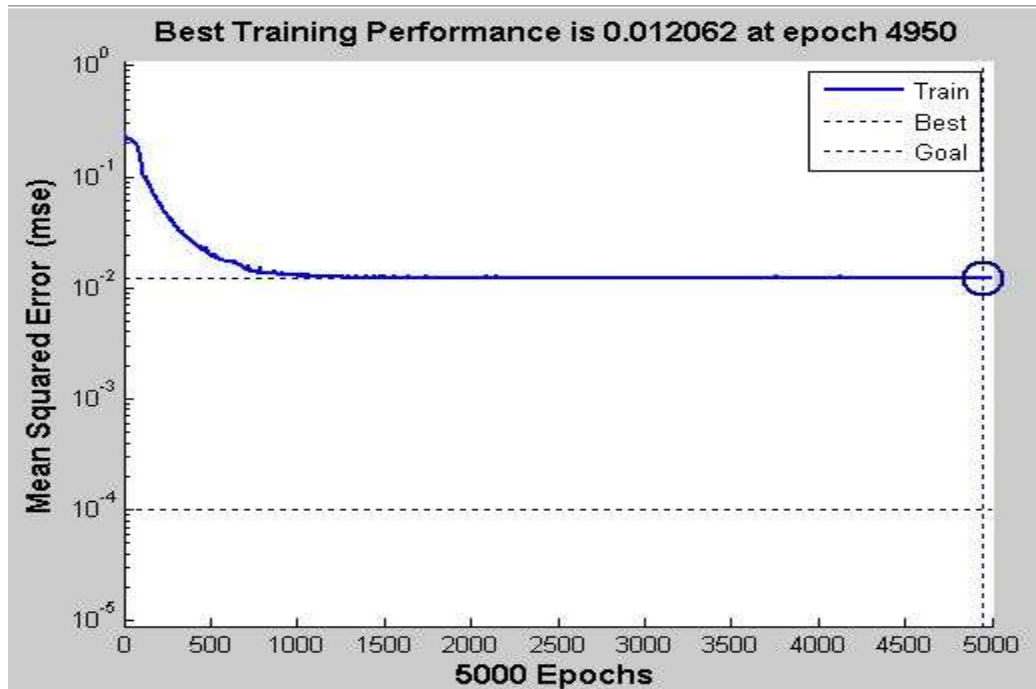
Figure 28: Learning curve of the training phase

This figure below represents the regression plot of the desired output (dotted line) and the actual output. As the actual output is far from the target as the error is increased. In this figure, it is remarked that the target and the actual output are overlapped which means that the error is minimized and the network well trained (training ratio = 100 %).
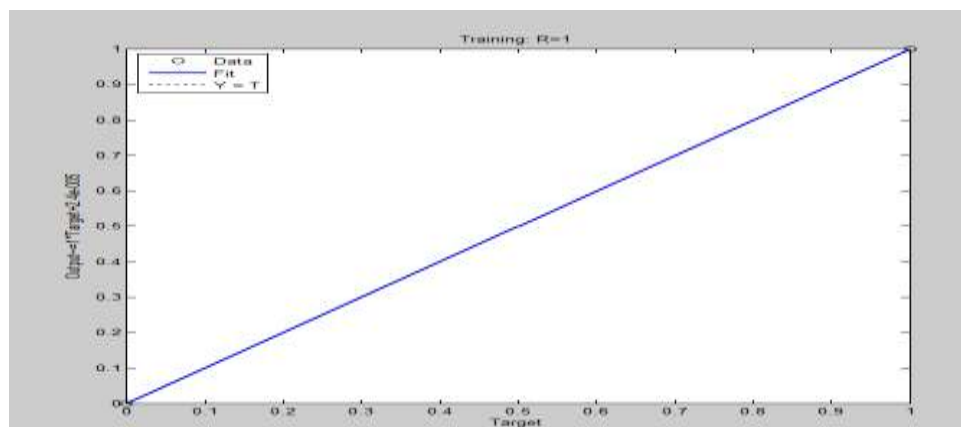

Figure 29: Actual versus target output

39

Table 6: Intervals of ROI extracted feature

Extracted features	Benign	Malignant
Roundness	0.85-0.99	0.25-0.84
Uniformity	0.98-1	0.81-0.89
Asymmetry	0.90-0.99	0.2-0.89
Compactness	0.2-0.7	0.72-1
Entropy	0.0503-0.304	0.347-0.593
Standard deviation	0.040565-0.238626	0.232185-0.439165
Mean	0.46-0.54	0.56-9

4.2.5 Experimental results and system performance

The proposed breast cancer identification system was implemented using 2.7 GHz PC with 4 GB of RAM, Windows 7 OS and Matlab 2013a software tools. The network was tested on a dataset of 300 images; 100 for normal, 100 for a benign tumor, and 100 for malignant tumor images. Table 6 represents the results obtained from two runs for the three different classes (Normal, Benign, malignant). This table below represents the number of images that were accurately recognized by the network in the training and the testing phase. It also shows the percentage of images that were not recognized during the testing phase. The images were obtained from DDSM public database of size 221*358 pixels. They were processed and then rescaled to 250*250 pixels for the purpose of fast computing. 7 features were then extracted and fed into the neural network as inputs. Figure 30 illustrates the process of feeding an image to the neural network including the processing phase.

The number of recognized images was divided by the total number of images with respect to each case set (Normal, benign, and malignant tumor). The result of this fraction is called the recognition rate, which is the efficiency of the neural network in identifying the breast cancer.

The experimental results of the intelligent breast cancer identification system were as follows: 100% using the training image set (300 images, 100 for each class). The overall identification rate was eventually calculated and the result is approximately 97% correct identification rate. Table 6 shows the intelligent breast cancer identification results in details.

Table 7: Breast cancer identification results

Breast image type	Image sets	Number of images	Identification rate	
Normal	Training set	100	100/100	100%
	Testing set	100	100/100	100%
Benign tumor	Training set	100	100/100	100%
	Testing set	100	96/100	96%
Malignant tumor	Training set	100	100/100	100%
	Testing set	100	94/100	94%
Total identification rate	All sets	600	597/600	97%

Table 8: Different identification rate for different input parameters

Learning rate	Momentum rate	Nb. Of hidden neurons	Epochs	Identification rate
0.04	0.36	5	10000	81 %
0.05	0.4	10	10000	85 %
0.02	0.7	15	10000	93 %
0.001	0.5	20	5000	97 %

4.3 Summary

The classification phase of the developed system was discussed in this chapter. The neural classifier and the backbropagation algoritm were also explained in details. Moreover, the results of training the designed system were presented through the learning curves and the tables that illustrate the identification rate obtained when testing the system.

CHAPTER FIVE
RESULTS AND DISCUSSION

5.1 Overview

This chapter provides simulation results of IBCIS. A comparison of the intelligent identification system and other related works that use the same database and similar image enhancement techniques for the identification of breast cancer is presented. The comparison is based on identification rate of the used system.

5.2 Results Discussion

In this thesis, an intelligent breast cancer identification system based image processing and neural network has been developed. The system aim is to classify the breast cancer mammograms into three types: normal, benign, and malignant tumor. It consists of two main phases: image processing and neural networks. In the first phase, the images were processed in order to extract the region of interest; then some features describing the tumor were extracted using pattern averaging. The extracted features were used as inputs for a neural classifier based on a backpropagation learning algorithm. After learning and getting good training recognition rate, the network was tested on different images and the identification rate was promising (97 %).
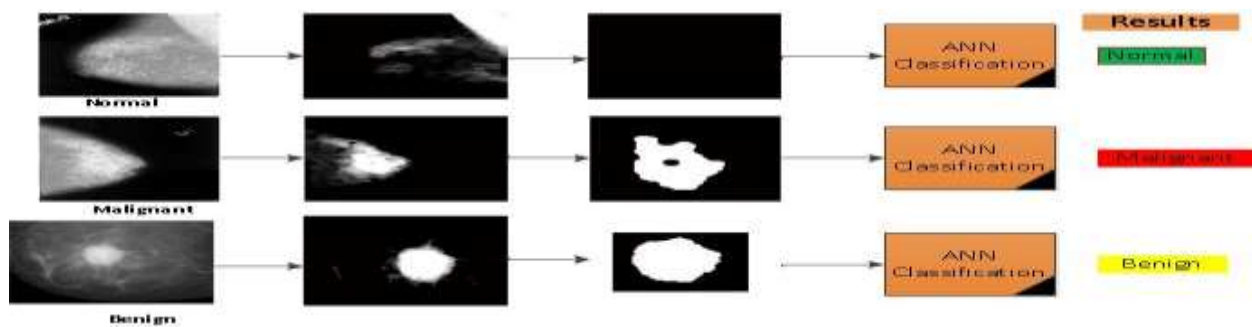

Figure 30: Some obtained accurate results of the system

5.3 Results of Comparison With Some Related Works

A system of early detection of breast cancer using the SVM classifier technique was developed by Anna Rejani and S.ThamaraiSelvi (Rejani and Selvi, 2009). The project is to develop a classification system able of classifying the breast cancer into benign and malignant using SVM classifier. The authors used different image processing techniques for reading and extracting the gray level information by image enhancement and for segmenting the breast tumor into the image. The images used were obtained from the same database DDSM. Then, the tumor region was extracted, and some morphological features are extracted to classify the breast tumor using SVM. The classification rate obtained by using this algorithm was 88.75%, which is lower than ours 97.1%.

Other work that has been done by Atef Boujelbenet and some authors (Boujelbenet et. al., 2012) is the classification of breast tumors into benign and tumors. This work is slightly similar to our research, since it uses both texture and shape features for the classification process. In this work, the authors used two methods to classify; multilayer perceptron and K-Nearest Neighbor. Since our classifier is the neural network, the comparison is between the accuracy obtained when they used the MLP as a classifier. The authors introduced a comparison of the classification rate obtained when extracting both the shape features and texture features. The results in terms of sensitivity and specificity tend to reach 90%. The results of that work have been tested using two algorithms of classification: KNN (K=7) and MLP. Using MLP and sharp features, the results were 89%. While using the texture features and MLP, they got also 89%.

Nasser basher and Mustafa Mohammed (Basher and Mohammed, 2013) developed a CAD system for the classification of breast cancer into benign and malignant. In their paper, a Computer Aided Diagnosis (CADx) system was implemented using the MATLAB for the classification of abnormal masses in digital mammograms using Support Vector Machines (SVM). The images used in their system were taken from the same database that we used DDSM. The developed identification system successfully achieved 93% as classification accuracy, which is considered as a lower result when compared with the developed work which is in the same research field 97%.

Table 9: Results comparison

Project title	Authors	Classifier used	Extracted fetaures	Classification accuracy
Early detection of breast cancer using SVM classifier technique	Y.Ireaneus Anna Rejani and S.Thamarai Selvi	SVM	Texture features	88.75%
Automatic Diagnosis of Breast Tissue	Atef Boujelben et al.	Neural network	Shape features	89%
Classification of Breast Masses in Digital Mammograms Using Support Vector Machines	Nasser basher and Mustafa H. Mohammed	SVM	Texture and shape features	93%
Proposed work accuracy		BPNN	Texture and shape features	97%

Figure 30 shows the structure of the designed image classification system. Here the feature extraction block extracts important features that describe the breast tumor in order then to be classified correctly. Some of the authors uses shape features in their systems only, while others use texture features only, or both of them. In the designed system, we extracted shape and texture features since they both describe the tumor. We used a neural network as a classifier, since it is more accurate in classification and figure 32 proofs that. We compared the neural network based classification system with the system based on SVM. The comparison shows that the classification accuracy of the designed system is better than the accuracy of other models. In addition, the above figure shows that using both shape and texture features results in higher classification accuracy.

5.4 Summary

In this chapter, a comparison with the designed system with other similar works was presented. All the compared systems used the same images database, however, the comparison was upon

the type of classifier used and also the types of extracted features. The results of the comparison shows that our intelligent system is more efficient than other similar intelligent systems.

CHAPTER SIX
CONCLUSION

In this thesis, an intelligent breast cancer identification system based on image processing and neural network has been developed. The system aimed to classify the breast mammography images into 3 classes: normal, benign tumor, and malignant tumor. The first phase of the designed system is the preprocessing of the mammograms using different image processing techniques in order to extract the clear region of interest which is the tumor. The techniques used are: image resizing, grayscale conversion, filtering, thresholding, segmentation, edge detection, tumor extraction, and morphological techniques. After processing, some features are extracted from the region of interest of the image. These features represent the shape and texture features of the images. Both types of features are important for classifying malignancy of the breast.

Classification is the next phase of the breast cancer identification system. The extracted features such as the asymmetry, roundness, compactness, uniformity, mean, and standard deviation are fed into the classifier. A backpropoagation algorithm is used for learning neural network based IBCIS. 600 images obtained from the public database DDSM. These images are used for the simulation of the IBCIS. Among them, 300 images representing the three classes are used for learning and minimizing the mean square error. The network was tested with other 300 images. The overall identification rate was obtained 97 %, which is high and sufficient for such medical application. Therefore, according to the obtained results, we can say that this intelligent and accurate identification system can be furthermore improved and then developed into real life application.

REFERENCES

Abiyev, R. (2012). Facial Feature Extraction Techniques for Face Recognition. *Journal of Computer Science, 10*, 2360–2365.

American Cancer Society. (2008). Cancer Facts and Figures. Retrieved December 20, 2014, from www.cancer.org/downloads/STT/2008CAFFfinalsecured.pdf.

American Cancer Society. (2014). Breast Cancer. Retrieved November 20, 2014, from http://www.cancer.org/acs/groups/cid/documents/webcontent/003090-pdf.pdf.

Anil K., Mao. J., and Mohiuddi, K. (1996). Artificial Neural Networks: A Tutorial. IEEE Computers.

Anthony, A., Michelle, L., Lonnemann, E., and Ritchie, L. (2013). Breast Cancer. Retrieved November 10, 2014, from http://www.physio-pedia.com/Breast_Cancer.

Basher, N., and Mohammed, M. (2013). Classification of Breast Masses in Digital Mammograms Using Support Vector Machines. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*, 200–210.

Boujelben, A., Tmar, H., Abid, Mohamed., and Mnif, J. (2012). Automatic Diagnosis of Breast Tissue. *Advances in Cancer Management*, 258–2270, doi: 10.5772/22565.

Church, J., Chen, Y., and Rice, S. (2008). A Spatial Median Filter for Noise Removal in Digital Images. *Southeastcon IEEE*, *30*, 618–623.

Clausi, D., and Zhao, Y. (2002). Rapid Co-occurrence Texture Feature Extraction Using a Hybrid Data Structure, *Computers and Geosciences*, *28*, 763–774.

Gebejes, A., and Huertas, R. (2013). Texture Characterization Based on Grey-Level Co-occurrence Matrix. *In Proceedings of the 3rd Conference of Informatics and Management Sciences,* (pp. 25–29). USA: New York University.

Gonzalez. R., and Woods, E. (2002). Digital Image Processing, Second Edition. Prentice-Hall.

Heath, M., Bowyer, K., Kopans, D., Moore, M., and Kegelmeyer, W. (2001). The Digital Database for Screening Mammography. *In Proceedings of the Fifth International Workshop on Digital Mammography* (pp. 212–218). San Diego: Convention Centre.

Helwan, A. (2014). ITDS: Iris Tumor Detection System Using Image Processing Techniques. *International Journal of Science and Engineering Reseach*, *3*, 45–50.

Ireaneus, Y., and Rejani, A. (2009). Breast Cancer Detection Using Multilevel Thresholding. *International Journal of Computer Science and Information Security, 6*, 35–40.

Jatoi, J., and Kaufman, M. (2010). Management of Breast Diseases. Springer-Verlag Heidelberg.

Kumar, G., and Bhatia, P. (2014). A Detailed Review of Feature Extraction in Image Processing Systems. *In Proceedins of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies* (pp.356–370). Spain: University of Cordoba.

Medscape. (2014)**.** Lumpectomy. Retrieved December 22, 2014, from http://emedicine.medscape.com/article/1830086-overview.

Mohanaiah, P., Sathyanarayana, P., and GuruKumar, L. (2013). Image Texture Feature Extraction Using GLCM Approach. *International Journal of Scientific and Research Publications*, *3*, 52–60.

National Breast Cancer Foundation. (2012). Causes of Breast Cancer. Retrieved july 25, 2014, from http://www.nationalbreastcancer.org/causes-of-breast-cancer

Padayachee, J., Alport, M., and Rae, W. (2007). Identification of The Breast Edge Using Areas Enclosed by Iso-intensity Contours. *Computerized Medical Imaging and Graphics*, *31*, 390–400.

Paramkusham, S., Rao, K., and Prabhakar, B. (2013). Automated Approach for Qualitative Assessment of Breast Density and Lesion Feature Extraction for Early Detection of Breast Cancer. *International Journal of Computing Science and Communication Technologies, 6*, 200–207.

Rederic, N., Faizon, F., and Sun, Y. (2000). Computerized Analysis of Asymmetry in Digitized Mammograms. *In Proceedings of the IEEE 26th Annual Northeast Bioengineering Conference,* (pp. 45–50). Canada: University of Waterloo.

Rojas, R. (1996). Neural Networks: A Systematic Introduction, Springer-Verlag.Davis.

Silva, W., and Menotti, D. (2012). Classification of Mammograms by the Breast Composition. *In International Conference on Image Processing, Computer Vision, and Pattern Recognition* (pp. 369–380). Lebanon: Lebanese University.

Society of North America. (2013). Mammography. Retrieved December 22, 2014, from http://www.radiologyinfo.org/en/info.cfm?pg=mammo.

Soh, L., and Tsatsoulis, C. (1999). Texture Analysis of SAR Sea Ice Imagery Using Grey Level Co-Occurrence Matrices. *IEEE Transactions on Geoscience and Remote Sensing*, *37*, 25–38.

Subashini, T., Ramalingam, V., and Palanivel, S. (2010). Automated Assessment of Breast Tissue Density in Digital Mammograms. *Computer Vision and Image Understanding*, *114*, 33–43.

Susangkoman, K. (2014). About Breast Cancer. Retrieved December 28, 2014, from http://ww5.komen.org/BreastCancer/Mammography.html#Findings-on-a-mammogram.

World Cancer Factsheet Form. (2012). Retrieved December 10, 2014, from http://www.cansa.org.za/files/2009/10/Global-factsheet-2012.pdfwho.

APPENDIX
SOURCE CODE

- **Image processing**

```
Close all
clear;
clc;
N=300;

saveData=zeros(7,300);

for k = 1:N
myFolder = 'D:\N.breast';
if ~isdir(myFolder)
errorMessage = sprintf('Error: The following folder does not exist:\n%s',
myFolder);
uiwait(warndlg(errorMessage));
return;
end
filePattern = fullfile(myFolder, '*.gif');
jpegFiles = dir(filePattern);
baseFileName = jpegFiles(k).name;
fullFileName = fullfile(myFolder, baseFileName)
fprintf(1, 'Now reading %s\n', fullFileName);
[I xmap]  = imread('bebreast25.gif');
[I xmap]  = imread(fullFileName);
    f = ind2gray(I,xmap);

imshow(f);  % Display image.
drawnow; % Force display to update immediately.

% Filteringgguding different filters******************
%%%% average filter*************************
i= fspecial('average',[3 3]);
a=imfilter(f,i);
figure, imshow(a),  title('average filtered image');
%pause

%********************************************************
%%%% Median filter*************************

m=medfilt2(a);
figure, imshow(m),  title('median filtered image'); %%% using median
filtering
%pause
%%%% adjust filter************************

im=imadjust(f,[0 1], [1 0]);        %% image adjustement, gamma =1, by
default....get the negative image
figure, imshow(im),  title('adjusted image');
```

```matlab
%pause

g2=imadjust(f,[0.5 0.75], [1 1]);    %% image adjustement, gamma =1, by
default..convert the intensities btw 0.5 and 0.75 to values btw 0 and 1.
% g3=im2uint8(mat2gray(log(1+double(f))));    %%%%logarithmic intensity
transformations
figure, imshow(g2), title('adjusted image #2');
%figure, imhist(f)
%pause
%%%%%%%%%%%%%%%%%%%%%%%%%%%%



%% treshold the image
level = graythresh(g2);
bw = im2bw(g2,level);
bw = bwareaopen(bw, 50);
figure, imshow(bw),title('treshold image');
%pause
% Canny
S5 = edge(bw, 'canny',0.3);
imshow(S5);title('Canny Edge Dectection [auto]');
size(S5);
%pause
ds = bwareaopen(S5,40);                          %# Remove small edge objects
imshow(ds); title('ds');                         %# Plot the remaining edges
%pause
%%%%Morphology technique, image erosion to erase the unwanted components
se = strel('disk',5);
I4 = imerode(g2,se);
I5 = imopen(I4,se);
imshow(I4),title('eroded image [auto]');
%pause

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%ROI exxtraction
props = regionprops(I4);
[~,ind] = max([props.Area]);
imshow(I4 == ind);
pause
% %%%%%%%%%%%%Features Detection....GLCM
[glcm] = graycomatrix(I4);
%stats = graycoprops(glcm,{'contrast','homogeneity','correlation','Energy'});
%%%%Entropy
E = entropy(glcm);
%%%%%%Standard Deviation
std = std2(glcm);
%%%%%%%Mean
mean = mean2(glcm);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
props = regionprops(I4, 'Area', 'Perimeter');
  areas = [props.Area];
  perims = [props.Perimeter];
  Assym = [props.sym]
  % computer circularities, and plot against areas
  circularities = 4 * pi * areas ./ perims .^ 2; % Dongsheng's formula
  plot(areas, circularities, 'ro')
```

```
%%%
Roundness = 4 * pi * areas ./ perims .^ 2; % Dongsheng's formula
compactness = (Perims .^ 2) ./ (4 * pi * area);

%tf=ischar(stats)
class(stats);
a=[ E; std ;mean];
b=[ Roundness  compactness Assym];
c = num2cell(b);
Ninput=[a;c];
Format shortG
Ninput=cell2mat(Ninput)
Ninput
Close all


saveData(:,k)=Ninput;
%     Ninput(:,k)=Ninput;


end
xlswrite('correctdata.xlsx', saveData,'sheet1');

closeall
```

- ## **Normalizing**

```
%M=normalizing(a,c)

X=xlsread('dataset.xlsx', 'sheet1');
nr = meshgrid(max(X),[1:size(X,1)]');
normdata=X./nr;
xlswrite('af.norm.xlsx', normdata,'sheet3');
%%%%%%%%%%%%%%%%%%%% to excel file
filename = 'trainset.xlsx';
writetable(M,filename,'Sheet',1,'Range','A');
close all
```

- ## **Neural Network**

```
clearall
closeall
clc

PATTERNS = [];
dataset = xlsread('dataset2.xlsx','sheet4');          %database%%% inputs

    PATTERNS = [dataset];



dis.out=xlsread('output.xlsx','sheet2');   %...outputs
% CREATING AND INITIATING THE NETWORK
net1=newff(PATTERNS,dis.out,20,{'logsig','logsig'},'traingdx');
net1 = init(net1);
net1.LW{2,1} = net1.LW{2,1}*0.01;
%net.b{2} = net.b{2}*0.01;
```

```matlab
% TRAINING THE NETWORK
net1.trainParam.goal = 0.01; % Sum-squared error goal.
net1.trainParam.lr = 0.005;  % Learning Rate.
%%0.007....................96.8%%
net1.trainParam.alpha = 0.27;                            %% 0.27
net.trainParam.show = 50; % Frequency of progress displays (in epochs).
net1.trainParam.epochs =10000;% Maximum number of epochs to train.

[net1,tr] = train(net1,PATTERNS,dis.out);     % Normal, Benign, Malignant

actout.normal=sim(net1,PATTERNS);
actout.normal
```