FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

DEVELOPMENT OF GRAPHICAL CODE-BASED ALGORITHMS FOR THE DETECTION OF ABNORMALITIES IN MAMMOGRAM IMAGES

Iman HAMADAMIN

Master Thesis

DEPARTMENT OF ELECTRIC AND ELECTRONICS ENGINEERING

AUGUST 2021

FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

Department of Electric and Electronics Engineering

Master Thesis

DEVELOPMENT OF GRAPHICAL CODE-BASED ALGORITHMS FOR THE DETECTION OF ABNORMALITIES IN MAMMOGRAM IMAGES

Author
Iman HAMADAMIN

Supervisor
Assoc. Prof. Dr. Hasan GÜLER

AUGUST 2021
ELAZIG

FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

Department of Electric and Electronics Engineering

Master Thesis

Title: Development of Graphical Code-Based Algorithms for the Detection of Abnormalities in Mammogram Images

Author: Iman HAMADAMIN

Submission Date: 07 July 2021
Defense Date: 10 August 2021

THESIS APPROVAL

This thesis, which was prepared according to the thesis writing rules of the Graduate School of Natural and Applied Sciences, Fırat University, was evaluated by the committee members who have signed the following signatures and was unanimously approved after the defense exam made open to the academic audience.

Signature

Supervisor:	Doç.Dr.Hasan GÜLER Firat University, Faculty of Engineering	Approved
Chair:	Prof.Dr.İbrahim TÜRKOĞLU Firat University, Faculty of Technology	Approved
Member:	Dr.Öğr.Üyesi Muhammet Emin ŞAHIN Yozgat Bozok University, Faculty of Engineering – Architecture	Approved

This thesis was approved by the Administrative Board of the Graduate School on

Signature

Assoc.Prof.Dr.Kürşat Esat ALYAMAÇ
Director of the Graduate School

DECLARATION

I hereby declare that I wrote this Master Thesis titled “Development of Graphical Code-Based Algorithms for the Detection of Abnormalities in Mammogram Images” in consistent with the thesis writing guide of the Graduate School of Natural and Applied Sciences, Firat University. I also declare that all information in it is correct, that I acted according to scientific ethics in producing and presenting the findings, cited all the references I used, express all institutions or organizations or persons who supported the thesis financially. I have never used the data and information I provide here in order to get a degree in any way.

10 August 2021

Iman HAMADAMIN

PREFACE

The breast cancer is the most common form of cancer in female population all over the world. However it appears in both genders but more percentage in women not men. This type of cancer if it is not detected early it will spread through all over the body and it will lead to death. Early detection is so important to deal with the undesired situation and treat the patient to be back in to the normal and good health condition. Nowadays, scientists, doctors and engineers are working together to make and find the best solutions for this serious issue. So many difficulties are facing them as the mass in the breast is so small to be palpable or it is hidden in the breast density, moreover, at the beginning of the disease the patient does not feel a serious pain in the breast so it is hard also to detect early. These all factors make the doctors and engineers to try their best to find better ways to solve the problem, this master is done as a small part of these revolution.

During studying master anyone faces dificultuies and needs effort to beat the struggling. First of all I want to thank “Allah” for giving me chance each day to develop myself until I reached today and give me the strength to continue. Then I thank my supervisor Assoc. Prof. Dr. Hasan GÜLER in Firat university, faculty of electric and electronic engineering for supporting me and anyone else stood by my side from the first day until the last day of studying master. Last but not least thanking my family especially my dear mother for being by my side always and leting me to be me as now.

Iman HAMADAMIN
ELAZIG, 2021

TABLE OF CONTENTS

Page

ABSTRACT

Development of Graphical Code-Based Algorithms for the Detection of Abnormalities in Mammogram Images

Iman HAMADAMIN

Master Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences

Department of Electric and Electronics Engineering

August 2021, Page: xi + 55

Nowadays, Mammography is the best available technique and method for early detection of breast cancer. The most common abnormalities in breast that may indicate breast cancer are appeared masses. Also, there are some signs that can lead to breast cancer diagnosis, such as architectural distortion and bilateral asymmetry.

Generally, an algorithm is used to detect and classify breast cancer in mammography images. Three stages are presented: (1) Image preprocessing and mass segmentation, (2) feature selection and extraction and (3) classification.

In the first stage which is preprocessing stage some image processing techniques can be used to enhance the image then segment the suspected mass inside the mammography image of the breast. Image pruning, applying smoothing and gaussian filters, thresholding and morphological operations are exaples and steps for image processing stage. The second stage is for selecting and extracting some features from segmented mass to use these features in the next and last step which is classification. Generally the features are related to the Intensity of the mass compared to the other parts of breast and background of the image. The last stage is classification and as it is clear from its name that it is a system for classifying the mass depending on the extracted features of the segmented mass from the previous step. There are some different algorithms for classifying such as support vector machine (SVM) and artificial neural network (ANN). In this system the ANN machine learning tool is used for the clasification stage to get the best result.

Keywords: Mammography, Segmentation, breast cancer detection, classification, benign and malignant.

ÖZET

Mamogram Görüntülerinde Anormalliklerin Tespiti için Grafik Kod Tabanlı Algoritmaların Geliştirilmesi

Iman HAMADAMIN

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Elektrik ve Elektronik Mühendislik Anabilim Dalı

Agustos 2021, Sayfa: xi + 55

Günümüzde mamografi, meme kanserinin erken teşhisi için en iyi teknik ve yöntemdir. Meme kanserine işaret edebilecek en yaygın anormallikler ortaya çıkan kitlelerdir. Ayrıca meme kanseri teşhisine yol açabilecek bazı işaretler vardır mesela; mimari bozulma ve iki taraflı asimetri.

Genel olarak, mamografi görüntülerinde meme kanserini tespit etmek ve sınıflandırmak için bir algoritma kullanılır. Üç aşama sunulmaktadır: (1) Görüntü önişleme ve kütle bölümleme, (2) özellik seçimi ve çıkarma ve (3) sınıflandırma.

Ön işleme aşaması olan ilk aşamada, görüntüyü iyileştirmek için bazı görüntü işleme teknikleri kullanılabilir ve ardından şüpheli kütleyi memenin mamografi görüntüsü içinde bölümlere ayırır. Görüntü budama, yumuşatma ve gauss filtreleri uygulama, eşikleme ve morfolojik işlemler görüntü işleme aşaması için örnekler ve adımlardır. İkinci aşama, bölümlenmiş kütleden bazı özellikler seçilir ve çıkarılır, bu özellikler, sınıflandırma aşaması olan son aşamada kütleyi sınıflandırmak için kullanılır. Son aşama sınıflandırmadır ve adından da anlaşılacağı üzere, bir önceki adımda parçalanmış kütlenin çıkarılan özelliklerine bağlı olarak kütleyi sınıflandırmak için bir sistemdir. Sınıflandırma için bazı farklı algoritmalar vardır mesela; Destek Vektör Makinesi (DVM) ve Yapay Sinir Ağı (YSA). Bu sistemde sınıflandırma aşamasında en iyi sonucu elde etmek için YSA makine öğrenme aracı kullanılmaktadır.

Anahtar Kelimeler: Mamografi, Segmentasyon, meme kanseri tespiti, sınıflandırma, iyi huylu ve kötü huylu.

LIST OF FIGURES

Page

LIST OF TABLES

Page

ABBREVIATIONS

ANN : Artificial Neural Network
ARCH : Architectural distortion
ASYM : Asymmetry
BANN : Bayesian Artificial Neural Network
BIRADS : Breast Imaging Reporting and Data System
BP : Back Propagation
BUS : Breast Ultrasound
CAD : Computer-Aided Detection
CC : Carnio-Caudal
CDF : Cumulative Distribution Function
CIRC : Well-defined/Circumscribed masses
CLAC : Calcification
CLAHE : Contrast Limited Adaptive Histogram Equalization
CT : Computed Tomography
FFDM : Full-Field Digital Mammography
FN : False Negative
FP : False Positive
GHE : Generalized Histogram Equalization
GLCM : Gray Level Co-occurrence Matrix
GVF : Gradient Vector Flow
IMAQ : Image Acquisition
k-NN : *k*-Nearest Neighbors
LAVIEW : Laboratory Virtual Instrument Engineering Workbench
LoG : Laplacian of Gaussian
MATLAB: Matrix Laboratory
MIAS : Mammographic Image Analysis Society
MISC : Other, ill-defined masses
ML : Maximum Likelihood
MLO : Medio-Lateral-Oblique
MRF : Markov Random Field
MRI : Magnetic Resonance Imaging
NI : National Instruments
NORM : Normal
ROC : Receiver Operating Characteristic
ROI : Region of Interest
SGLD : Spatial Gray Level Dependence
SOM : Self-Organizing Map
SPIC : Spiculated masses
SVM : Support Vector Machine
TN : True Negative
TP : True positive

1. INTRODUCTION

The most common form of cancer in the female population is breast cancer [1]. However it exists in both genders but it affects women more than men. We cannot make an exact statistic for the entire world because of the lag in development in some countries, this type of cancer occurs after the age of 40 in both genders but with different ratio, last years it has affected more than 200.000 women and 2000 men per year in U.S.

Recently, there is an increase in the rate of the affected women from breast cancer. This type of cancer alone accounts for about 22% of the female cancers and approximately 15% of mortality among women having cancer [1]. Although the exact reasons are still not clear but some experienced people think that the reasons are age and sex, they have felt the existence of the masses lately that is why early detection is main way to decrease this rate. Without early detection the affected women may feel the existence of abnormal mass via self-examination when it is metastasized and reached palpable size, so doubtless early detection is really important when the mass is still localized inside breast because if we do not care about it and ignore it, it leads the patient to death.

If we can detect the cancer cells soon the treatment process will be easier and better, that is why nowadays scientists, doctors and engineers all are working together to deal with such a problem and detect any abnormality in breast to avoid letting women and men being attacked by breast cancer.

There are some stages for breast cancer, generally it divided to three stages; local breast cancer, regional breast cancer and distant breast cancer. Each stage has its own specializes depending on the ability of metastasizing of mass cells.

There are two types of masses in the breast, the mass or tumor that is normal called benign mass and the mass or tumor that supposed to be cancer called malignant mass.

In this thesis some different algorithms have been proposed to detect breast cancer. The images have been provided from MIAS database which is a database that a lot of researchers have been used them in their research studies. The program that will be used for programming the system is LabVIEW program which is a graphical code based program. 80 images will be chosen from the database; 40 images with benign masses and 40 images with malignant masses.

Each image will pass through a long process that consists three main stages; image preprocessing and mass segmentation, feature selection and extraction and classification.

In each stage some steps will be applying on the image; in the first stage some different image preprocessing techniques will be applying on the image to develop and enhance it. Image filtering, thresholding, opening and closing in morphological operations, removing undesired regions and segmentation are some examples of image processing.

In the second stage, some features will be selected and extracted from the segmented suspected mass; the features are related to the density of the mass. Standard Deviation, Contrast, Mean Intensity, Skewness, Uniformity, Smoothness and Entropy are the features that will be extracted from the segmented mass.

The last stage is classification, in classification the extracted features will be processed and compared and will be the input for the classifier to calculate the necessary calculation and decide on the segmented mass if it is a benign mass or a malignant mass, almost all of the normal mass are the benign type and the cancer cells are named as malignant mass.

2. BREAST ABNORMALITIES

In this chapter breast abnormalities are explained in detail with its stages and different classes, for this purpose it is necessary to explain breast anatomy and its specifications and how mammography is connected to this topic.

2.1. Overview

As a starting point toward better understanding of breast cancer it is important to know how cancer in general develops. Cancers occurs when control of the division of normal cells is lost and they start to invade other healthy tissues which takes place when a single cell or a group of cells escape from the usual control that regulates cellular growth when they start to multiply, spread and form a mass [2].

When the mass is formed, it can be considered as benign or malignant depending on its shape and behavior. When abnormal growth is restricted to a single and circumscribed mass of cells, it is known as benign. The term “cancer” is used to describe malignant masses which not only can invade surrounding tissues but also have the ability to spread or “metastasize” to distant areas of the body. When the breast masses reach a palpable size, this means that they are metastasized [2].

Properties such margins and shapes help to define masses. For instance, masses with round and smooth margins indicate that they are benign while malignant masses have speculated, rough or blurry boundaries [3].

Mammogram images have important advantages and at the same time few disadvantages as well. Confusing, misconception and misinterpretation of non-cancer masses as cancer and missing some real cancer cells inside the dense of breast during the process of detection are counted as disadvantages. The way of early detection is to take mammography images and work on them but sometimes it is not accurate perfectly and it happens when radiologists take a large number of images and examine them by applying the same algorithm on all of them.

Computer-aided detection (CAD) systems have been recently developed as powerful tools toward identifying malignant from benign masses. An efficient interpretation can be measured through high accuracy of detection and/or diagnosis while maintaining high productivity with a large number of mammography images [4].

Some of the difficulties of detecting abnormal masses in breast are as the following:

1- The existence of noise in the image due to low-quality of radiology device taking mammography images or due to the existence of some irrelative tapes on the image causes the necessity of cutting most images but not all of them.

2- The existence of pectoral muscle and absence in some other making it hard to apply same process and forcing to deal each case differently.
3- In some cases mass is hidden in the fatty breast because of the intensity in the area.

2.2. Breast Anatomy

Figure 2.1 illustrates the structure of the female breast [1]. It consists of ducts, lobules, fat, nipple, pectoralis major muscle and rib cage. The milk producer glands are the lobules, ducts are transferring the milk to nipple during breastfeeding. Amount of fat is giving shape and size to the breast while it is filling the space between lobules, ducts and skin of the breast.

Breast cancer usually begins from two locations: the ducts which are the passages that drain milk from lobules to nipple or the lobules [5].

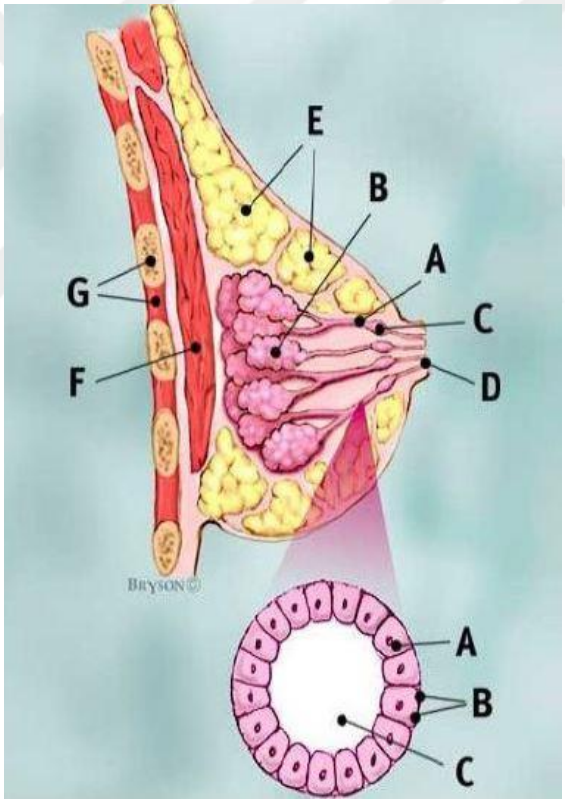

Figure 2.1. Breast Anatomy

From above figure; A: Ducts, B: Lobules, C: Dilated section of duct to hold milk, D: Nipple, E: Fat, F: Pectoralis major muscle, G: Chest wall/rib cage. Enlargements: A: Normal duct cells, B: Basement membrane, C: Lumen (center of duct).

2.3. Stages of Breast Cancer

The stages of breast cancer can be classified to three stages depending on the danger and the distance between cancer cells and original tumor, in other words, measurement of development in metastasis.

2.3.1. Local Breast Cancer

In the first stage, breast cancer is still local. The cancer is still located inside the breast in lobules and ducts and it is not invading the neighborhood tissues. In other words, the normal tissues beyond the breast are not affected with this type of breast cancer [6].

2.3.2. Regional Breast Cancer

The second stage of breast cancer is regional breast cancer. This stage occurs when the cancer cells start the invasion of neighbor tissues and try to reach to the under arm lymph nodes. The lymph nodes are small organs that filter the body from foreign substances. The lymphatic system consists of lymph nodes and ducts that form a network. Its main work is to fight against the foreign substances in the body and filter them [7].

2.3.3. Distant Breast Cancer

In the third stage, which is termed as distant breast cancer, cancer cells are invasive and get into the lymph nodes. They also have a pathway into other parts of the body such lungs, distant lymph nodes, skin, bones, liver and brain [8].

2.4. Breast Cancer Screening and Mammography

All of the women have to take breast cancer screen every while because it is the best way for early detection of tumor if exists. This test can be applied on everyone no matter if the person is healthy or suspected to have abnormal masses in breast, the output of the test can make the person and doctors sure about having any abnormal tumor in breast or even small masses that they cannot be palpable with hands and cannot be seen with eyes. Generally, this test is done to the one whose cancer mass is still a local breast cancer if exists.

Nowadays, many effective methods and devices are developed for detecting the breast cancer. These methods are X-ray mammography, ultrasonography, trans-illumination, thermography, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) all of which are used for breast cancer diagnosis. It has been found that X-ray mammography is the best and the most effective method used for detecting breast cancer [9].

A mass is a lesion that occupies a space in the breast and it can be seen on at least two projections or viewpoints (Carnio-caudal CC and medio-lateral-oblique MLO). The view of a mammogram of Carnio-caudal CC is taken from above while the view of a mammogram of medio-lateral-oblique MLO is taken as oblique or angled view [10].

Nowadays, most of hospitals have digital mammography that can digitize images. The process of digitizing the image is done by capturing the image using electronic X-ray detector that converts it into a digital image to be seen on a computer monitor. Figure 2.2 shows CC and MLO views of the same breast of a subject [2].

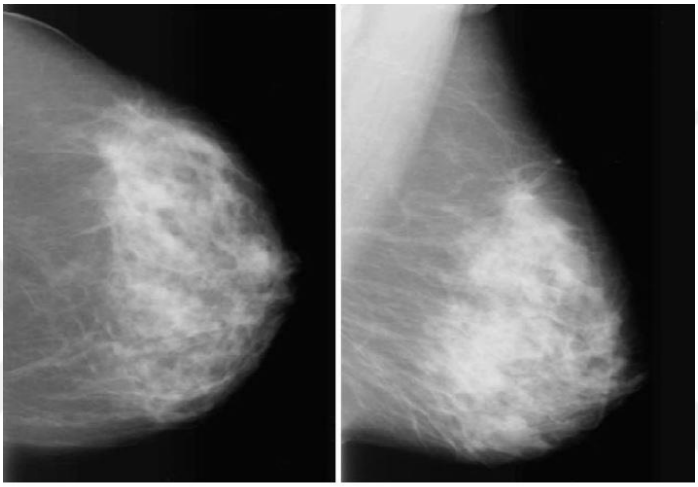

Figure 2.2. Two Basic Views of Mammographic Image; (a) Cranio-cuadal (CC) and (b) medio-lateral oblique (MLO) mammograms of the same breast of a subject.

The process of capturing a mammography image from the patient is done by exposing the patient's breast to X-ray beam source and compressing the breast by breast compression paddle. Compressing the breast is very important and the reason behind that is after compressing the breast, scattered radiation is reduced, motion is eliminated, a uniform density distribution is created and thereby increasing the visibility of details in the image. Figure 2.3 shows how mammography images are taken from the patient [11].

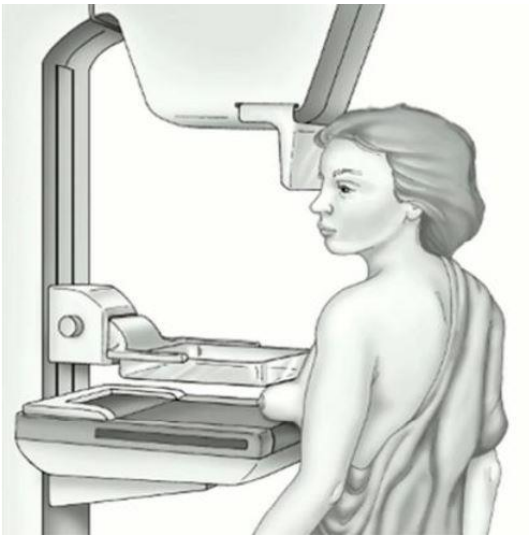

Figure 2.3. Mammogram / Cranio-cuadal (CC)

Mammography has been recommended for the asymptomatic women who have their breast cancer screening periodically done. It offers high quality images at a low radiation dose, and is currently the only widely accepted imaging method used for routine breast cancer screening [9].

In the process of diagnosing there are several characteristics associated with breast cancer disease indicator, including; micro-calcifications, masses, architectural distortions and any asymmetry between breasts. Although the presence of pain in breast.

In general, microcalcifications are small size lesions, which can develop during the early levels of breast cancer. These lesions have a size typically in the range 0.05 to 1 mm see Figure 2.4 Because of their very small sizes, microcalcifications can be quite difficult and rather challenging to detect. Their presence in the breast zone can vary and they often can be distinguished as bright. On the other hand, these lesions can come in different sizes and shapes and their distribution can be different from patient to patient. In addition, on occasions there can be seen low contrast in their color because of the reduced intensity difference between the suspicious spots and the area surrounding these lesions. It will be very difficult to detect all these factors. Furthermore, the proximity to the surrounding tissues will add to the challenge in their detection. In the cases of dense tissues, suspicious areas can hardly be detected because of the tissue superimposition. The following sections give some examples of the types of these lesions. Some anatomic structures such as fibrous strands, breast borders or hypertrophied lobules are similar to microcalcifications in the mammographic image [12].

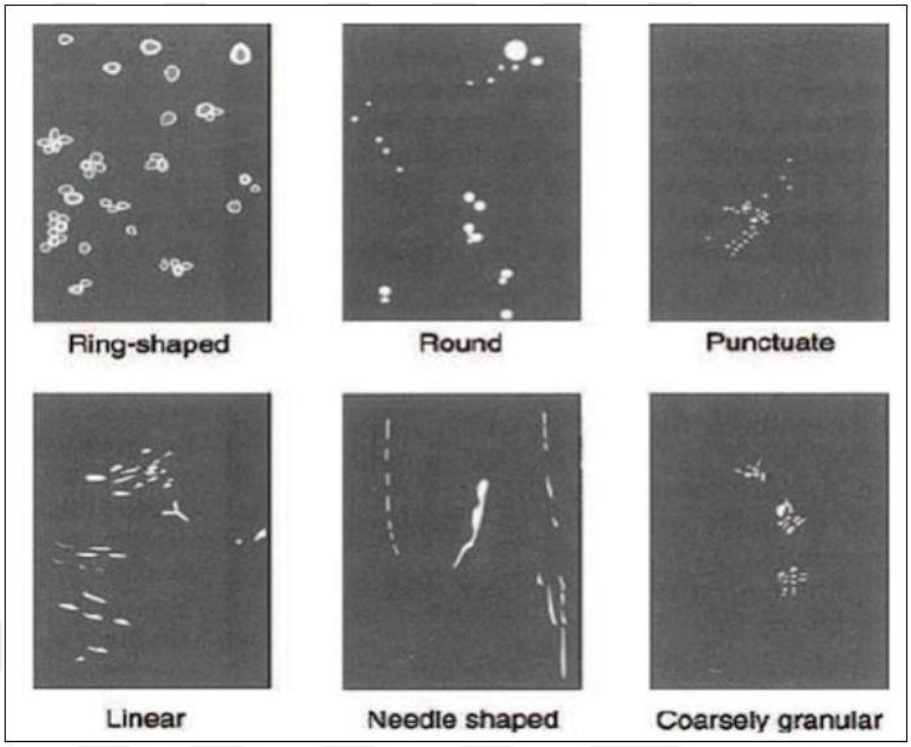

Figure 2.4. Type of microcalcifications commonly seen on mammographic images [13].

Research work and scientific studies highlight that the correlation between the presence of microcalcifications and breast cancer is high. Moreover, when the microcalcifications appear in clusters, this is especially profound. Therefore, an accurate detection for microcalcifications is fundamental to an early detection for the majority of breast [14].

Generally, benign calcifications come in round bigger elliptical shapes with uniform size however, non-uniformed , small, polymorphic and radiate calcifications, with heterogeneous volume and morphology have higher chance to being malignant [15] see it in Figure 2.5

Masses appear as intensive regions of different characteristic and volume. They can be lobular, circular, oval or non-uniform/speculated and their boundaries can be [15], Figure 2.6

- Circumscribed, these are well-define and distinctly demarcated borders.
- Obscured, these are hiding behind superimposed or adjacent tissue.
- Micro-lobulated, these have undulating circular borders.
- Ill-defined, these are poorly defined scattered borders.
- Speculated, these are radiating thin lines.

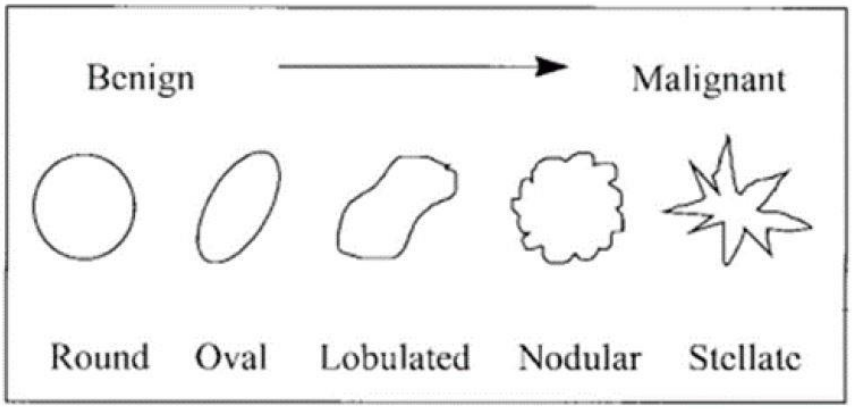

Figure 2.5. Morphologic spectrum of mammographic masses [16]

Earlier studies have indicated which depending on the morphology; the masses usually found to have several chance of malignancy. For instance, the speculated and ill-defined boundaries have higher chance of malignant [15]. A benign kind commonly, shown with the existence of elliptical or circular masses. Other studies reported that the great variability of the mass appearance could be a huge challenge as well as an obstacle to a correct mammography analysis [16]. Furthermore, the derangement of the normal configuration of the parenchyma in an arbitrary or radiating or pattern called architectural distortions, without a visible center or mass. These are very hard to discover it because of it is very variable [17].

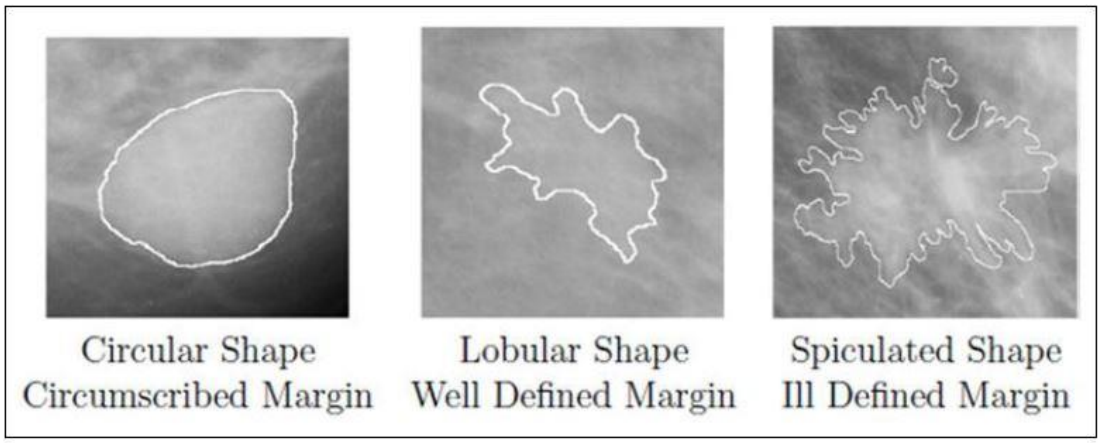

Figure 2.6. Mass examples with different shapes and borders [18]

There are some difficulties in breast scanning and abnormal mass detection because of the difference of breast fatty.

The American College of Radiology Breast Imaging Reporting and Data System (BIRADS) becoming a standard on the assessment of mammographic images and uses four categories for density evaluation [19]:

- BIRADS I: the breast is almost entirely fatty.
- BIRADS II: there is some fibro-glandular tissue.
- BIRADS III: the breast is heterogeneously dense.
- BIRADS IV: the breast is extremely dense.

All the types are shown in figure 2.7:

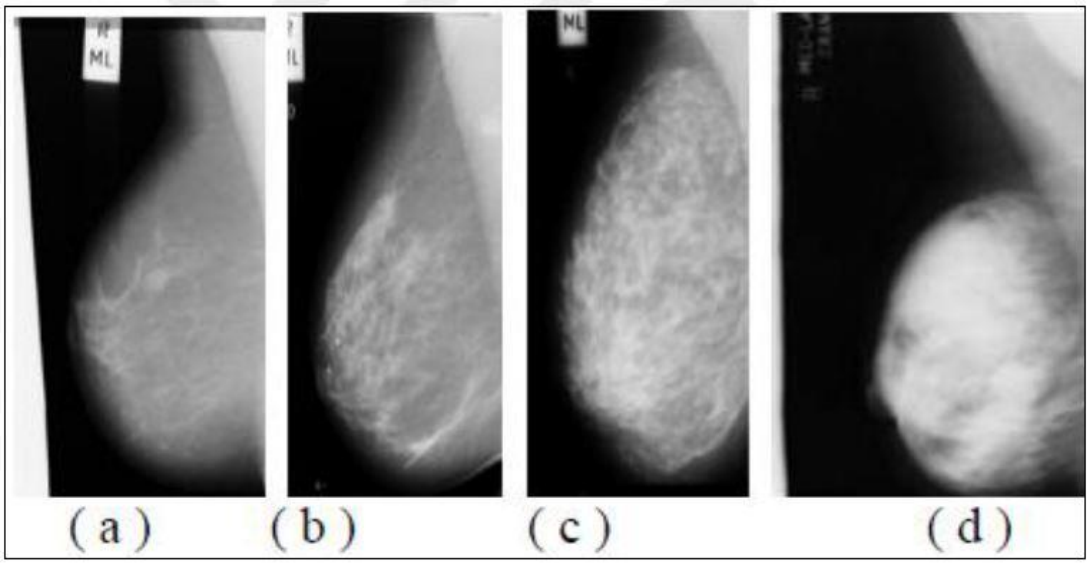

Figure 2.7. Examples of Mammograms, where the breast density increases from (a) BIRADS I to (d) BIRADS IV

2.5. Severity of Abnormality

There are two types of abnormal masses in the breast, they called benign masses and malignant masses.

2.5.1. Benign Masses

The benign masses are considered to be in the first stage of breast cancer stages because they lack the ability to metastasize and lack invasive properties of cancer as it is shown in figure 2.8. Although benign masses can have side effects but many kinds of this type of mass are not harmful to human health condition and they are not life threatening.

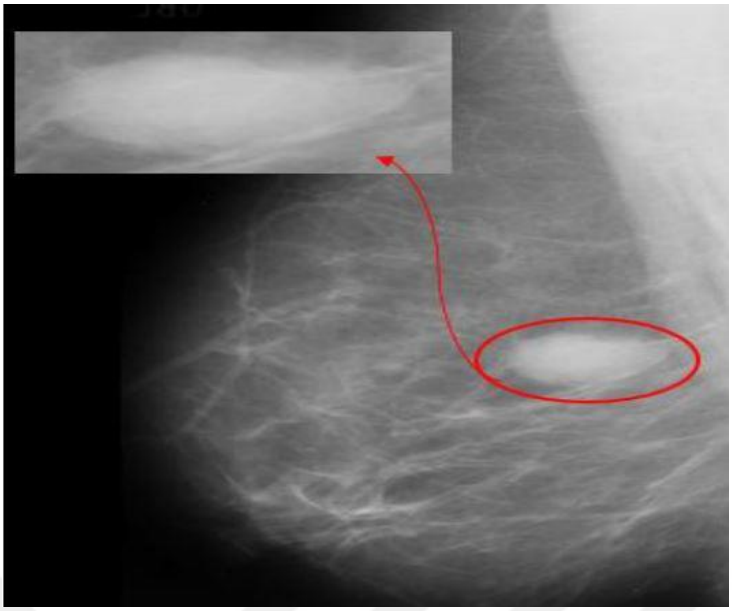

Figure 2.8. Benign Mass

2.5.2. Malignant Masses

The malignant masses are considered to be the second and third stage of breast cancer stages because they are not self-limited in growth as it is shown in figure 2.9. They have the capability of invading the neighboring tissues around the breast and spreading to distant regions of the body. Therefore, the term “cancer” is used for malignant masses which are usually more serious and more dangerous for human health condition.

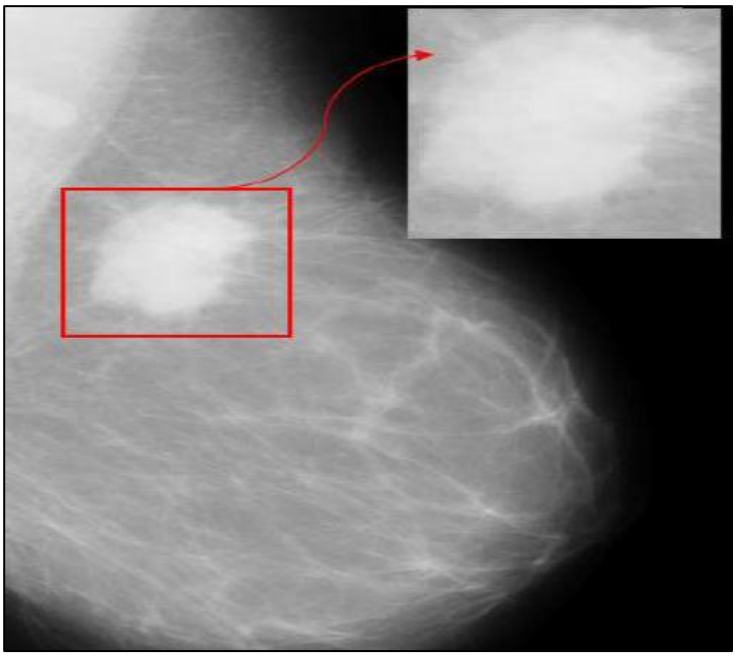

Figure 2.9. Malignant Mass

2.6. Classes of Abnormalities

There are some classes of abnormalities found after the mammography images are taken. Expert radiologists can determine the type of the abnormality just from its appearance and then they can determine whether the mass in the mammography image is benign or malignant mass. Breast cancer screening mammograms found these types of abnormalities: calcification (i.e., macro and micro-calcifications), spiculated masses, well-defined/circumscribed masses, architectural distortion, asymmetry breast tissue and other miscellaneous findings [3].

In mammography images, the appearance of macro-calcifications looks as large white dots distributed randomly within the breast. It is found that half of women over 50 and 10 women under that age have macro-calcifications in their breast. Macro-calcifications are considered as noncancerous and for that reason no need to do follow-up care [20] as it is shown in figure 2.10. Although Micro-calcifications are usually not an indicator or result of breast cancer but they can become dangerous when they are clustered in a group and appear in a certain pattern, when they grouped together, they are considered as a start indicator of breast cancer.

If calcifications are detected in the breast, doctors categorize them into three types in order to be treated. The first category is benign calcifications which are considered harmless and no need to do treatment for them. Another is probably benign calcifications. It is found that more than 98% of this type as noncancerous. Typically, they are monitored every six months for at least one year. If there is no change found after a year of follow-up, doctor's recommendation is to have a routine mammogram once a year [20].

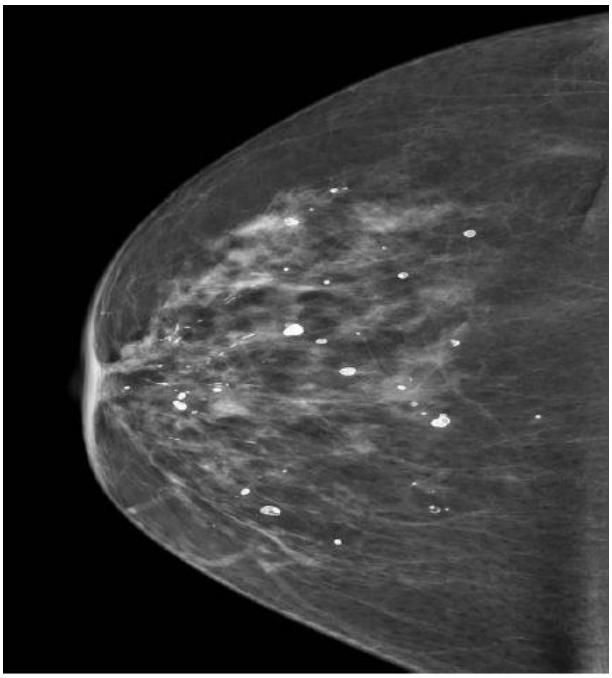

Figure 2.10. Calcification

A spiculated mass is considered as the most dangerous class of abnormality since it is one of the primary indicators of cancer [21]. This type of masses can be anywhere inside the human body but often found in breasts or lungs. When these spiky masses are found in anywhere inside body even in the breasts doctor's recommendation is to give a biopsy to confirm whether they are malignant or benign. If they are malignant, the treatment can range from excision to radiation, as it is shown in figure 2.11.

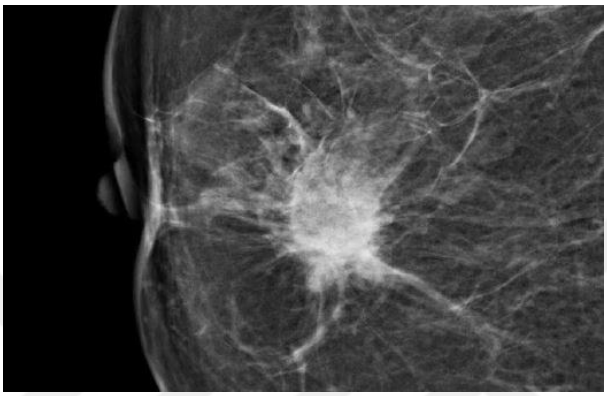

Figure 2.11. Spiculated Mass

Well-defined/circumscribed masses are another class of abnormalities found in breast cancer screening and mammography. Another term can be used for this type of masses which is circumscribed carcinoma. This term refers to ductal carcinoma that appears as circumscribed on a mammogram. Although circumscribed carcinoma is less frequently seen than typical spiculated carcinoma, it has both types of severity of abnormality benign and malignant. Circumscribed carcinoma includes medullary, invasive ductal carcinoma and other types [22], as it is shown in figure 2.12.

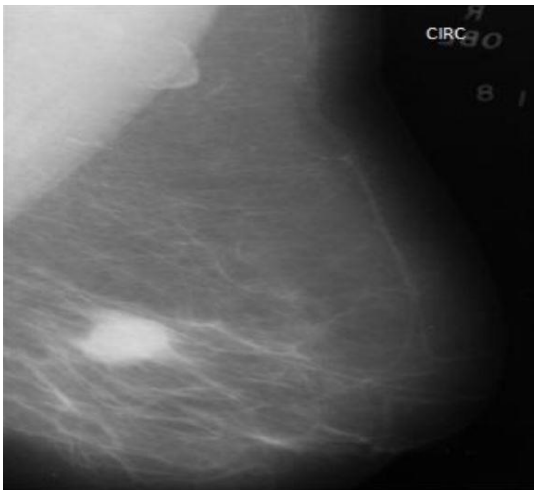

Figure 2.12. Circumscribed Mass

Architectural distortion is the last class of abnormalities discussed in this chapter. It is considered as the third most common class of abnormalities according to its appearance. It is found that 6% of abnormalities have this type. The incidence of architectural distortion is small compared to calcifications and visible mass. However, when it is exist in the mammography image, it is difficult to be detected and diagnosed because of its variability in presentation [23] as it is shown in figure 2.13.

The appearance of architectural distortion in mammography images seems like a disruption in the structure of the breast itself. The most interesting thing in this class of abnormality is that there is no mass to indicate and name it abnormal mass but the distortion appears as a stellate shape or with radiating spiculations like the masses found in spiculation case.

A scar inside the breast formed from a previous surgery which is benign can be interpreted as architectural distortion. Although the reason of architectural distortion can be a result of a benign disease, it is found that almost 80% of the detected masses are a result of invasive breast cancer [23].

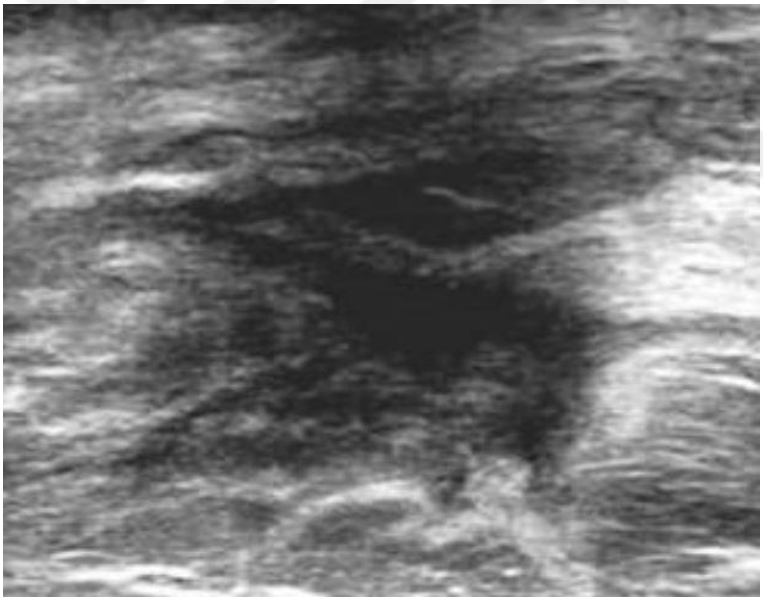

Figure 2.13. Architectural Distortion

3. IMAGE PROCESSING IN BREAST CANCER

There are many different algorithms proposed in literature for breast cancer detection and classification thesis. This thesis presents the most successful methods used in the literature. The most recent breast cancer detection and classification algorithms can be presented in three sections, as the following: image preprocessing and mass segmentation, feature selection and extraction and classification. In this part all of the algorithms and methods have been used in the project are explained in detail. The terms are used to make the process easy for understanding with the mathematical expressions. During these three stages mathematical formulas are implemented to get the desired results.

3.1. Image Processing and Mass Segmentation

The aim of image preprocessing stage is to modify and enhance the original image to improve its visual appearance without destroying or damaging the important features of the image for diagnosing of abnormalities.

Jumaat, A.K. et al. [24] worked on forty different breast ultrasound images obtained from palace of the golden horses screening center in Malaysia. They used Adobe Photoshop CS2 as the first step of preprocessing of the image in order to crop the regions of interest. They obtained the new size of the images as 64 × 64 pixels. Then, they applied median filter on the images in order to remove speckle noise. The last step of their preprocessing stage was applying histogram stretching method to enhance the contrast of the images.

Rahmati, P. et al. [25] used mammography images in their study, their first step was applying the enhanced version of the contrast limited adaptive histogram equalization (CLAHE) which they referred to as the Fuzzy CLAHE. FCLAHE minimizes the statistical change of the intensity so the mammography image is enhanced as the result.

Shi, X. et al. [26] used breast ultrasound (BUS) images which are in low contrast. By using the developed Multi-peak GHE (Generalized Histogram Equalization) they enhanced the images then changed the order of gray levels in the image so the enhancement procedure was made completely controllable as the result.

Yao, Y. [27] worked on Magnetic Resonance Imaging (MRI). In his study, he used sharpening spatial filters, sharpening frequency filters and Sobel operator to enhance the details of the image and to detect the mass inside the image. He used a 5 × 5 mask approximation to Laplacian of Gaussian (LoG). Then, according to that mask he filtered the input image.

Kom, G. et al. [28] recognized that the masses in the mammography images have different sizes and shapes with different densities. By using the local adaptive thresholding the mass areas

are segmented. Small and large windows located around the pixels are used to calculate the adaptive threshold.

Varela, C. et al. [29] also applied adaptive thresholding in their study to segment the suspicious masses. They clustered the pixels according to the Cumulative Distribution Function (CDF). They obtained the binary image, the values of the pixels which have intensity values greater than 93% are set to one. Otherwise, their value is zero. Their system could detect 135 true masses from the total number of test images which is equal to 138. The images were previously enhanced with an iris filter.

Jumaat, A.K. et al. [24] used balloon snake method to segment the suspicious masses found in ultrasound images. Active contour (snake) method is a combination of two operations; mathematical optimization conception and computer technology, so they used this active counter. The boundaries of the object are traced by a computer generated curve.

Nguyen, A. et al. [30] used Sobel operator for detecting the mass in the breast in combination with some MATLAB morphological operations. They dilated the image using a vertical and horizontal line structuring elements of length three, filled the holes in the ROI and used opening to remove the small objects.

Shi, X. et al. [26] used Markov Random Field (MRF) which is one of the segmentation methods. The pixels were classified according to the labels put on each pixel. This statistical method used the local neighborhood relationship to represent the global relationship, this made the system more consistent and more tolerant to noise.

Rahmati, P. et al. [25] used maximum likelihood active contour model using level sets (MLACMLS) algorithm to segment the suspicious mass. The main aim of their algorithms was to estimate a segmentation contour that differentiates between the foreground and background regions. For separation the mass from the background region they used Gamma distribution.

Dominguez, A.R. and Nandi, A.K. [31] used multilevel-thresholding segmentation method. They segmented regions by converting the original images to binary images at multiple threshold levels. The results of their method concluded that 30 levels were sufficient to segment all the mammography images with intensity values in the range [0,1] with step size of 0.025.

Zou, F. et al. [32] proposed a method that uses gradient vector flow (GVF) field. After the mammography images were enhanced by using adaptive histogram equalization, the boundaries of the region of interest (ROI) were determined according to GVF field component with the larger entropy.

In image processing stage image enhancement methods are implemented to process the image to make the result be more suitable than the original image. It brings out some features in region of interest that are invisible or difficult to notice in the original image. Image enhancement

techniques include: image filtering, thresholding, morphological operations and some more simple operations that will be showed in later paragraphs.

Histogram

The histogram of a digital image in general is a discrete function that represents the number of pixels in each different gray level which called intensity level. It can be written as follows:

$$h(r_k) = n_k \tag{3.1}$$

Where r_k represents the k^{th} gray level and n_k is the number of pixels in an image having intensity level r_k. The range of intensity level for 8-bit grayscale image will be [0, L-1]; i.e., the range will be [0 - 255] since$L = 2^k$. The histogram can be normalized and its range becomes [0 - 1] according to:

$$P(r_k) = n_k/n \tag{3.2}$$

Where n is the total number of pixels in the image.

Figure 3.1 shows grayscale image histogram. The x-axis corresponds to the gray level values r_k and the y-axis corresponds to the number of pixels in those gray scales n_k. The zero value in the histogram indicates the black pixels in the image and L-1 value which is equal to 255 indicates the white pixels in the image. While other values indicate the distribution of [0 - L-1] intensities in the image in other words they are the gray levels from dark to light gray, histogram sample is shown in figure 3.1.

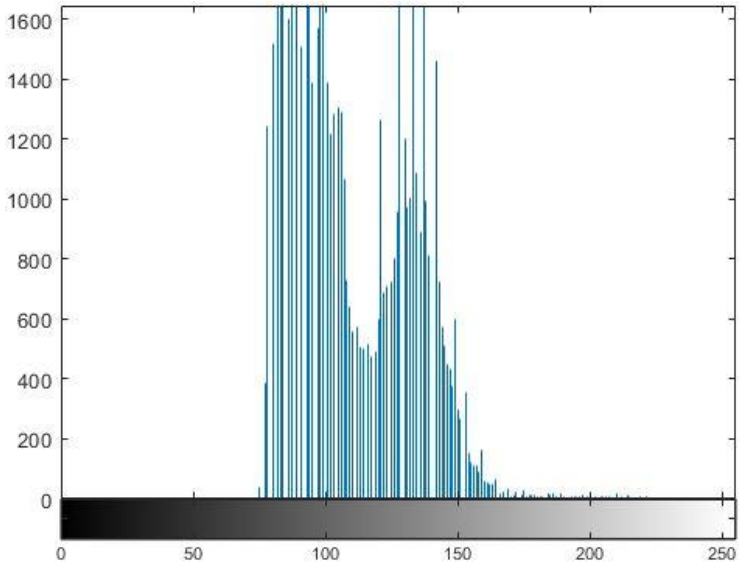

Figure 3.1. Histogram of grayscale image

Kernel Family Filters

In Kernel block, there are four matrixes that can be used as filters after being convolved with the original image; the block has some inputs as code for being able making the exact desired filter. The block reads a predefined kernel. This code consists of three separate units: Kernel Family, Kernel Size, and Kernel Number. If the code is already known, it can be entered directly with Kernel Code. Kernel family determines the type of matrix. This value corresponds to the thousandth unit in the researched code. The matrix types are shown in table 3.1:

Table 3.1. Kernel family filters

Gradient (1)	Specifies the kernel family as Gradient
Laplacian (2)	Specifies the kernel family as Laplacian
Smoothing (3)	Specifies the kernel family as Smoothing
Gaussian (4)	Specifies the kernel family as Gaussian

Kernel size determines the horizontal and vertical matrix size. The values are 3, 5, and 7, corresponding to the convolutions 3 × 3, 5 × 5, and 7 × 7 supplied in the matrix catalog. This value corresponds to the hundredth unit in the researched code. Kernel number is the matrix family number. It is a two-digit number, between 0 and n, belonging to a family and a size. A number of predefined matrices are available for each type and size.

Gradient Kernels contains the following filters: Prewitt Filters, they have the following kernels. The notations West (W), South (S), East (E), and North (N) indicate which edges of bright regions they outline.

Table 3.2. Prewitt filters (3*3)

#0 W/Edge	**#1 W/Image**	**#2 SW/Edge**	**#3 SW/Image**
-1 0 1 -1 **0** 1 -1 0 1	-1 0 1 -1 **1** 1 -1 0 1	0 1 1 -1 **0** 1 -1 -1 0	0 1 1 -1 **1** 1 -1 -1 0
#4 S/Edge	**#5 S/Image**	**#6 SE/Edge**	**#7 SE/Image**
1 1 1 0 **0** 0 -1 -1 -1	1 1 1 0 **1** 0 -1 -1 -1	1 1 0 1 **0** -1 0 -1 -1	1 1 0 1 **1** -1 0 -1 -1
#8 E/Edge	**#9 E/Image**	**#10 NE/Edge**	**#11 NE/Image**
1 0 -1 1 **0** -1 1 0 -1	1 0 -1 1 **1** -1 1 0 -1	0 -1 -1 -1 **0** 1 1 1 0	0 -1 -1 1 **1** -1 1 1 0

#12 N/Edge	#13 N/Image	#14 NW/Edge	#15 NW/Image
-1 -1 -1	-1 -1 -1	-1 -1 0	-1 -1 0
0 **0** 0	0 **1** 0	-1 **0** 1	-1 **1** 1
1 1 1	1 1 1	0 1 1	0 1 1

Sobel filters, they are very similar to the Prewitt filters except that they highlight light intensity variations along a particular axis that is assigned a stronger weight. The Sobel filters have the following kernels. The notations West (W), South (S), East (E), and North (N) indicate which edges of bright regions they outline.

Table 3.3. Sobel filters (3*3)

#16 W/Edge	#17 W/Image	#18 SW/Edge	#19 SW/Image
-1 0 1	-1 0 1	0 1 2	0 1 2
-2 **0** 2	-2 **1** 2	-1 **0** 1	-1 **1** 1
-1 0 1	-1 0 1	-2 -1 0	-2 -1 0
#20 S/Edge	**#21 S/Image**	**#22 SE/Edge**	**#23 SE/Image**
1 2 1	1 2 1	2 1 0	2 1 0
0 **0** 0	0 **1** 0	1 **0** -1	1 **1** -1
-1 -2 -1	-1 -2 -1	0 -1 -2	0 -1 -2
#24 E/Edge	**#25 E/Image**	**#26 NE/Edge**	**#27 NE/Image**
1 0 -1	1 0 -1	0 -1 -2	0 -1 -2
2 **0** -2	2 **1** -2	-1 **0** 1	1 **1** -1
1 0 -1	1 0 -1	2 1 0	2 1 0
#28 N/Edge	**#29 N/Image**	**#30 NW/Edge**	**#31 NW/Image**
-1 -2 -1	-1 -2 -1	-2 -1 0	-2 -1 0
0 **0** 0	0 **1** 0	-1 **0** 1	-1 **1** 1
1 2 1	1 2 1	0 1 2	0 1 2

Table 3.4. Gradient filters (5*5)

#0 W/Edge	**#1 W/Image**	**#2 SW/Edge**	**#3 SW/Image**
0 -1 0 1 0	0 -1 0 1 0	0 0 1 1 1	0 0 1 1 1
-1 -2 0 2 1	-1 -2 0 2 1	0 0 2 2 1	0 0 2 2 1
-1 -2 **0** 2 1	-1 -2 **1** 2 1	-1 -2 **0** 2 1	-1 -2 **1** 2 1
-1 -2 0 2 1	-1 -2 0 2 1	-1 -2 -2 0 0	-1 -2 -2 0 0
0 -1 0 1 0	0 -1 0 1 0	-1 -1 -1 0 0	-1 -1 -1 0 0
#4 S/Edge	**#5 S/Image**	**#6 SE/Edge**	**#7 SE/Image**
0 1 1 1 0	0 1 1 1 0	1 1 1 0 0	1 1 1 0 0
1 2 2 2 1	1 2 2 2 1	1 2 2 0 0	1 2 2 0 0
0 0 **0** 0 0	0 0 **1** 0 0	1 2 **0** -2 -1	1 2 **1** -2 -1
-1 -2 -2 -2 -1	-1 -2 -2 -2 -1	0 0 -2 -2 -1	0 0 -2 -2 -1
0 -1 -1 -1 0	0 -1 -1 -1 0	0 0 -1 -1 -1	0 0 -1 -1 -1
#8 E/Edge	**#9 E/Image**	**#10 NE/Edge**	**#11 NE/Image**
0 1 0 -1 0	0 1 0 -1 0	0 0 -1 -1 -1	0 0 -1 -1 -1
1 2 0 -2 -1	1 2 0 -2 -1	0 0 -2 -2 -1	0 0 -2 -2 -1
1 2 **0** -2 -1	1 2 **1** -2 -1	1 2 **0** -2 -1	1 2 **1** -2 -1
1 2 0 -2 -1	1 2 0 -2 -1	1 2 2 0 0	1 2 2 0 0
0 1 0 -1 0	0 1 0 -1 0	1 1 1 0 0	1 1 1 0 0
#12 N/Edge	**#13 N/Image**	**#14 NW/Edge**	**#15 NW/Image**
0 -1 -1 -1 0	0 -1 -1 -1 0	-1 -1 -1 0 0	-1 -1 -1 0 0
-1 -2 -2 -2 -1	-1 -2 -2 -2 -1	-1 -2 -2 0 0	-1 -2 -2 0 0
0 0 **0** 0 0	0 0 **1** 0 0	-1 -2 **0** 2 1	-1 -2 **1** 2 1
1 2 2 2 1	1 2 2 2 1	0 0 2 2 1	0 0 2 2 1
0 1 1 1 0	0 1 1 1 0	0 0 1 1 1	0 0 1 1 1

Table 3.5. Gradient filters (7*7)

#0 W/Edge	**#1 W/Image**	**#2 SW/Edge**
0 -1 -1 0 1 1 0	0 -1 -1 0 1 1 0	0 1 1 1 1 1 0
-1 -2 -2 0 2 2 1	-1 -2 -2 0 2 2 1	1 2 2 2 2 2 1
-1 -2 -3 0 3 2 1	-1 -2 -3 0 3 2 1	1 2 3 3 3 2 1
-1 -2 -3 **0** 3 2 1	-1 -2 -3 **1** 3 2 1	0 0 0 **0** 0 0 0
-1 -2 -3 0 3 2 1	-1 -2 -3 0 3 2 1	-1 -2 -3 -3 -3 -2 -1
-1 -2 -2 0 2 2 1	-1 -2 -2 0 2 2 1	-1 -2 -2 -2 -2 -2 -1
0 -1 -1 0 1 1 0	0 -1 -1 0 1 1 0	0 -1 -1 -1 -1 -1 0
#3 S/Image	**#4 E/Edge**	**#5 E/Image**
0 1 1 1 1 1 0	0 1 1 0 -1 -1 0	0 1 1 0 -1 -1 0
1 2 2 2 2 2 1	1 2 2 0 -2 -2 -1	1 2 2 0 -2 -2 -1
1 2 3 3 3 2 1	1 2 3 0 -3 -2 -1	1 2 3 0 -3 -2 -1
0 0 0 **1** 0 0 0	1 2 3 **0** -3 -2 -1	1 2 3 **1** -3 -2 -1
-1 -2 -3 -3 -3 -2 -1	1 2 3 0 -3 -2 -1	1 2 3 0 -3 -2 -1
-1 -2 -2 -2 -2 -2 -1	1 2 2 0 -2 -2 -1	1 2 2 0 -2 -2 -1
0 -1 -1 -1 -1 -1 0	0 1 1 0 -1 -1 0	0 1 1 0 -1 -1 0
#6 N/Edge	**#7 N/Image**	
0 -1 -1 -1 -1 -1 0	0 -1 -1 -1 -1 -1 0	
-1 -2 -2 -2 -2 -2 -1	-1 -2 -2 -2 -2 -2 -1	
-1 -2 -3 -3 -3 -2 -1	-1 -2 -3 -3 -3 -2 -1	
0 0 0 **0** 0 0 0	0 0 0 **1** 0 0 0	
1 2 3 3 3 2 1	1 2 3 3 3 2 1	
1 2 2 2 2 2 1	1 2 2 2 2 2 1	
0 1 1 1 1 1 0	0 1 1 1 1 1 0	

Table 3.6. Laplacian Filters (3*3)

#0 Contour 4	**#1 +Image*1**	**#2 +Image*2**
0 -1 0	0 -1 0	0 -1 0
-1 4 -1	-1 5 -1	-1 6 -1
0 -1 0	0 -1 0	0 -1 0
#3 Contour 8	**#4 +Image*1**	**#5 +Image*2**
-1 -1 -1	-1 -1 -1	-1 -1 -1
-1 8 -1	-1 9 -1	-1 10 -1
-1 -1 -1	-1 -1 -1	-1 -1 -1
#6 Contour 12	**#7 +Image*1**	
-1 -2 -1	-1 -2 -1	
-2 12 -2	-2 13 -2	
-1 -2 -1	-1 -2 -1	

Table 3.7. Laplacian filters (5*5)

#0 Contour 24	**#1 +Image*1**
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1
-1 -1 **24** -1 -1	-1 -1 **25** -1 -1
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1

Table 3.8. Laplacian filters (7*7)

#0 Contour 48	**#1 +Image*1**
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1
-1 -1 **48** -1 -1	-1 -1 **49** -1 -1
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1
-1 -1 -1 -1 -1	-1 -1 -1 -1 -1

Table 3.9. Smoothing filters (3*3)

0 1 0 1 0 1 0 1 0	0 1 0 1 1 1 0 1 0	0 2 0 2 1 2 0 2 0	0 4 0 4 1 4 0 4 0
1 1 1 1 0 1 1 1 1	1 1 1 1 1 1 1 1 1	2 2 2 2 1 2 2 2 2	4 4 4 4 1 4 4 4 4

Table 3.10. Smoothing filters (5*5)

1 1 1 1 1	1 1 1 1 1
1 1 1 1 1	1 1 1 1 1
1 1 **0** 1 1	1 1 **1** 1 1
1 1 1 1 1	1 1 1 1 1
1 1 1 1 1	1 1 1 1 1

Table 3.11. Smoothing filters (7*7)

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 **0** 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 **1** 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Table 3.12. Gaussian filters (3*3)

0 1 0 1 2 1 0 1 0	0 1 0 1 4 1 0 1 0	1 1 1 1 2 1 1 1 1
1 1 1 1 4 1 1 1 1	1 2 1 2 4 2 1 2 1	1 4 1 4 16 4 1 4 1

Table 3.13. Gaussian filter (5*5)

1 2 4 2 1 2 4 8 4 2 4 8 16 8 4 2 4 8 4 2 1 2 4 2 1

Table 3.14. Gaussian filter (7*7)

1 1 2 2 2 1 1 1 2 2 4 2 2 1 2 2 4 8 4 2 2 2 4 8 **16** 8 4 2 2 2 4 8 4 2 2 1 2 2 4 2 2 1 1 1 2 2 2 1 1

All of the tables and filter matrixes are taken from the library of LabVIEW program, all tables as; table 3.2, table 3.3, table 3.4, table 3.5, table 3.6, table 3.7, table 3.8, table 3.9, table 3.10, table 3.11, table 3.12, table 3.13 and table 3.14.

Thresholding

Thresholding is a simple method used for segmenting the breast in the mammography image. It is called global since it is based on the global information of the image like histogram and a single threshold value is selected for the whole image. The global threshold value can be found easily because the intensity values of the abnormality regions are greater than the surrounding tissue [33].

A binary image is obtained after applying thresholding in which the intensities of the grayscale image are partitioned. The pixels which are greater than the threshold value are classified to be the pixels of the breast and their values are one in other words the white pixels and the pixels which are less than the threshold value are classified to be the pixels of the background and their values are zero in other words the black pixels. The following expression explains the global thresholding:

$$g(x,y) = \begin{cases} 1 & if\ f(x,y) \geq T \quad (ROI - Breast) \\ 0 & Otherwise \quad\quad (Background) \end{cases} \tag{3.3}$$

Where $g(x,y)$ indicates the binary image resulted from segmentation, $f(x,y)$ indicates the input grayscale image and T indicates the selected threshold value. The initial estimate for threshold value T is selected to be between [150- 200] for all the mammography images to segment the breast.

Morphological Operation

Morphological operations are very useful in image processing for extracting, describing and improving the shapes of regions of interest. They can be used in pre-processing and post-processing operations. Most of the time, morphological operations are used in binary images since they rely on the relative ordering of pixels and not on their numerical values [34]. Dilation and erosion are the two major morphological operations used in this thesis and most of researches done so far [35].

Dilation is one of the most basic morphological operations. It is used in order to thicken or grow the region of interest or to bridge the small gaps between neighboring regions. When dilation is applied, the boundaries of regions of interest will be smoother [35]. Dilation can be defined as follows:

$$A \oplus B = \{z \mid \left(\hat{B}\right)_z \cap A \neq \emptyset\} \tag{3.4}$$

The above definition indicates that a set A can be dilated by another set B and the result is the dilation $A \oplus B$. This means that B which is a small shape or template called structuring element (STREL) can dilate the set A by reflecting B about its origin $\left(\hat{B}\right)$ and translating this reflection by z, then moving the structuring element over the set A providing that there is an overlapping between A and $\left(\hat{B}\right)_z$ by at least one element. The term convolution can be used since B can be

considered as a dilation mask which is moving over the set A. As a result of dilation, new pixels are added to the boundaries of the region of interest of set A, as it is shown in figure 3.2.

The following example shows the result of dilating a simple binary image containing one rectangle object by a structuring element of type line with degree 45. The gray box in the middle of B indicates the origin of the structuring element B.

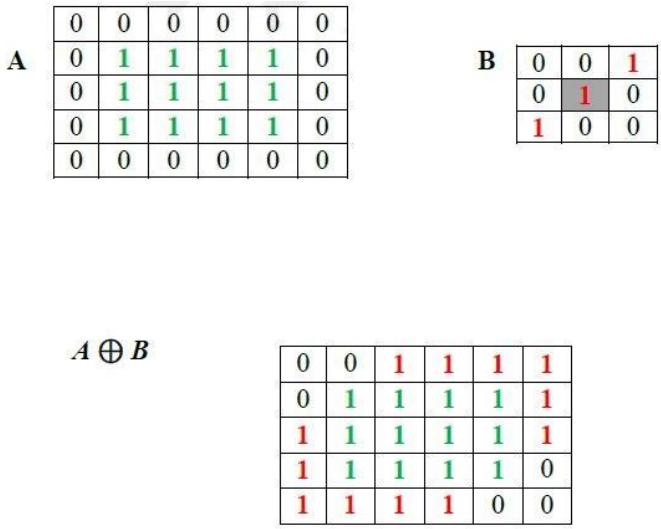

Figure 3.2. Dilation

Erosion is another basic morphological operation. It is used in order to eliminate the irrelevant details of the region of interest. Erosion has the opposite effect of dilation. While dilation thickens or grows the region of interest by adding new pixels to the boundaries of the object, erosion thins or shrinks the region of interest by removing pixels from boundaries [16]. Erosion can be defined as follows:

$$A \ominus B = \{z \mid (B)_z \subseteq A\} \tag{3.5}$$

The above definition indicates that a set A can be eroded by another set B and the result is the erosion $A \ominus B$. This means that the structuring element B can erode the set A by convolving B over the set A providing that the translated structuring element $(B)_z$ fits onto A, i.e., B is totally contained within A. As a result, holes and gaps between different regions become larger, the small details will be eliminated and the region of interest will be shrunk.

The following example shows the result of erosion. A simple binary image containing one rectangle object is eroded by a structuring element of type vertical line. The gray box in the middle of the structuring element B indicates the origin of it, as it is shown in figure 3.3.

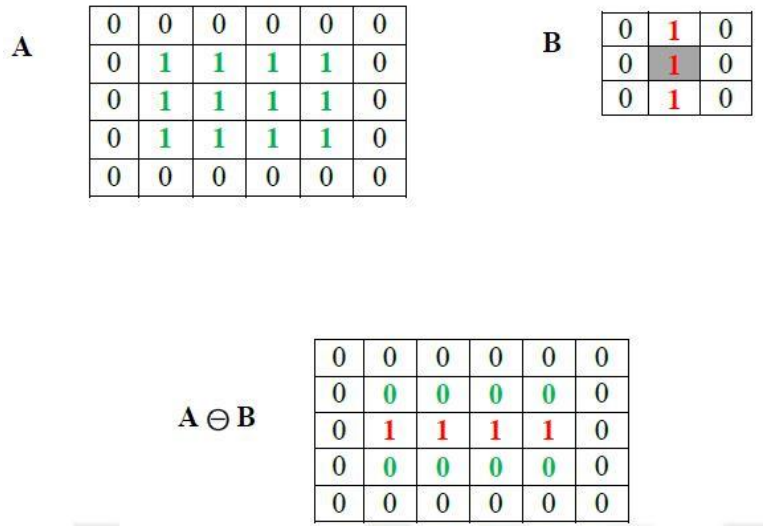

Figure 3.3. Erosion

Opening and closing are two morphological operations resulted from the two basic morphological operations dilation and erosion. They are a combination of dilation and erosion. Opening is an erosion operation followed by dilation while closing is a dilation operation followed by erosion [17]. Opening and closing can be defined as follows:

$$A \circ B = (A \ominus B) \oplus B \tag{3.6}$$

$$A \bullet B = (A \oplus B) \ominus B \tag{3.7}$$

Opening is used in order to smooth the outer contour of the object and make it more visible. It is called opening because it opens a gap and breaks the narrow connections between the neighboring regions that are connected with a thin bridge of pixels. Also, if there is a thin protrusion in the image, opening can eliminate it. On the other hand, closing is used in order to remove holes and gaps existing between the regions smoothing the object contour.

Border Rejection and Mass Segmentation

In Labview there is a block for rejecting all the parts linked to the border of the image. The principle of working for this part is like this; after having a binary image which means black and white image, each white part connected to the first and last rows and columns of image pixels are rounded to zero, it causes removing and rejection all connected white parts to the border.

3.2. Feature Selection and Extraction

The aim of feature selection and extraction stage is to select the most effective features that can accurately differentiate between malignant and benign masses. Then, extract the feature set from breast cancer masses for classification stage.

Shi, X. et al. [26] divided the features extracted from the suspicious mass into three categories: textural, fractal dimensions and histogram-based features. The number of the extracted features obtained is: 140 texture features, 5 fractal dimensions and 6 histogram-based features. They applied regression method in order to decrease the number of the selected features and that method produced an optimal subset of 13 features including eight texture features, three fractal dimensions and two histogram-based features.

R.M. et al. [36] proposed methods in their study to obtain shape features extracted from the turning angle functions of contours. Shape-based features are useful in describing the shape of the boundary of the masses found in the breast and they give important details of shape related to spicules and lobulations

Features of shape and margin of the mass can be used to differentiate between malignant and benign masses. Figure 3.4 shows the well-known shapes and margins of the mass [5].

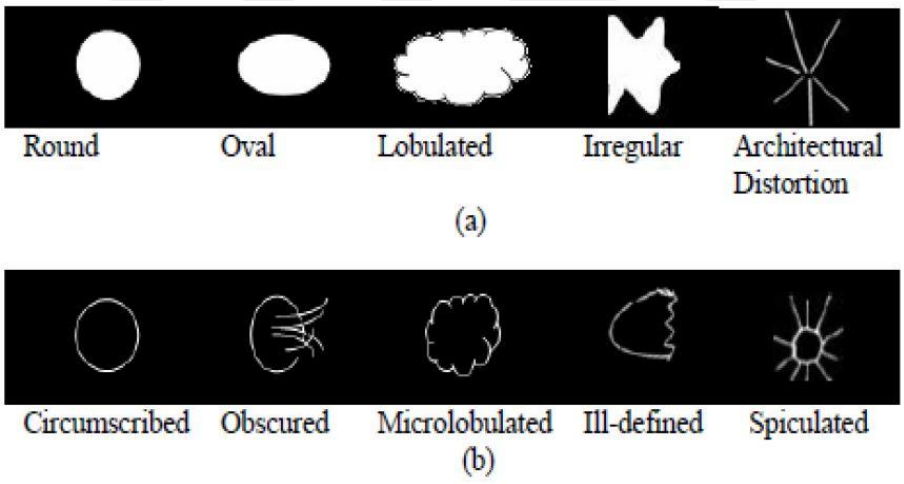

Figure 3.4. Shapes and Margins of Mass (a) Shapes (b) Margins

Yuan, Y. et al. [37] worked on full-field digital mammography (FFDM) database. They classified the extracted features into three groups: the first group included characterizing spiculation, margin, shape and contrast of the mass. The second group included texture features extracted from Gray Level Co-occurrence Matrix (GLCM). The last group included a distance feature calculated as a Euclidean distance from the nipple to the center of the mass.

Bellotti, R. et al. [38] worked on a large number of mammography images (3369 image). They used texture features extracted from GLCM, which is also known as spatial gray level

dependence (SGLD) matrix to characterize the region of interest of the image. They selected eight features that remain unchanged under monotonic transformation.

Timp, S. et al. [39] improved the character of mass by adding information about the mass behavior over time in order to differentiate between malignant and benign masses. They proposed temporal features extracted from the current mammographic images which are combined with the features calculated from the corresponding region in mammographic image taken in previous exam to provide temporal information. After that, they divided the extracted temporal features into two kinds: different features and similar features. Different features measured changes in feature values between corresponding regions in the prior and the current view. Similar features measured whether two regions are comparable in appearance. Finally, they concluded that the change of benign masses is slow and the appearance of two consecutive screening mammograms is similar. Malignant masses may change considerably and become more suspicious by time.

Some researchers do not only use one category of features. They can use many categories of features to classify the mass. For example, Ball, J.E. and Bruce, L.M. [21] used shape-based features and texture features to classify the mass in the region of interest. Moreover, they used another feature which is the patient's age.

After the mass segmentation it is turn to extract some features from the mass to select them and use them for classification. Features as; Standard deviation, Mean Intensity, Mean contrast, Skewness, Entropy, Smoothness and Uniformity are proper features to extract from the mass to classify it as Benign or Malignant. The features are extracted from the histogram of the intensity levels of the segmented mass.

Standard Deviation

It is a measure of average contrast of the segmented mass. It can be calculated from the following expression:

$$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2} \tag{3.8}$$

Mean Intensity

It is a measure of average intensity of the segmented mass. It can be calculated from the following expression:

$$m = \sum_{i=0}^{L-1} z_i \, p(z_i) \tag{3.9}$$

Where z indicates a random intensity value, $p(z)$ indicates the histogram of the grayscale image and L indicates the number of gray (intensity) levels. The values of pixels in 8-bit images will be in the range [0 $L-1$]; i.e., the values of pixels will range between [0 - 255] .

Contrast

Contrast adjustment is another method used to enhance the image. The pixels of the image have different values of intensity which can lead to different types of regions with low contrast and high contrast regions. The objective of contrast adjustment method is to increase the contrast between the different regions. Contrast can be measured from Maximal value minus Minimal value.

Skewness

It is also known as skewness of the histograms. It measures the symmetry of the histogram. The value of skewness will be 0 for symmetric histograms, positive for histograms skewed to the right (about the mean) and negative for histograms skewed to the left. It can be calculated from the following expression:

$$\mu_3 = \sum_{i=0}^{L-1}(z_i - m)^3\, p(z_i) \tag{3.10}$$

Entropy

It is a measure of randomness. The following expression shows the equation of entropy:

$$e = \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \tag{3.11}$$

Smoothness

It is a measure of the relative smoothness of the intensity in the segmented mass. The value of smoothness (R) is equal zero for a region of contrast intensity and reaches one if the region has large excursions in the values of its intensity levels. Smoothness can be calculated from the following expression:

$$R = 1 - \frac{1}{1+\sigma^2} \tag{3.12}$$

Uniformity

It features the uniformity. This measure is maximum when all gray levels are equal (maximum uniform) and decreases from there for the inequality. The following expression explains how the uniformity is calculated:

$$U = \sum_{i=0}^{L-1} P^2(z_i) \tag{3.13}$$

3.3. Classification

The aim of doing classification stage is to classify the mass as benign or malignant based on the selected and extracted features.

Shi, X. et al. [26] used fuzzy support vector machine as the classification tool. Then, they used five performance metrics to evaluate the classification results. Those metrics are: accuracy, sensitivity, specificity, positive predictive value and negative predictive value. Their system achieved 91.67% sensitivity, 96.08% specificity, 94.25% accuracy, 94.29% positive predictive value and 94.23% negative predictive value.

Cascio, D. et al. [40] used artificial neural network ANN as a classifier in their study. They selected the number of input neurons to be 12, the number of output neuron to be one and the number of the hidden neurons was tuned to obtain the best classification performance. The performance of their proposed ANN system was found to be Az = 0.862 ± 0.007 where Az is the average area under the curve of receiver operating characteristic (ROC).

Song, J.H. et al. [15] also used ANN as a classifier to differentiate between malignant and benign masses. They worked on 54 ultrasound images (20 malignant and 34 benign). The extracted features were margin sharpness, margin echogenicity, angular continuity and age of patients. The performance of their proposed system was found to be Az = 0.856 ± 0.058.

Oliver, A. et al. [41] used three different ways in mammogram classification. They used a *k*-Nearest Neighbors (k-NN) classifier, a decision tree classifier and a Bayes classifier based on the combination of the first two classifiers.

Ball, J.E. and Bruce, L.M. [21] used *k*-Nearest Neighbors (k-NN) and maximum likelihood (ML) classifiers to classify the masses as malignant or benign. When the number of nearest neighbors is 1 or 2, the accuracy of the classifier was 93% with three FP and one FN. Using ML classifier, they achieved 92% overall accuracy with three FP and two FN.

Chen, C.M. et al. [42] used hierarchical ANNs as a classifier and the performance of their proposed system was found to be Az = 0.9840 ± 0.0072. Although the performance of hierarchical ANNs classifier is high, it has a main disadvantage which is time consuming step of training the whole data.

Li, H. et al. [43] used Bayesian artificial neural network (BANN) classifier. To differentiate between malignant and benign masses, they combined the selected features and the process of

combination gave them better results than the individual selected features. The performance of their proposed system was found to be Az = 0.83.

Classification is the last stage in which the extracted region of interest is classified. Many techniques are used to classify the region of interest. Support vector machine (SVM) and artificial neural network (ANN) are some of the most frequently used classifiers [44]. As the classification is the last stage in breast cancer detection algorithm. Here the selected and extracted features are fed into a classifier in order to classify the segmented mass as malignant or benign mass.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning technique which means that the input data and the class of the object are fed into the learning machine. The main point of support vector machine is to design a hyper-plane that can classify all the training data into two classes by separating them. In order to keep the two classes of samples separated the width and the orientation of hyper-plane can be changed.

It can be found that many hyper-planes can be applicable in separating the two classes of data. However, when they leave the margin between the two classes maximum, they are considered as optimal [45].

Artificial Neural Network (ANN)

ANN is a classifier in which its construction is composed of mathematical models similar to nervous system [33]. ANN is a classifier used to determine whether the segmented suspected mass is benign or malignant according to the extracted features. Its construction is composed of mathematical models and algorithms similar to nervous system of human.

ANN has three main layers: the input layer, the hidden layer and the output layer. Each layer is composed of neurons. Figure 3.5 shows the components of neural network.

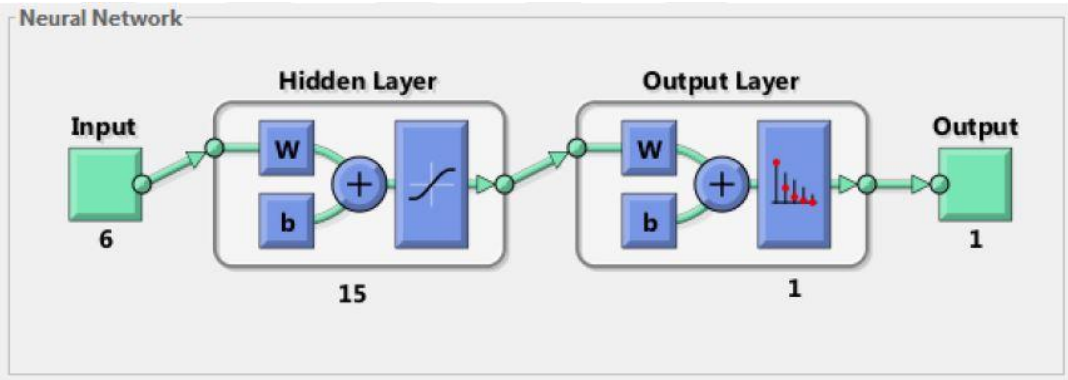

Figure 3.5. Neural Network

In breast cancer detection, three types of artificial neural networks are frequently used. These types are: back-propagation neural network, self-organizing map (SOM) and hierarchical ANN [42].

The first type is BP neural network. It is a feed-forward ANN with supervised learning process. This means that the input data fed into the learning machine should be labeled with the type of the class of mass to get the desired output. The number of layers and neurons is not necessary to be fixed because it depends on the purpose of using the neural network.

The second type of artificial neural networks is SOM. It is a totally unsupervised method which means that the training data is not provided with labels so it is not a good choice for this type of projects. The last type of artificial neural networks is a hierarchical ANNs and its principle of working is different from what is proper for this project. In this thesis, Back-propagation ANN is selected because it is robust and widely applicable. While the other methods need to define some parameters in the process of constructing the model.

4. DEVELOPMENT OF GRAPHICAL CODE BASED BREAST CANCER DETECTION AND CLASSIFICATION

In this chapter all materials and methods that used in this thesis are explained. Materials that have been used in this thesis are a database of mammographic images and a software program of Graphical-code based programming language. The dataset of mammography images used in this thesis is called Mammographic Image Analysis Society (MIAS) Mini Mammographic Database. It has earlier version which has been digitized at 50 μm pixel size but in the new version Mini-MIAS database, the images have been sampled at 200 μm pixel size, 8 bit gray-level quantization and then clipped/padded so that the size of each image became 1024 × 1024 pixels. The MIAS database has been used in several research studies and now it is considered as a benchmark database [43]. Table 4.1 shows the meaning of each column of the mammography images in MIAS database.

Table 4.1. Characteristics of the MIAS Database Mammograms

Columns	Explanation	Phrases
1st column	Reference number	MIAS database reference number
2nd column	Character of background tissue	F Fatty G Fatty-glandular D Dense-glandular
3rd column	Class of abnormality present	CALC Calcification CIRC Well-defined/circumscribed masses SPIC Spiculated masses MISC Other, ill-defined masses ARCH Architectural distortion ASYM Asymmetry NORM Normal
4th column	Severity of abnormality	B Benign M Malignant
5th & 6th column	Coordinates	*x,y* image-coordinates of center of abnormality
7th column	Radius	Approximate radius (in pixels) of a circle enclosing the abnormality

In this database images are taken from 161 patients, each single patient has left mammogram (even numbers of filename) and right mammogram (odd numbers of filename). All the views of mammography images are taken as MLO views. The total number of mammography images in MIAS database is 322 cases. They are divided into normal images and images that have

abnormalities. The total number of normal mammography images is 207 while the other 115 have different classes of abnormalities.

The programming language used in this thesis is LabVIEW. Laboratory Virtual Instrument Engineering Workbench (LabVIEW) is a system-design platform and development environment for a visual programming language from National Instruments (NI). The graphical language is named "G" not to be confused with G-code.

In this thesis LabVIEW version 2017 is used and all the functions are implemented in LabVIEW. When using LabVIEW, data can be analyzed, algorithms can be developed, images can be processed and applications can be created. Also, LabVIEW has built-in math functions and tools which give a fast solution. It has many applications such as: image processing, computer vision, signal processing, control system, statistics, etc. Nowadays, LabVIEW is used by many researchers, engineers and scientists.

For such a project some different toolkits are needed to install in order to make all the process work perfectly, the toolkits are; vision toolkit, real time toolkit, analytics and machine learning toolkits and advanced signal processing toolkit.

In this chapter the results of applying each step in the system is showed. There are three main stages that each one of them has some steps to be done. The three main stages are; Image processing and mass segmentation, Feature selection and extraction and classification.

The images of system in block diagram and the results in front panel is showed to make it clear what is the result of applying each step. First, the steps of the Image processing and mass segmentation are done. Then, the features extracted in the segmented mass. Finally the classification is done.

4.1. Image Processing and Mass Segmentation

The NI Vision toolkits and NI-IMAQ toolkits are needed to make this main stage because of the built-up block diagrams in these toolkits, generally the algorithms that can be applied on images are used in the parts and steps of this stage. From Image loading into the program to the Mass segmentation step are showed in this part. All necessary parts are shown for understanding how the program works and how the process goes on.

4.1.1. Image Loading and Pruning

In our steps we take the image with the reference code of (mdb184) in MIAS database of Mammography images as a sample to apply the steps on it. The first step is load the image of mammography in to the program, as shown in figure 4.1 and figure 4.2.

After loading the image it is turn to prune it because there are some unnecessary spaces from the first and last columns in some of the images in the database. Determining the four numbers as [x1, y1, x2, y2] for the rectangle of the interested region, as shown in figure 4.3 and figure 4.4.

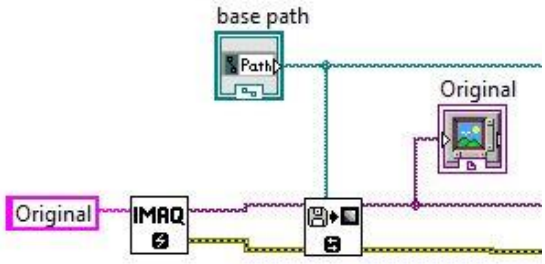

Figure 4.1. Original Image loading and showing (Block Diagram)

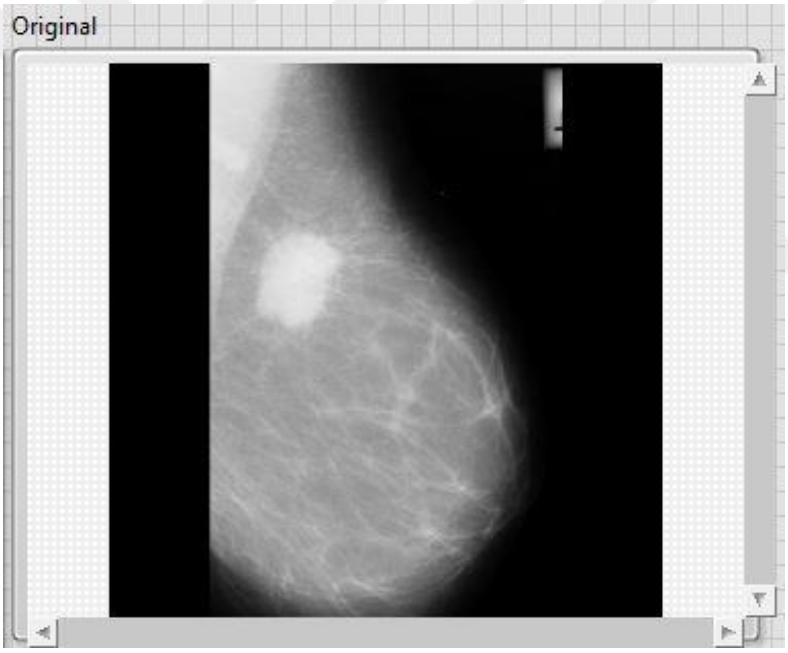

Figure 4.2. Original Image (Front Panel)

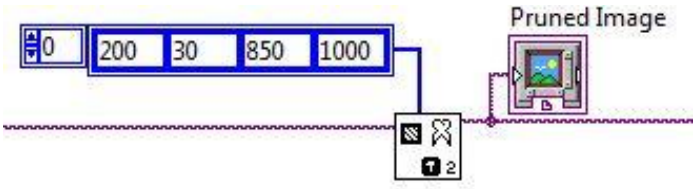

Figure 4.3. Pruning the Image (Block Diagram)

Figure 4.4. Pruned Image (Front Panel)

4.1.2. Applying Filters on the Image

After pruning the image it is time to apply some filters on the image to make it more suitable for other next steps. The first filter is (Smoothing) from Kernel families. The process is convolution the matrix of Smoothing and the image, as shown in figure 4.5, figure 4.6, figure 4.7 and figure 4.8.

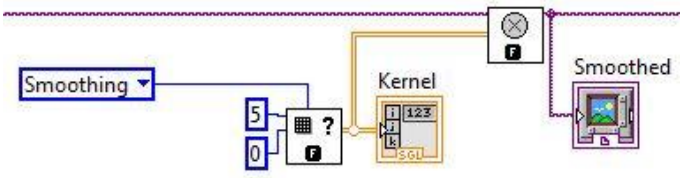

Figure 4.5. Smoothing the Image (Block Diagram)

Figure 4.6. Smoothed Image (Front Panel)

Second filter is Gaussian filter in Kernel family.

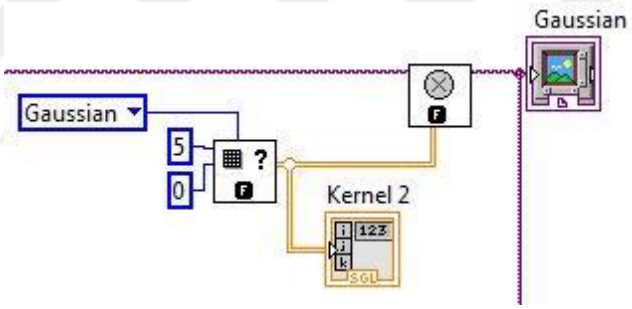

Figure 4.7. Gaussian filter on the Image (Block Diagram)

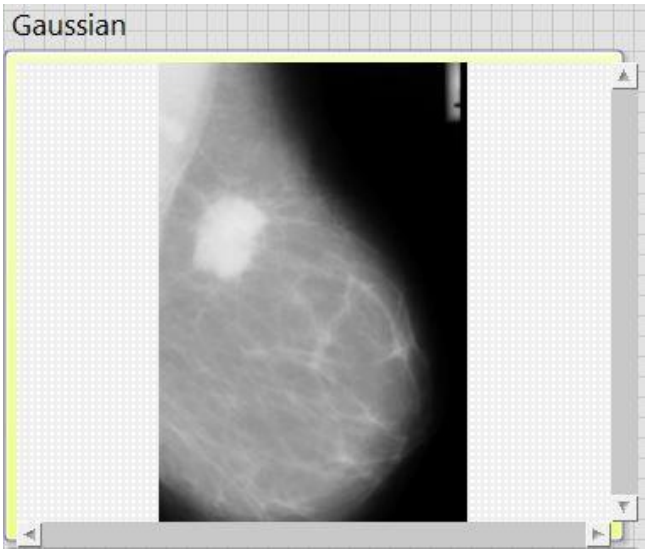

Figure 4.8. Gaussian filter on the Image (Front Panel)

4.1.3. Thresholding

Thresholding is applied on the grayscale image for changing it to binary image. The only thing to do is determining the value of threshold, as shown in figure 4.9 and figure 4.10.

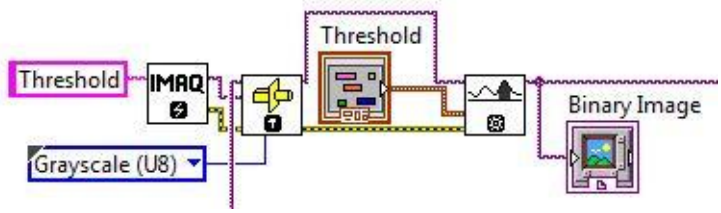

Figure 4.9. Thresholding and getting the Binary Image (Block Diagram)

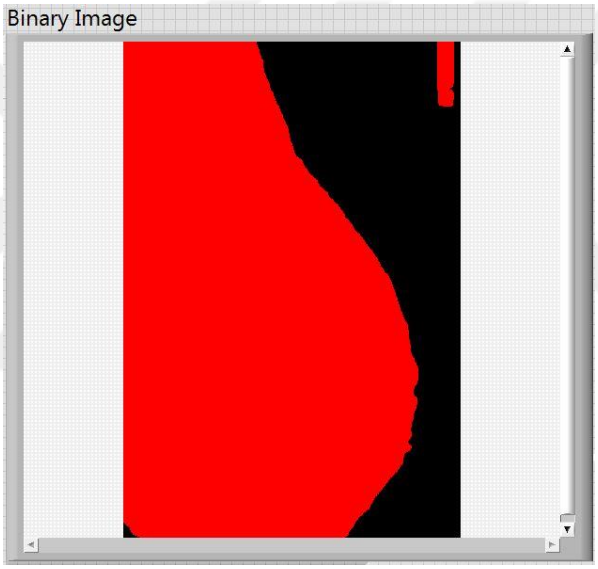

Figure 4.10. Binary Image (Front Panel)

4.1.4. Preprocessed Image

The morphological operations have been applied on the binary image according the system demand the Erosion and Dilation operations are applied on the image for making it more available for next step, in other hand the last version of grayscale image before changing it to binary image is token as (Masked image). Then, these two images which are the grayscale image and the binary image multiplied to each other to get the preprocessed image, as shown in figure 4.11 and figure 4.12. The preprocessed image is ready to segment any region of interest from it.

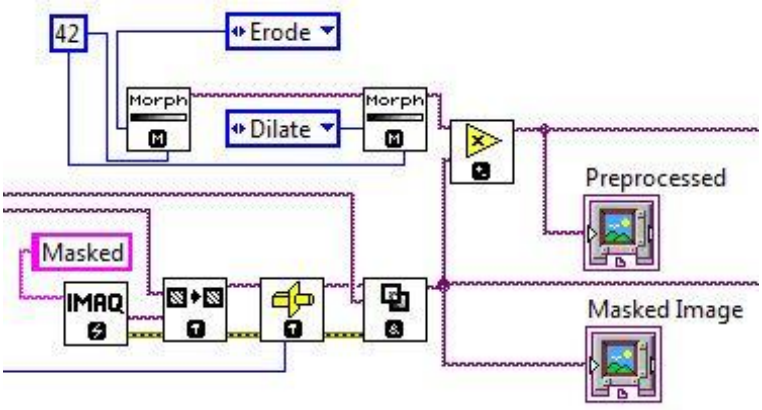

Figure 4.11. Morphological operations on the binary image and multiplying it to the grayscale masked image to get the preprocessed image (Block Diagram)

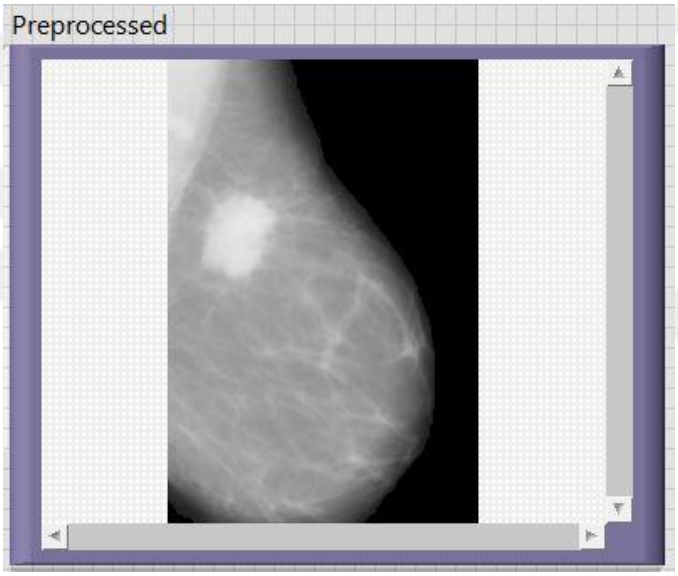

Figure 4.12. Preprocessed Image (Front Panel)

4.1.5. Removing Unnecessary Parts

After getting the preprocessed image, threshold value is determined and thresholding is applied on the preprocessed image to get only the bright regions as a binary image, in other words, the output of this step is a binary image that only the bright parts are ON, all other parts are OFF. The bright parts can be pectoral muscle, some fatty parts inside breast because of the intensity and the suspected mass. Then, this binary image is fed to the border rejection block diagram which removes all ON parts connected to the border of the image; the pectoral muscle is removed in this way. After that, the image is fed to particle remover filter to remove all small particles and pixels that are not the region of interest. Finally, what remained as a bright region is only the suspected mass, as shown in figure 4.13 and figure 4.14.

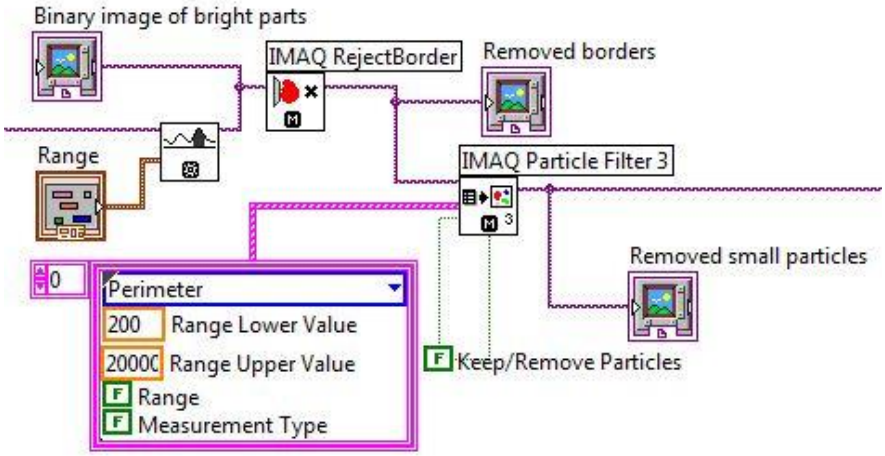

Figure 4.13. Removing Unnecessary Parts (Block Diagram)

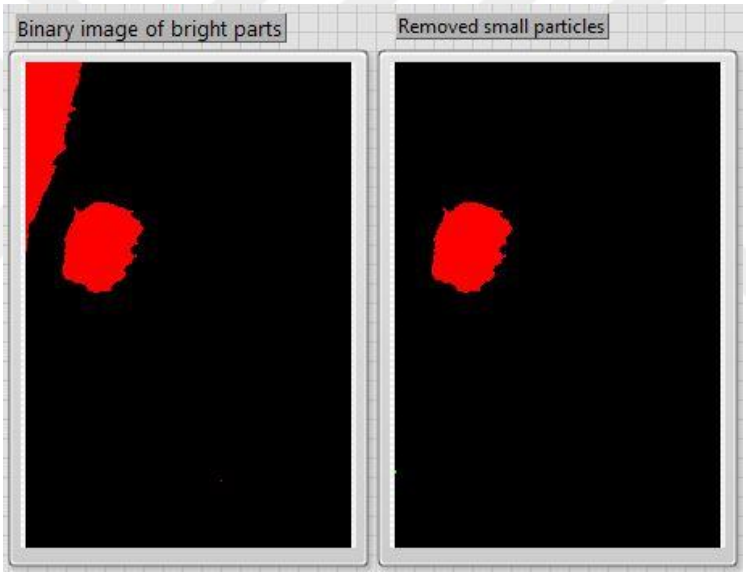

Figure 4.14. Removed Unnecessary Parts (Front Panel)

4.1.6. Mass Segmentation

After getting the binary image of only the mass as it can be called the mask image for the suspected mass, the next step is to multiply the preprocessed image with this mask image of the suspected mass. The result of the multiplication is the segmentation of the mass because the only shown part is the mass. Then, applying Gaussian filter on the result of multiplication to make the mass be detected perfectly, as shown in figure 4.15 and figure 4.16. Finally he segmented mass is ready to take out some features from it.

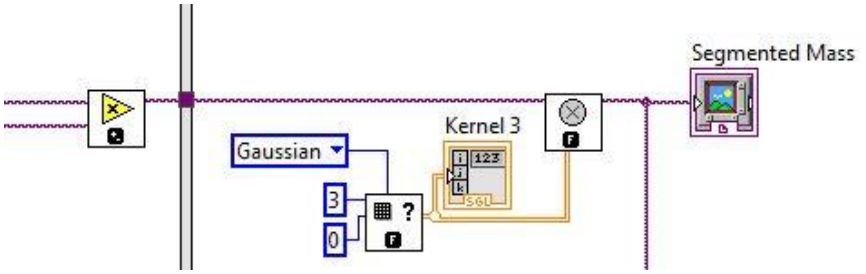

Figure 4.15. Mass Segmentation (Block Diagram)

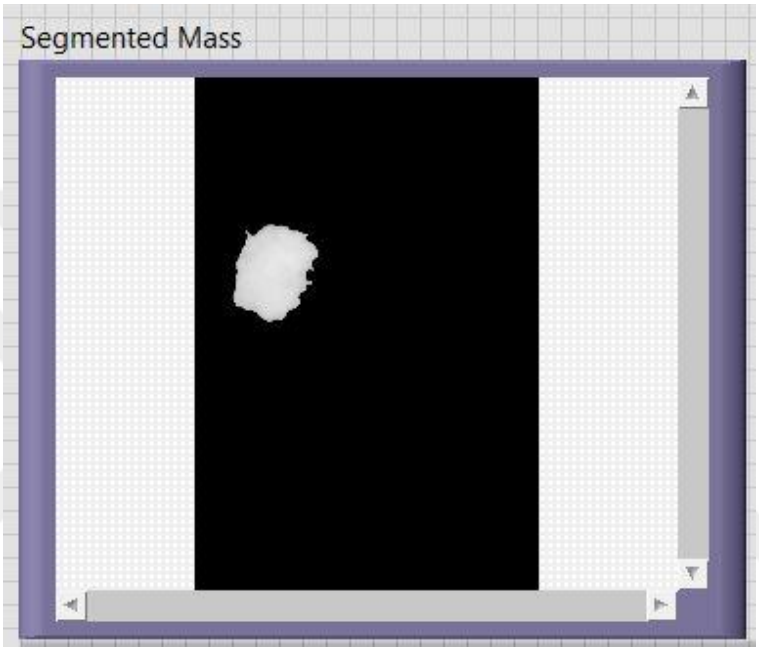

Figure 4.16. Segmented Mass (Front Panel)

4.1.7. Histogram

The histogram block diagram already exists in the program, so by using this block the histogram of the image can be taken, as shown in figure 4.17 and figure 4.18.

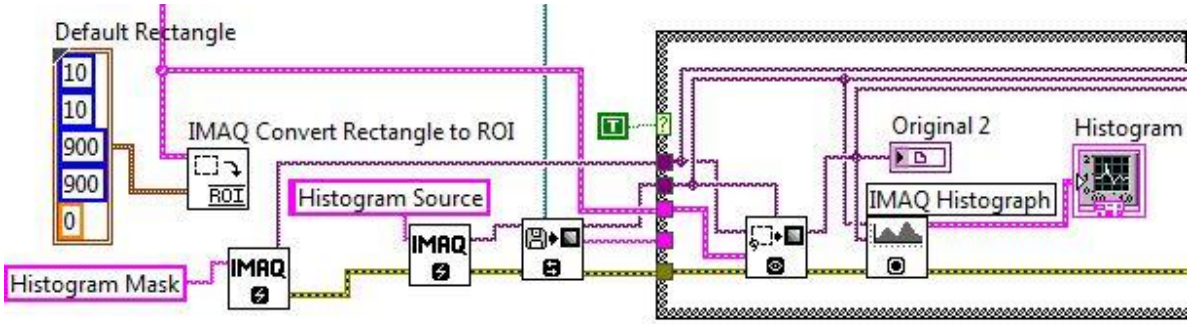

Figure 4.17. Histogram (Block Diagram)

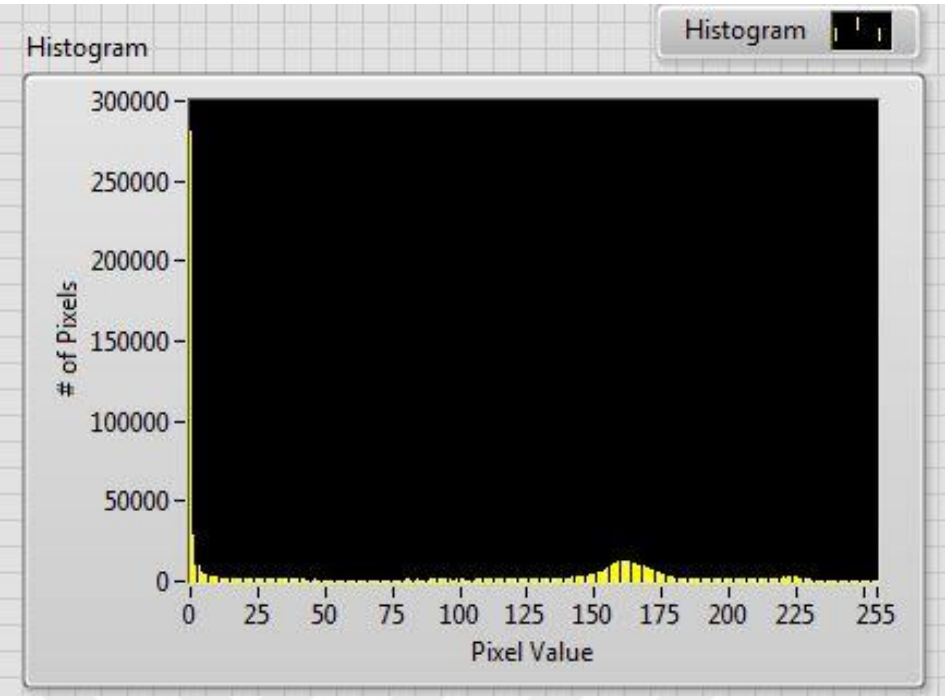

Figure 4.18. Histogram (Front Panel)

4.2. Feature Selection and Extraction

In this stage seven different features are selected and extracted from the suspected mass to classify it later based on these features. From the beginning some steps are done as following and showed in Figure 4.19.

1. 'IMAQ GetFileInfo' returns the image type
2. 'IMAQ Create' creates an image buffer 'image' with the previously determined image type
3. 'IMAQ ReadFile' loads an image from disk into the image buffer
4. 'IMAQ ContructROI' opens a GUI where the user can interactively select a ROI to analyze
5. 'IMAQ Overlay ROI' overlays the selected ROI onto the displayed image
6. 'IMAQ Create' creates an additional image buffer 'mask' for the mask
7. 'IMAQ ROIToMask2' converts the ROI (from step 4.) to a mask and saves the mask to the image buffer 'mask'
8. Since the intensity later calculated 'IMAQ Histogram' needs a grayscale image as input 'IMAQ Cast Image' is used to convert the color image to grayscale
9. Analyze the grayscale image with 'IMAQ Histogram' and select the 'Mean Value' that returns the mean intensity of the selected image section
10. Dispose the image buffer with 'IMAQ Dispose'

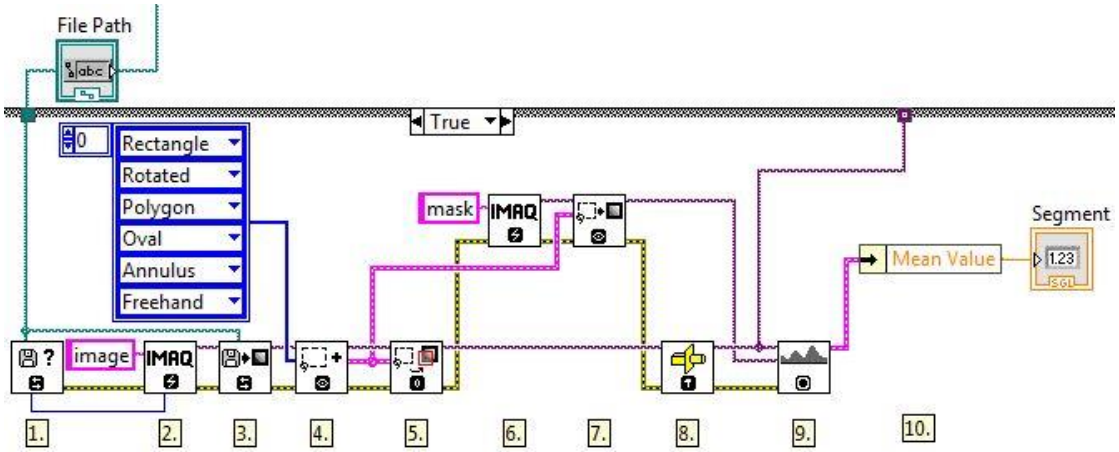

Figure 4.19. Preparation for feature extraction (Block Diagram)

The next step is showing the block diagram of each seven features, Entropy and Uniformity are in Figure 4.20, Skewness is in Figure 4.21, Mean Intensity, Contrast, Standard Deviation and Smoothness is in Figure 4.22.

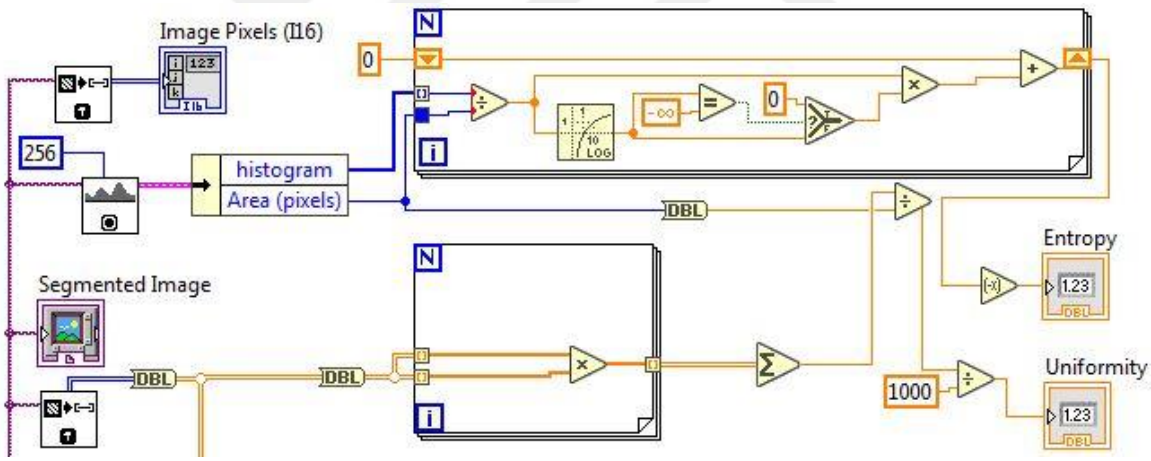

Figure 4.20. Entropy and Uniformity (Block Diagram)

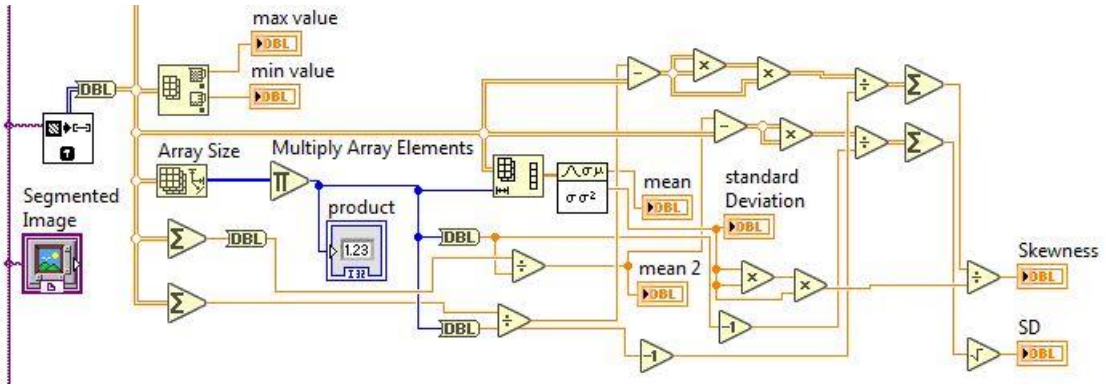

Figure 4.21. Skewness (Block Diagram)

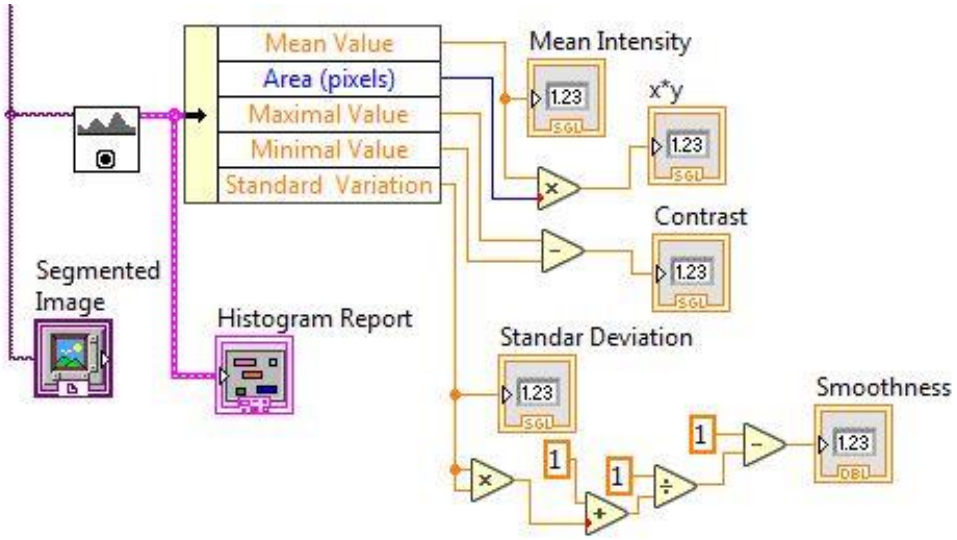

Figure 4.22. Mean Intensity, Contrast, Standard Deviation and Smoothness (Block Diagram)

4.3. Classification

After extracting features from the mass which is the region of interest in this system now it is turn to use these features to determine the mass whether it is a benign mass or malignant mass, in other words the extracted features from the previous stage is used in the last stage which is the classification stage. Artificial Neural Network (ANN) is a powerful tool in classifying to classify the suspected segmented mass into two classes; benign or malignant.

The structure of ANN classifier consists of three main layers: input layer, hidden layers and output layer. In this system the input layer is the extracted features from the previous stage that will be fed into the classifier. The second layer is the hidden layer which defines the number of hidden neurons in the neural network in general. The third and last layer is the result of the classifier that here in this system it will show the segmented mass is benign mass or malignant. Moreover, ANN consists of two main processes; the first process is training data by taking samples it also called the learning stage of process and the second process is testing the remained part of data, in this system the data is the mammography images. The model in the program is shown in figure 4.23.

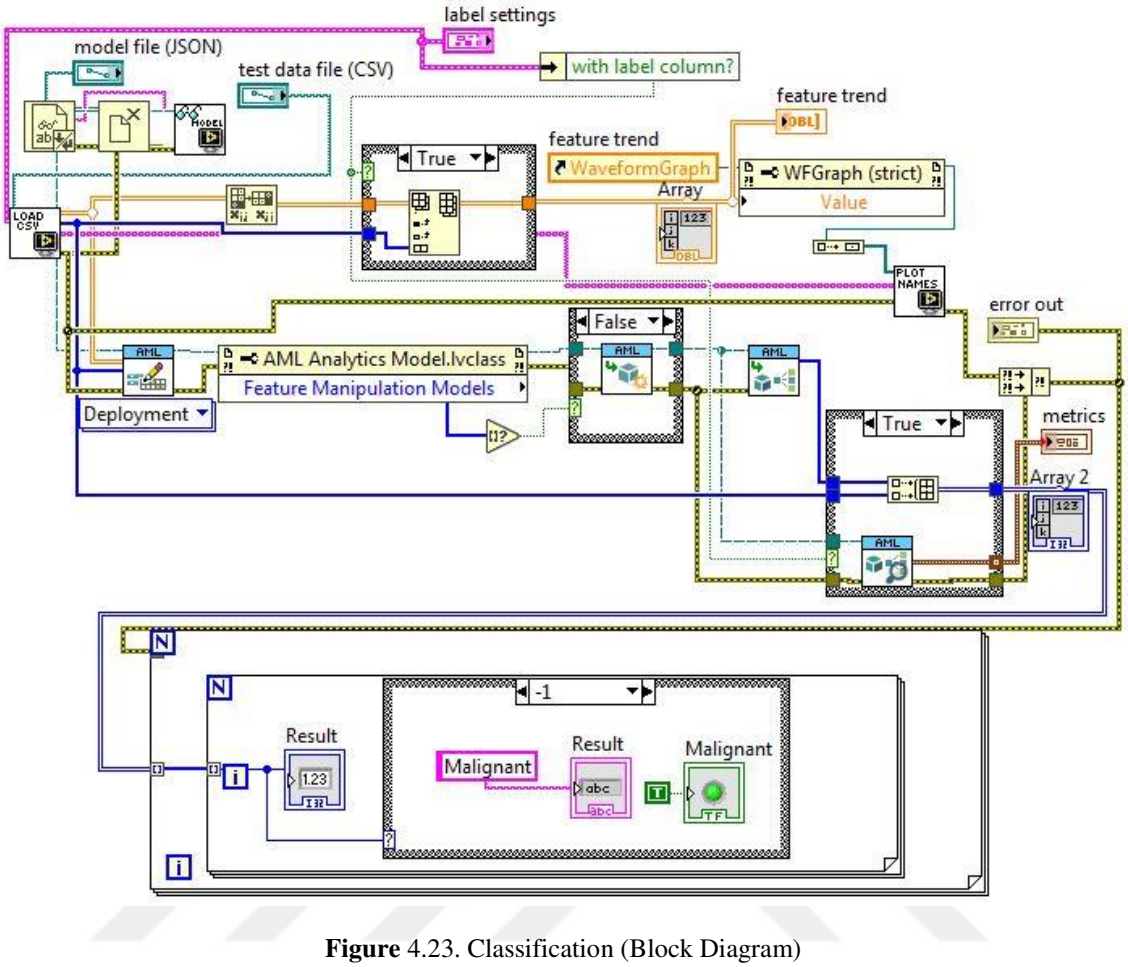

Figure 4.23. Classification (Block Diagram)

In the database, 80 images are chosen to be worked on in this system, 40 benign masses and 40 malignant masses. In the training process of neural network 62.5% of data (50 images; 25benign and 25 malignant samples) are trained into the neural network system as the input, after they are processed in the hidden layers the output supposed to be 1 for benign mass and -1 for malignant mass.

The network parameters are set to: feed-forward neural network is chosen as the type of propagation to train the network with 10 layers of hidden layers and sigmoid activation as hidden layer type because it is a non-linear model and it is suitable for such a system. Non-linear means that the output of neuron is the dot product of some inputs and plus bias and then put into a sigmoid function.

After all the 50 samples are trained into the system, a matrix obtained that called confusion matrix as a measure of performance of the neural network system. Confusion matrix consists of four values which are True Negative (TN), False Negative (FN), False Positive (FP) and True Positive (TP). The following matrix shows the confusion matrix and Table 4.2 defines the meaning of each part of measure in confusion matrix.

$$Confusion\ Matrix = \begin{bmatrix} \text{TN} & FN \\ FP & TP \end{bmatrix} \quad (4.1)$$

Table 4.2. Measured and meaning of each part in confusion matrix

Measures	Meaning
True Negative (TN)	The mass is defined as benign by biopsy and is also classified as benign by the neural network.
False Negative (FN)	The mass is defined as malignant by a biopsy but it is classified as benign by the neural network.
False Positive (FP)	The mass is defined as benign by a biopsy but it is classified as malignant by the neural network.
True positive (TP)	The mass is defined as malignant by a biopsy and is also classified as malignant by the neural network.

In training process, 50% of images (25 benign images) are benign in the matrix which means (TN) and the other 50% of images (25 malignant images) are malignant in the matrix which means (TP). The percentages of (FN) and (FP) are zero. The 62.5% of the samples are trained without any error. The performance of the neural network in the training process is 100%.

The second process of the neural network is testing the mammography images in the classifier to determine whether they are benign masses or malignant masses. 37.5% of the samples (30 images; 15 benign images and 15 malignant images) are tested.

In the result, from the 15 benign masses only one image is misinterpreted and from 15 malignant masses only two images are missed. The overall result is showed in the following matrix:

$$Confusion\ Matrix = \begin{bmatrix} 39 & 2 \\ 1 & 38 \end{bmatrix} \quad (4.2)$$

The confusion matrix shows that 77 mammography images are correctly classified. The 39 mammography images in the upper left are true negative (TN), they are those masses that defined as benign by a biopsy and is also classified as benign by the neural network. The 38 mammography images in the lower right are true positive (TP), they are those masses that defined as malignant by a biopsy and is also classified as malignant by the neural network. However, two mammography images are missed as false negative (FN) and classified as benign by the neural network although they are defined as malignant by a biopsy and another m1ammography image is misinterpreted as false positive (FP) as malignant by the neural network although it is defined as benign by a biopsy. This mistake is happened because the extracted features from these three images are mixed between the features of benign mass and features of malignant mass, in other words; some of features are closed to benign masses but others are closed to malignant masses so the system will be confused to put it in which class.

The missed images are the image with the reference number of (mdb028 - mdb134) that are measured as false negative, the misinterpreted image is the image with the reference number of (mdb218) that is measured as false positive.

It is clear that it is not good for the system to have a high value in FN and FP so in order to have an accurate diagnosis, the values of FN and FP should be small because high value of FN means that malignant masses are missed in the handling process and high value of FP means that benign masses are misinterpreted as cancer.

The obtained confusion matrix indicates some important performance measuring of the neural network classifier. These measures are sensitivity, specificity and accuracy. They can be measured by the following expressions:

$$Sesitivity = \frac{TP}{TP+FN} \tag{4.3}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4.4}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.5}$$

After measuring the sensitivity, specificity and accuracy the results are as the following: The neural network classifier has sensitivity of 95%, specificity of 97.50% and finally the accuracy of the classifier is 96.25%.

Two examples of the system's main GUI (one of them benign case and the other malignant case) are shown in figure 4.24 and figure 4.25.

The result of Extracted features and the result of the classification are showed in the table 4.3 and table 4.4, 40 cases of benign masses and 40 cases of malignant masses are chosen to work on in this thesis.

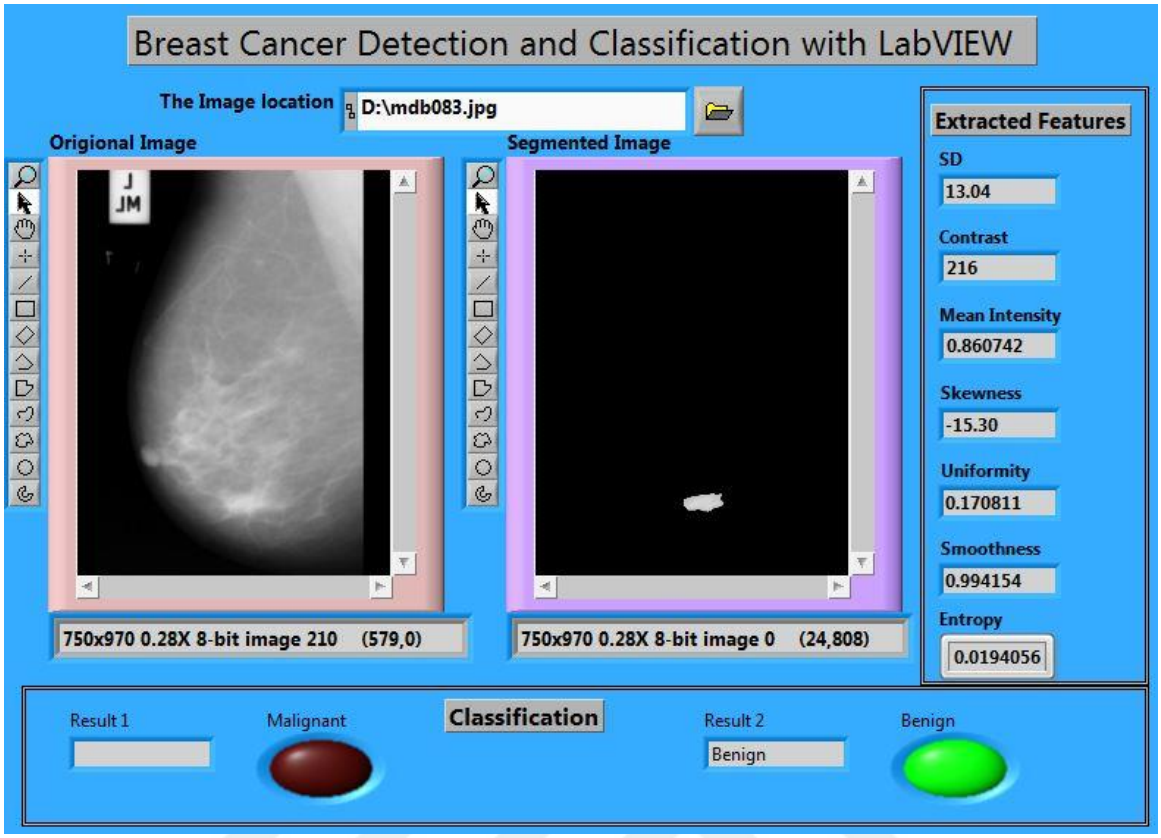

Figure 4.24. Last GUI view of the system / Benign Sample (Front Panel)

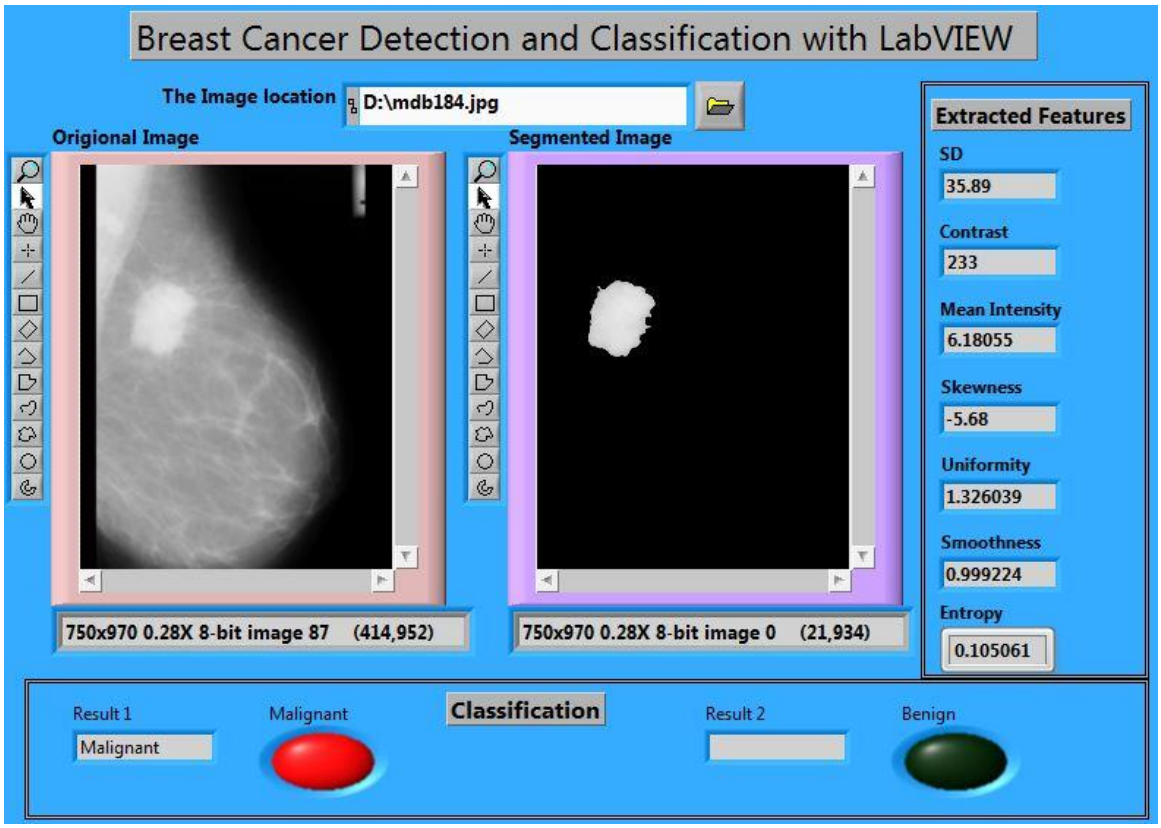

Figure 4.25. Last GUI view of the system / Malignant Sample (Front Panel)

Table 4.3**.** Extracted features from 40 benign images

Image	**Selected and Extracted Features**							**Classification**	
Reference No.	Standard Deviation	Contrast	Mean Intensity	Skewness	Uniformity	Smoothness	Entropy	Original	Estimated
mdb010	27.128	4.51	199	-6.003	0.999	0.756	0.095	B	B
mdb013	45.602	11.325	221	-3.826	1	2.208	0.184	B	B
mdb015	13.488	0.95	205	-14.309	0.995	0.183	0.021	B	B
mdb017	23.049	2.849	204	-8.091	0.998	0.539	0.054	B	B
mdb021	33.892	6.583	203	-5.047	0.999	1.192	0.123	B	B
mdb025	11.156	0.661	198	-17.056	0.992	0.125	0.014	B	B
mdb080	15.194	1.467	186	-10.5	0.996	0.233	0.038	B	B
mdb081	42.052	8.923	229	-4.539	0.999	1.848	0.15	B	B
mdb083	13.041	0.861	216	-15.303	0.994	0.171	0.019	B	B
mdb091	9.262	0.513	179	-18.336	0.988	0.086	0.013	B	B
mdb099	11.778	0.739	207	-16.169	0.993	0.139	0.017	B	B
mdb104	16.755	1.429	209	-11.76	0.996	0.283	0.028	B	B
mdb121	42.736	9.521	228	-4.309	0.999	1.917	0.16	B	B
mdb126	67.676	26.88	223	-2.157	1	5.303	0.388	B	B
mdb127	9.387	0.47	205	-20.301	0.989	0.088	0.011	B	B
mdb132	10.648	0.654	198	-16.518	0.991	0.114	0.018	B	B
mdb133	12.208	0.963	177	-12.954	0.993	0.15	0.026	B	B
mdb145	35.345	6.49	221	-5.323	0.999	1.291	0.115	B	B
mdb150	21.636	2.677	201	-8.047	0.998	0.475	0.058	B	B
mdb160	13.221	0.986	199	-13.638	0.994	0.176	0.024	B	B
mdb165	26.676	3.902	202	-6.739	0.999	0.727	0.073	B	B
mdb175	16.325	1.416	211	-11.636	0.996	0.269	0.03	B	B
mdb193	39.274	7.708	225	-4.936	0.999	1.602	0.123	B	B
mdb195	10.911	0.6	210	-18.427	0.992	0.119	0.013	B	B
mdb198	20.243	2.045	216	-9.908	0.998	0.414	0.039	B	B
mdb199	20.026	2.108	210	-9.522	0.998	0.405	0.043	B	B
mdb204	16.085	1.555	186	-10.473	0.996	0.261	0.037	B	B
mdb207	33.965	6.061	218	-5.479	0.999	1.19	0.107	B	B
mdb218	34.461	6.102	223	-5.529	0.999	1.225	0.109	B	M
mdb219	14.084	1.119	201	-12.807	0.995	0.2	0.027	B	B
mdb222	36.359	6.666	219	-5.318	0.999	1.366	0.107	B	B
mdb227	8.847	0.427	199	-21.162	0.987	0.078	0.01	B	B
mdb236	66.086	24.124	222	-2.404	1	4.949	0.325	B	B
mdb244	57.904	17.46	229	-3.047	1	3.658	0.254	B	B
mdb248	11.496	0.748	190	-15.602	0.992	0.133	0.018	B	B
mdb252	17.052	1.85	179	-9.308	0.997	0.294	0.045	B	B
mdb290	36.166	7.098	209	-4.931	0.999	1.358	0.126	B	B
mdb312	10.629	0.58	212	-18.689	0.991	0.113	0.013	B	B
mdb314	16.506	1.633	194	-10.263	0.996	0.275	0.039	B	B
mdb315	48.575	11.686	229	-3.942	1	2.496	0.176	B	B

Table 4.4. Extracted features from 40 malignant images

Image	**Selected and Extracted Features**							**Classification**	
Reference No.	Standard Deviation	Contrast	Mean Intensity	Skewness	Uniformity	Smoothness	Entropy	Original	Estimated
mdb023	21.716	2.405	217	-9.075	0.998	0.477	0.046	M	M
mdb028	18.364	1.696	214	-10.828	0.997	0.34	0.034	M	B
mdb058	9.163	0.453	208	-20.711	0.988	0.084	0.011	M	M
mdb072	23.955	2.789	226	-8.558	0.998	0.582	0.052	M	M
mdb075	7.099	0.348	227	-21.421	0.981	0.051	0.012	M	M
mdb092	9.541	0.539	184	-17.996	0.989	0.091	0.014	M	M
mdb095	21.397	2.451	212	-8.811	0.998	0.464	0.052	M	M
mdb102	22.438	2.5	221	-8.982	0.998	0.51	0.048	M	M
mdb105	87.503	42.187	241	-1.605	1	9.437	0.515	M	M
mdb110	25.033	3.08	222	-8.075	0.998	0.636	0.056	M	M
mdb111	31.75	4.969	224	-6.283	0.999	1.033	0.084	M	M
mdb120	21.959	2.578	211	-8.555	0.998	0.489	0.052	M	M
mdb125	49.339	13.209	219	-3.512	1	2.609	0.212	M	M
mdb130	28.315	4.045	220	-6.944	0.999	0.818	0.071	M	M
mdb134	13.344	0.953	202	-14.093	0.994	0.179	0.022	M	B
mdb141	25.657	3.995	187	-6.384	0.998	0.674	0.082	M	M
mdb148	56.235	19.102	221	-2.649	1	3.527	0.327	M	M
mdb171	87.781	42.626	239	-1.587	1	9.523	0.512	M	M
mdb178	33.406	6.064	220	-5.408	0.999	1.153	0.117	M	M
mdb179	29.169	3.779	240	-7.736	0.999	0.865	0.05	M	M
mdb181	34.233	7.356	202	-4.516	0.999	1.226	0.152	M	M
mdb184	35.887	6.181	233	-5.681	0.999	1.326	0.105	M	M
mdb202	42.386	10.128	214	-4.008	0.999	1.899	0.18	M	M
mdb206	30.373	5.892	189	-5.026	0.999	0.957	0.129	M	M
mdb209	38.634	8.318	212	-4.476	0.999	1.562	0.155	M	M
mdb211	49.636	14.046	222	-3.293	1	2.661	0.247	M	M
mdb213	28.255	5.216	181	-5.283	0.999	0.826	0.113	M	M
mdb216	83.16	39.91	236	-1.624	1	8.508	0.511	M	M
mdb231	19.055	2.651	161	-7.218	0.997	0.37	0.07	M	M
mdb233	35.512	7.878	202	-4.339	0.999	1.323	0.165	M	M
mdb238	34.318	7.599	188	-4.364	0.999	1.235	0.151	M	M
mdb239	75.957	31.058	233	-2.056	1	6.734	0.379	M	M
mdb241	42.164	9.494	214	-4.247	0.999	1.868	0.152	M	M
mdb245	24.639	3.617	210	-6.81	0.998	0.62	0.079	M	M
mdb249	21.092	2.301	212	-9.17	0.998	0.45	0.046	M	M
mdb253	84.173	43.274	223	-1.448	1	8.958	0.51	M	M
mdb256	53.043	16.436	221	-2.955	1	3.084	0.283	M	M
mdb270	23.252	3.103	231	-7.478	0.998	0.55	0.065	M	M
mdb271	24.133	3.298	215	-7.337	0.998	0.593	0.074	M	M
mdb274	13.378	1.238	166	-10.871	0.994	0.181	0.035	M	M

5. CONCLUSIONS

Breast cancer is the most common form of cancer in the female population in all over the world with different rates in different places and countries. In a lot of countries women were affected and died because of this type of cancer. In this thesis an algorithm is developed to detect suspected masses in mammography images and classify the mass as benign or malignant to help the radiologists for providing an accurate diagnosis and minimize misinterpretation.

Three main stages are performed in general, each stage consists some different steps inside. The three main stages are; image preprocessing and mass segmentation, feature selection and extraction and classification.

The images used in this thesis were from MIAS database, and the program was LabVIEW version2017 with some specified toolkits. The original image loaded into the program and first operation was pruning the image because some of the images have black spaces in the first and last columns that should be cut to avoid the parts out of region of interest. From the kernel family both of smoothing and Gaussian filters applied to the pruned image to enhance the image before thresholding it, the next step was to set a value of threshold to make a binary image, any pixel value higher than the threshold value set to one and lower pixel values set to zero. Morphological operations such as Erosion and Dilation were applied to the binary image to make it more suitable for the next step by sharpening and smoothing the image. The output image of morphological operation multiplied to the original image and the result of multiplication is the preprocessed image. For the second time threshold is applied to the image with a different value to make only the bright parts remain in the image such as; suspected mass, pectoral muscle and in some images the sticker on one of the sides from upper of the image. Two built up blocks in the program is used to remove undesired parts as pectoral muscle and the sticker, reject border block used to remove the pectoral muscle and keep/remove particles block used to remove any remained small particles in the image except of the suspected mass. After removing undesired bright parts, this image multiplied to the preprocessed image to get only the mass, in other words to segment the mass. Applying Gaussian filter in the last step just for smoothen the segmented mass.

After segmenting the mass it is turn to select and extract some features from the mass. Seven features selected, extracted and then fed to the artificial neural network as classifier to classify the mass as benign mass or malignant mass. The seven features were; Standard Deviation, Contrast, Mean Intensity, Skewness, Uniformity, Smoothness and Entropy.

The last step was classifying the masses depending on the extracted features, from 80 images 77 images classified correctly. Only 3 images classified in a wrong way that two of them were from malignant masses and the last one was from benign masses

Form 40 benign images 39 images classified as benign only one image misinterpreted and classified as malignant. From 40 malignant images 38 images classified correctly only two images missed and classified as benign.

In this thesis all trained and tested images are obtained from Mammographic Image Analysis Society (MIAS) Mini Mammographic Database and the system is implemented in Laboratory Virtual Instrument Engineering Workbench (LabVIEW) version 2017 as graphical code-based program. The result of the method obtained from the confusion matrix achieved 95% sensitivity, 97.50% specificity and 96.25% accuracy.

Although, the main difference between this thesis and other works done so far is that graphical code is used to program the system but it still could obtain good results as it is obvious from the percentage of sensitivity, specificity and accuracy.

In future work, the proposed system will be continued and develop more to get better results. Different algorithms can be used in image processing and mass segmentation to get the exact region of interest which is the suspected mass. And different features can be selected and extracted from the segmented mass to feed them into the classifier as input to get a more accurate classification system. In other words, try to increase the values of sensitivity, specificity and accuracy that it causes having a better system of detection and classification for breast cancer. The system can be developed by testing on different and larger database.

REFERENCES

[1] Heber, D. (2006), "*Breast Cancer. In Nutritional Oncology"*, G. L. Blackburn, V. L. W. Go, & J. Milner (Eds.), 2nd edition, ISBN 978-0-12-088393-6, Academic Press. (pp. 393-404).

[2] Rangayyan, R.M. (2005), "*Biomedical Image Analysis",* Biomedical Engineering Series. CRC Press LLC, ISBN 0-8493-9695-6

[3] American College of Radiology (ACR) (2003), *"ACR Breast Imaging and Reporting Data System"*, Fifth edition, Reston VA, American College of Radiology.

[4] Jalalian, A. et. al. (2013), *"Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review",* Clinical Imaging, 37(3), 420-426.

[5] Kopans, D.B. (2007), "*Breast imaging*", Lippincott Williams & Wilkins, 3rd edition, ISBN 978-0781747684

[6] Smith, R.A. et. al. (2003), *"American Cancer Society guidelines for breast cancer screening", CA:* a cancer journal for clinicians, 53(3), 141-169.

[7] DeSantis, C. et. al. (2013), *"Breast Cancer Facts & Figures 2013-2014",* American Cancer Society, 1-38, Inc.

[8] American Joint Committee on Cancer (2010), *"Breast Cancer Staging",* 7th edition. New York: Springer; 347–369.

[9] Tang, J. et. al (2009), *"Computer-aided detection and diagnosis of breast cancer with mammography",* recent advances. Information Technology in Biomedicine, IEEE Transactions on, 13(2), 236-251.

[10] Rafferty, EA. et. al. (2006), *"Breast Tomosynthesis: One View or Two?"* Presented at RSNA, Session SSG01-04 Breast Imaging (digital tomosynthesis).

[11] American Cancer Society (2016), *"Breast Cancer Facts & Figures 2016"*, Atlanta: American Cancer Society, Inc.

[12] Sankar, D., and Thomas, T., (2010). *"A New Fast Fractal Modeling Approach for the Detection of Microcalcifications in Mammograms",* J Digit Imaging. 23(5): 538-46.

[13] Gunderman, R.B., (2006). Essential radiology. Clinical presentation, pathophysiology, Germany.

[14] Li, H., Liu, R., and Lo, S. (1997). *"Fractal Modeling and Segmentation for the Enhancement of Microcalcifications in Digital Mammograms",* IEEE Trans on Medical Imaging, 16(6): 785-798.

[15] Akay, M., (2006). *Automatic mass segmentation in mammographic images*, Universitat de Girona,Spain.

[16] Kaushak, C., (2007). *"Breast Mammography: Pictorial Review"*, Indian Journal for the Practising Doctor, 4(1): 3-4.

[17] Mini,G., and Thomas, T., (2003). *"A neural network method for mammogram analysis based on statistical features"*, Convergent Technologies for the Asia-Pacific Region Conference , In TENCON 2003 IEEE.

[18] Rodrigues, D.L., (2008). *"An Application of Hough Transform to Identify Breast Cancer in Images"*, VIPimage, 55 (31): 3319-4910.

[19] Medical Image Understanding and Analysis 2007, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.692.2635&rep=rep1&type=pdf#page=234 ,20 July 2016.

[20] American Cancer Society (2014), *"Mammograms and Other Breast Imaging Tests"*, Atlanta: American Cancer Society, Inc.

[21] Ball, J.E. and Bruce, L.M. (2007), *"Digital mammogram spiculated mass detection and spicule segmentation using level sets",* In Engineering in Medicine and Biology Society, EMBS 2007. 29th Annual International Conference of the IEEE (pp. 4979-4984). IEEE.

[22] Ramos, RM. et. al. (2015), *"Well-circumscribed breast carcinoma. Keys to face the challenge of malignant tumors with a benign appearance",* European Congress of Radiology.

[23] Ayres, F. J. and Rangayyan, R.M. (2005), *"Characterization of architectural distortion in mammograms",* Engineering in Medicine and Biology Magazine, IEEE, 24(1), 59-67.

[24] Jumaat, A.K. et. al. (2010), *"Segmentation of masses from breast ultrasound images using parametric active contour algorithm"*, Procedia-Social and Behavioral Sciences, 8, 640-647. Elsevier

[25] Rahmati, P. et. al. (2012), *"Mammography segmentation with maximum likelihood active contours"*, Medical image analysis, 16(6), 1167-1186. Elsevier

[26] Shi, X. et. al. (2010), *"Detection and classification of masses in breast ultrasound images",* Digital Signal Processing, 20(3), 824-836. Elsevier

[27] Yao, Y. (2004), *"Segmentation of breast cancer mass in mammograms and detection using Magnetic Resonance Imaging",* IEEE Image Processing Soc. J.

[28] Kom, G. et. al. (2007), *"Automated detection of masses in mammograms by local adaptive thresholding",* Computers in Biology and Medicine, *37*(1), 37-48. Elsevier

[29] Varela, C. et. al. (2007), *"Computerized Detection of Breast Masses in Digitized Mammograms"*, Computers in Biology and Medicine 37(2), 214–226. Elsevier

[30] Nguyen, A. et. al. (2009), *"Characterizing image properties for digital mammograms",* In Proceedings of the Third Australasian Workshop on Health Informatics and Knowledge Management-Volume 97 (pp. 19-24). Australian Computer Society, Inc.

[31] Dominguez, A.R. and Nandi, A.K. (2008), *"Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection",* Computerized Medical Imaging and Graphics, *32*(4), 304-315. Elsevier

[32] Zou, F. et. al. (2008), *"Gradient vector flow field and mass region extraction in digital mammograms"*, In Computer-Based Medical Systems, CBMS'08. 21st IEEE International Symposium on (pp. 41-43). IEEE.

[33] Cheng, H.D. et. al. (2006), *"Approaches for automated detection and classification of masses in mammograms",* Pattern recognition, *39*(4), 646-668. Elsevier

[34] Mencattini, A. et. al. (2008), *"Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing",* Instrumentation and Measurement, IEEE Transactions on, 57(7), 1422-1430.

[35] Gonzalez, R.C. and Woods, R.E. (2002), *"Digital Image Processing",* Prentice-Hall, Inc., 2nd edition, ISBN 0-201-18075-8. International Edition.

[36] Rangayyan, R.M. et. al. (2006), *"Feature extraction from the turning angle function for the classification of contours of breast tumors",* In IEEE Special Topic Symposium on Information Technology in Biomedicine, (Vol. 4).

[37] Yuan, Y. et. al. (2008), *"Correlative feature analysis of FFDM images",* In Medical Imaging (pp. 69151L-69151L). International Society for Optics and Photonics.

[38] Bellotti, R. et. al. (2006), *"A Completely Automated CAD System for Mass Detection in a Large Mammographic Database"*, Medical Physics 33(8), 3066–3075.

[39] Timp, S. et. al. (2007), *"Temporal change analysis for characterization of mass lesions in mammography",* Medical Imaging, IEEE Transactions on, 26(7), 945-953.

[40] Cascio, D. et. al. (2006), *"Mammogram segmentation by contour searching and mass lesions classification with neural network",* Nuclear Science, IEEE Transactions on, 53(5), 2827-2833.

[41] Oliver, A. et. al. (2008), *"A novel breast tissue density classification methodology",* Information Technology in Biomedicine, IEEE Transactions on, 12(1), 55-65.

[42] Chen, C.M. et. al. (2003), *"Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks"*, Radiology 226(2):504-514.

[43] Suckling, J. and Parker, J. (1994), *"The mammographic image analysis society digital mammogram database",* In: Exerpta Medica, International Congress Series, vol. 1069. p. 375–378.

[44] Song, J.H. et. al. (2005), *"Artificial neural network to aid differentiation of malignant and benign breast masses by ultrasound imaging",* Proceedings of SPIE, p. 2957-2960.

[45] Huang, Y.L. and Chen, D.R. (2005), *"Support vector machines in sonography: application to decision making in the diagnosis of breast cancer",* Clinical Imaging; 29(3):179-184

CURRICULUM VITAE

Iman HAMADAMIN

[redacted]