REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

AUTOMATIC SEGMENTATION OF BREAST CANCER ON MAMMOGRAM IMAGES USING IMAGE PROCESSING TECHNIQUES

SAFWET ALAN ALDAWERI

M.Sc. THESIS

DEPARTMENT OF COMPUTER ENGINEERING
PROGRAM OF SYSTEMS ENGINEERING

ADVISER
ASSIST.PROF. DR.GÖRKEM SERBES

İSTANBUL, 2016

REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

Automatic Segmentation of Breast Cancer on Mammogram Image Using Image Processing Techniques

A thesis submitted by SAFWET ALDAWERI in partial fulfillment of the requirements for the degree of **MASTER OF COMPUTER ENGINEERING** is approved by the committee on 20/11/2016 in Department of Computer Engineering.

Thesis Advisor

Asst. Prof. Dr. Gorkem SERBES

Yıldız Technical University

Approved By the Examining Committee

Asst. Prof. Dr. Gorkem SERBES

Yıldız Technical University ______________________

Prof. Dr. Nizamettin AYDIN, Member

Yıldız Technical University ______________________

Asst. Prof. Dr. Cemal OKAN SAKAR, Member

Bahcesehir University ______________________

ACKNOWLEDGEMENTS

First, I would like to thank God for giving me strength and patience to work on this thesis. It would not have been possible without God's guidance and support.

I would also like to thank Dr.Gorkem SERBES for his supervision and valuable participation, discussions and contributions towards my project. Their input and advice have steered me in the right direction for reaching a successful results.

Here I must acknowledge the Ministry of Higher Education in the Republic of Iraq for their great help and financial support towards my Master studies.

Finally, my special and profound thanks go towards my beloved father Mr. ALAN ALI and my mother Mrs. RAJWA HASAN for providing me with their encouragement, unconditional support and wishes to achieve success in my higher studies. This accomplishment would not have been possible without their great help and understanding. Last not least my great thanks to my brothers (NASARAT, RAFAT) and them family and for sure my only beloved sister (DALAL) and her family also.

November 2016

Safwet Alan ALDAWERI

TABLE OF CONTENTS

Page

LIST OF SYMBOLS

†in	Input image
†out	Output image
C	Weight of high pass filter
†lp†	Low pass filter
†hp†	High pass filter
⊕	Dilation
⊖	Erosion
○	Opening
•	Closing

LIST OF ABBREVIATIONS

ARCH	Architectural distortion
ASYM	Asymmetry
BIRADS	Breast Imaging Reporting and Data System
BW	Black And White
CAD	Computer Aided Detection
CADx	Computer Aided Diagnosis
CALC	Calcification
CIRC	Well-defined/circumscribed masses
EM	Expectation Maximization
FCM	Fuzzy C-Mean
FDA	Food and Drug Administration
MIAS	Mammogram Images Analysis Society
MISC	Other, ill-defined masses
MPM	Maximizer of Posterior Margins
NORM	Normal
SPIC	Speculated masses
SVM	Support Vector Machine

LIST OF FIGURES

Page

LIST OF TABLES

ABSTRACT

AUTOMATIC SEGMENTATION OF BREAST CANCER ON MAMMOGRAM IMAGES USING IMAGE PROCESSING TECHNIQUES

Safwet ALDAWERI

Department of Computer Engineering

MSc. Thesis

Advisor Asst. Prof. Dr. Görkem SERBES

There are several types of cancers affecting a large number of people around the world. The main goal of this study is to work on one type of cancer common for a significant percentage of women, that is breast cancer. This disease claims the life of so many women every year, yet early diagnosis of this type of cancer can be quite effective in offering the appropriate medical treatment in a timely fashion. Hence, it could be a very useful tool in helping saving more women's lives.

Specialists usually use mammography as a method for examining and detecting breast cancer at its earlier stage. This diagnostic technique often employed to detect different lesions of the disease including calcifications, well defined (circumscribed) masses, Spiculated masses, and so on. Recent development of science and technology has led to the use of Computer Aided Detection (CAD) as a part of diagnostic procedures that will lead to increasing success rates in early diagnosis of the disease. CAD systems will be advantageous in also decreasing the number of misdiagnosed cases. In order to produce diagnostic information, CAD systems first acquire images using appropriate modalities and then use sophisticated image analysis algorithms.

In this thesis, we first give a broad background on breast cancer and discuss its influence on the society. The main target of this master research study is to investigate the automatic enhancement and segmentation of lesions in mammographic images. Towards this goal, we first describe and discuss implementation and employment of ‘un-sharp masking' as an image enhancement technique. Then, we introduce image segmentation techniques based on morphological operations. Finally, we apply these techniques on mammogram images and show/discuss our results.

All experimental work were performed on a set of mammograms with different breast densities obtained from mini Mammogram Images Analysis Society (MIAS Database) and the procedures were developed in MATLAB R2013 environment.

Key words: Breast Cancer; Image processing; Mammography; Medical imaging; MIAS; CAD

ÖZET

Mamografi Görüntülerinden Görüntüİşleme Tekniklerini Kullanılarak Göğüs Kanserinin Otomatik Tespiti

Safwat Alan ALDAWERI

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı: Yrd. Doç. Dr. Görkem SERBES

Dünya üzerinde geniş insan kitlelerini etkileyen çok sayıda kanser türü vardır. Bu tezin amacı, kadınların önemli bir kısmında görülen Göğüs Kanseri üzerinde çalışmaktır. Bu hastalık her yıl çok fazla sayıda kadının hayatına mal olmakla birlikte, erken teşhis doğru tedavinin uygulanmasında yüksek düzeyde etkili olmakta ve bu sayede birçok kadının hayatı kurtulabilmektedir.

Uzmanlar hastalığın erken evresinde tanı ve muayene amaçlı olarak çoğunlukla Mamografi yöntemini kullanırlar. Bu tanı yöntemi, çoğunlukla kalsifikasyonlar da dâhil hastalığın farklı değişimlerini (lezyonlarını) tespit etmek için kullanılmaktadır. Son dönemde bilim ve teknoloji alanındaki gelişmeler, mamografi sürecinde Bilgisayar Destekli Tanılamanın (BDT) önünü açmıştır ki bu da bu hastalığın erken teşhisindeki başarı oranını arttırmaktadır. BDT kullanımını ölümlü vaka sayısının azalmasına ve kadınların hayatının kurtulmasına yardımcı olmaktadır. BDT uygulama süreci bilgisayarlı algoritmaların toplanması ve kullanılmasını, yani doku bozukluklarının (lezyonların) otomatik görüntü analizi yoluyla işlenmesini içermektedir.

Bu yüksek lisans tezinin amacı, mamografik görüntülerdeki kitlenin otomatik netleştirilmesi ve segmentasyonunu incelemektir. Tez kapsamsında öncelikle göğüs kanserinin arka planı ve birçok kadının hayatını ne şekilde etkilediği üzerinde durulmuştur. Bu amaçla, ilk olarak "keskin olmayan maskeleme" gibi görüntü iyileştirme tekniklerinin uygulaması ve kullanılması tartışılmıştır. Daha sonra, biçimsel operasyonlar kullanılarak gerçekleştirilen görüntü segmentasyon tekniklerinden bahsedilmiştir. Son olarak ise bu yöntemler mamografi görüntüleri üzerinde denenmiş ve sonuçlar değerlendirilmiştir. Tüm deneyler mini Mamografi Görüntü Analiz Topluluğu veri-tabanından alınmış farklı yoğunluktaki mamografi görüntüleri üzerinde gerçekleştirilmiş ve tüm algoritmalar MATLAB R2013 platformunda gerçeklenmiştir.

Anahtar Kelimeler:Göğüs Kanseri; Görüntü İşleme; Mamografi; Tıbbi Görüntüleme; MIAS; BDT.

CHAPTER 1

INTRODUCTION

1.1 Literature Review

Significant work and research about detecting the tumor or cancer abnormality in breast tissue done at the follow report:

A work by [Sheshadri HS, Kandaswamy] [1] for image segmentation using MPM (Maximizer of the posterior margins), the base of segmentation on tissue feature would classify the breast texture under various classes. These algorithms evaluate the region properties of the mammogram images and the process would classify the image into important segment.

Other work by [B.MonicaJenefer &V.Cyrilraj] [2] the mammogram enhancement can be gain by removing the noise and make efficient the quality of the image using speckle noise removal and EM algorithm respectively. The most recognized partition technique exercises are Modified Watershed Segmentation process. The features are extract from the segmented tumor region and classify the regions utilizing the Support Vector Machine SVM classifier.

Also a work by [S.SahebBasha & 2dr.K.Satya Prasad] [3]using Morphological operators and finally fuzzy c- means clustering (FCM) algorithm has been performed for intensity – based segmentation.

A work by [Nalini Singh &Ambarish G Mohapatra&GurukalyanKanungo] [4] this study that shows the outcome of applying image-processing threshold, edge based and watershed segmentation in breast cancer on mammogram image and as well presents a case research between them based on time consumption and simplicity.

Another works by [Mussarat Yasmin, Muhammad Sharif and Sajjad Mohsin] [5]. This research work is an attempt to highlight the obtainable breast cancer detect methods based on image processing and supply an overview on the affordability, reliability and outcomes of each technique.

At last, work by [Célia Freitas da Cruz] [6] the main goal of this thesis dissertation match to the automatic for both enhancement and segmentation of microcalcifications in mammographic images. This dissertation includes implementation and application of automatic image cropping; image enhancement methods such as unsharp masking, adaptive unsharp masking, with the evaluation of multi different parameters. The techniques evaluated with emphasis on mass enhancement on real mammographic images, where the adaptive techniques had better performance. Image segmentation techniques also implemented and applied, such as Morphological operation. The segmentation technique edge detectors and regions growing of selected areas had higher sensitivity, while edge detection and threshold and difference of Gaussians had higher accuracy, precision and F-measure. A dataset additionally created with the features extracted of the segmented objects and preliminary classification studies performed.

1.2 Objective of the Thesis

The main aim of writing this thesis is to diagnose breast cancer via mammogram image automatically helping radiologist in detection and classification of breast cancer in accordance with image processing techniques as well as new advances in this field of science. This system helped to speed up the discovery of the disease in addition to reducing the cost and not exposed the patient to take a biopsy from his body as possible to hurt him.

1.3 Hypothesis

There are many methods of detecting the mass region of breast cancer in mammography images. This research study proposed new method that combines different image processing techniques to extract mass regions. These techniques are pre-processing images using crop image depend on the information that given in the dataset, image enhancements by using two techniques (unsharpmask and adjust image), image segmentations using morphological operations, and feature extraction using regionprop technique.

CHAPTER 2

GENERAL INFORMATION

2.1 Overview

Serious concerns about the numbers of deaths associated with cancer in general exist. Even though scientists and researched are making effort to find efficient medicine to cure this disease, there is remarkable break through to analyze the main reasons that cause it to happen. Statistics imply that breast cancer affects one in every eight women. Reports also claim that men are also diagnosed breast cancer. However, high percentage of this disease remains within the women population. A malignant tumor, which initiates in cells of the breast, and at later stages, spreads to other tissues in the affected area. The yearly number of women who are embodying with breast cancer is more than 50000 in the UK. The number of deaths caused by breast cancer is also similar in the United States every year [7].

Efforts are underway by Scientists and Medical researchers to investigate the development of appropriate medical and diagnostic treatments for women affected by the disease. These efforts go back as far as the 16th century. Their focus is mostly on understanding more about the physics of why and how it initiates in human body with the purpose to make advances in the development of medicines that could cure this deadly disease.Unlike any other diseases, it is advantageous when the symptoms of breast cancer diagnosed and detected at the very early stages. Early diagnose with appropriate medical treatment and help can save the lives of tens among thousands of women every year. Recently, information technology is widely used as part of several methods to detect breast cancer. Medical Image Processing developed and it is now widespread to help specialist in detecting this disease. Patient of breast cancer can take advantage of this technology by consulting their doctors as soon as they suspect unusual growth of a tumor any other symptoms of this disease.

Detecting early symptoms for breast cancer by using medical imaging of the affected area will require efficient and suitable image processing equipment and software as well as a professional staff running this advanced equipment. What is more, these parameters will in no doubt increase the success rate of detecting the disease at its early stage causing an increase in the chances of eliminating the disease from the patients.

This research study outlines and provides an insight for the available different breast cancer detect methods together with their numerous impacting parameters. Each technique is unique discussion on its nature and targets are present including a special kind of scenario. The details, pros, and cons for each methodology are also present in the next sections [8].

2.2 Breast Cancer

The types of pathologies, which can affect the breast, are different. However, it has been stated by studies that the imageology of the breast is the main element, which contributes to the formation of breast cancer [9]. Although the causes of breast cancer are to some cases remain unclear, the disease is frequently explain by an un-controlled malignant growth. The growth typically initiates in the cells within the breast tissues, which in a healthy and normal situation can be governed and controlled to a specific order without causing any problem. In the case of breast cancer, it is impossible to control the regeneration and growth of the cells. Thus, as the reparation process does not function, eventual mutations fails leading the creation of a cancer tumor.

Once a cancer tumor formed, an increase in the size of tumor follows as well as all related side effects for the patient. The stage of tumor evolution varies among patients. Other health parameters play an important role for each individual person. It is mostly agreed the key factor of controlling the spread of breast cancer disease is early detection and diagnose, as soon as the patient notices any unusual growth in the breast areas. This will also help in drawing on the best and appropriate medical treatment to terminate the disease from becoming a terminal one and therefore increasing the opportunities to save lives. Failure to seek medical advice at early breast cancer stages will encourage it to spread through the body, damaging more tissues and organs. Unfortunately, statistics shows that more patients who died by breast cancer because of not taking medical help towards the early detection and diagnose. Recovery is very slim for those patients who tested positive for cancer after it stayed in their body un-detected. The earlier that disease is detected the better opportunity of survival can be achieved. In addition, there are several other parameters, which also can affect how breast cancer influences people's life. Family history seen as one good possibility of inheriting breast cancer. More elements, for example rotundity, early menarche, late menopause, null parity and chest radiation exposition, abnormal cells in fibrocystic disease and hormone standing by therapy, can play an important role in the formation of breast cancer [9] [5]. Research work conducted by Gunderman has claimed that patients who are over 50 years of age are more prone to have high risks of getting this type of cancer.

The importance of detection of breast cancer was highly emphasized earlier amongst women. Good examples are followed within the developed and developing countries in order to place screening tests and women at the age of 40 and over are encouraged to use the advantage of this scheme to enhance the diagnose at the earliest potential time when the disease is at its early stage.

2.2.1 Breast Cancer Lesions

There are several characteristics associated with breast cancer. These include micro-calcifications, masses, and architectural distortions. In addition, any asymmetry between breasts can also be applies good indicator during the diagnoses of breast cancer disease.

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 2.1. 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 [11].

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 [12]. 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 [13] Figure 2.2.

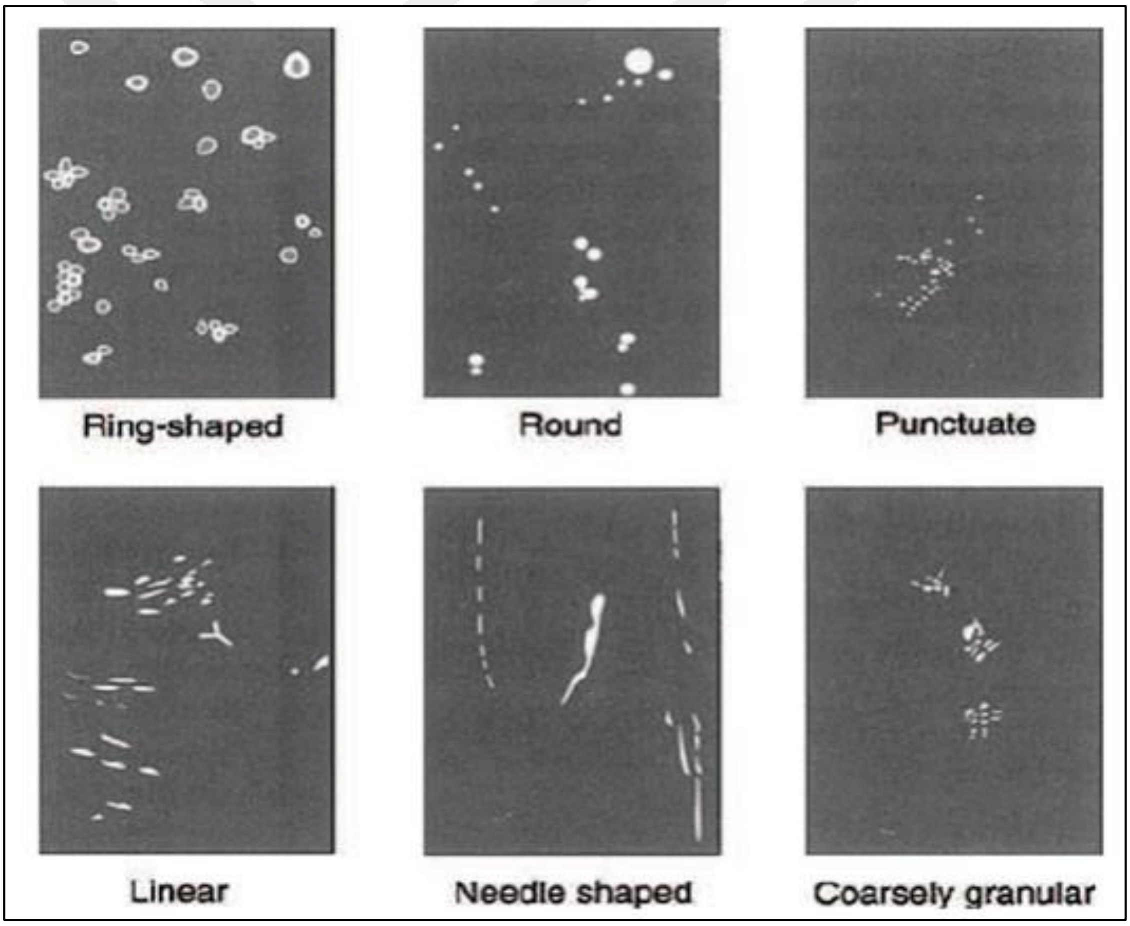


Figure 2.1 – Type of microcalcifications commonly seen on mammographic images [9].

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 [13], Figure 2.3 and Figure 2.4:

- 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.

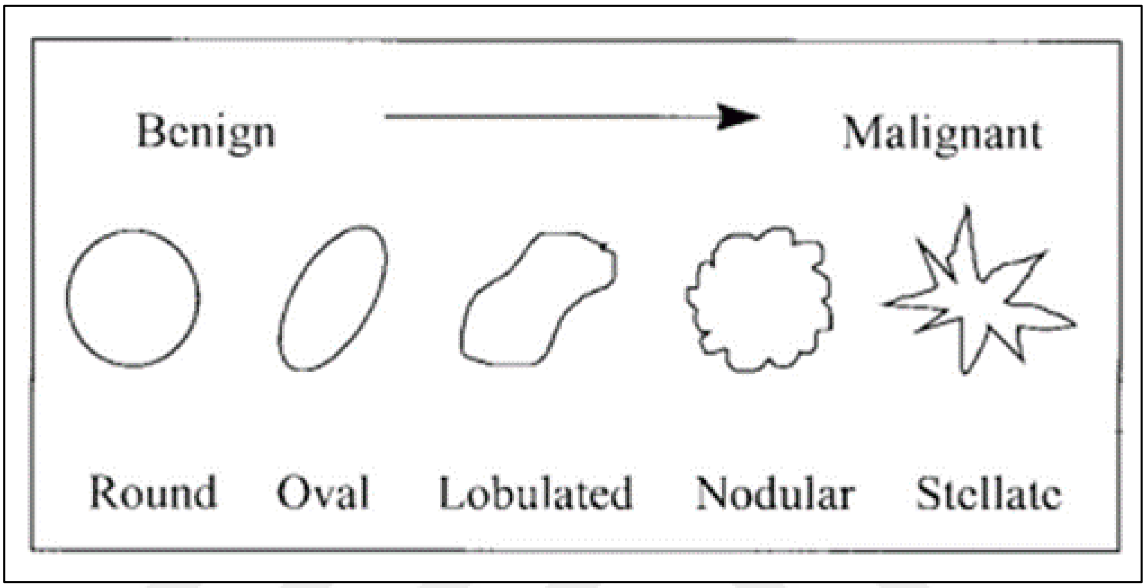


Figure 2.2 – Morphologic spectrum of mammographic masses [14]

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 [13]. 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 [15]. Some masses can incorporate microcalcifications, as in Figure 2.5.

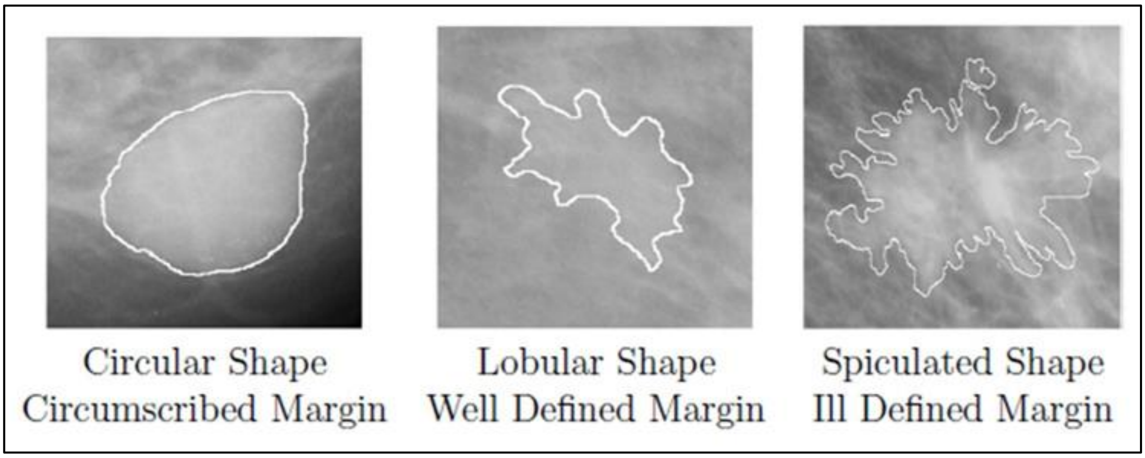


Figure 2.3– Mass examples with different shapes and borders [16]

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 [15].

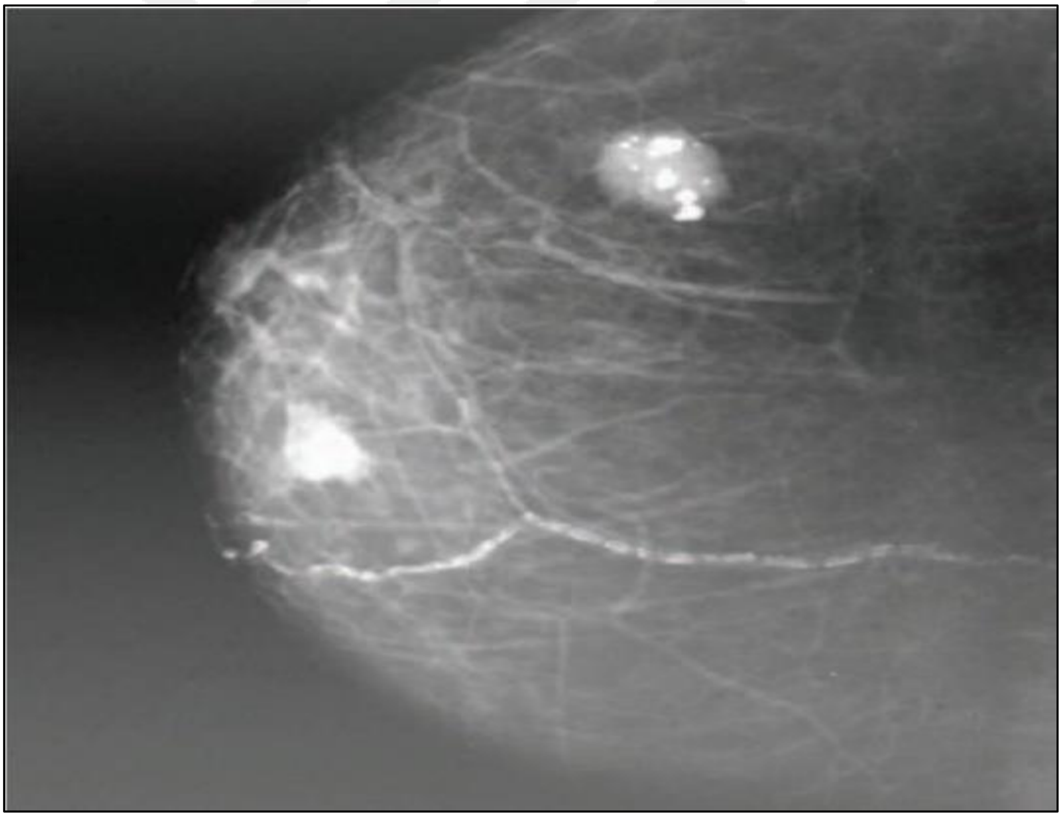


Figure 2.4– A craniocaudal view of the right breast demonstrates benign vascular calcifications as well as two well-circumscribed masses containing “popcorn” calcifications classic for involution fibro adenomas [9]

2.2.2 The types of Breast Cancer

There several criteria for identifying the different types belong to breast cancer. Breast cancer could be named after the breast tissue where the cancer was originated (glands, ducts, fat tissue or connective tissue). Another criteria for distinguishing the types is the extent of the cancer spread (non-invasive/in situ or invasive/infiltrating) [9].

Eastman, Wald, &Crossin [17] claimed that carcinoma in situ tumor is an early form of carcinoma (invasive malignant tumor due to muted epithelial cells) detected in an early stage and with the absence of invasion in surrounding tissues. After the initiation stage of cancer, another phase takes place through the spread of the disease to other UN affected areas. An infiltrating cancer begins in the glands or ducts, spread to healthy surrounding tissue. Usually, this type of cancer can have a variety of appearances. Indeed, the severity of its spread can be different amongst different patients.

Both in situ and infiltrating cancers can be ductal and lobular, depending on the breast cancer location. Ductal carcinoma arises from the epithelial cells that line the breast milk ducts. In the ductal carcinoma in situ, cancer cells have not permeated the basement membrane of the ducts. In the mammographic images is characterized by fine microcalcifications; however, the degree of cancer infiltration is not generally visible [3] Studies have reported that the infiltrating ductal carcinoma is the general frequent breast cancer types, being the cause of almost 80% of cases. A tumor irregular mass is characteristic in the mammography of this type of cancer. Another type of breast cancer known as lobular carcinoma initiates in the milk glands, in the terminal lobules. Approximately, 10% of breast cancer is lobular carcinoma [6]. The lobular carcinoma in situ and therefore is hardly detected in mammography.

Figure 2.5– Invasive Ductal Carcinoma showing micro lobulated borders and microcalcifications [16].

Once cancer established within the breast area, it wills then starts to infiltrate and spread to other organs. When cancer spreads for other parts into the body through blood and lymph circulation, it called metastization. When the ductal carcinoma invades the skin for the nipple, it called Paget’s disease.

Inflammatory breast cancer corresponds to an aggressive tumor that invaded the dermal lymphatic’s [9] representing about 1 to 4% of the breast cancer. This cancer usually presents breast inflammation. Medullary breast carcinoma arises from the stromal cells of the breast [9]. Mucinous carcinoma is associated with large amounts of cytoplasmic mucin [9] the last two types of cancer generally experience lower ability to create metastasis than the ductal and lobular.

2.3 Mammography

For the detection of the disease, X-rays are use as diagnostic tool in mammography for the examination of human breast. These examinations are record as specialized images, which are then observe by radiologists for any possible abnormality. In the following sections, few techniques are discussed that use mammography for early detection of breast cancer. The mammogram is the most efficient and cheapest tools used to detect clinically occult illness, being the only image-based method recommended for breast cancer screening [18] Mammography helped in great way of reduce the breast cancer mortality in a well-organized screening program over the population, being the breast cancer detection technique that most reduces mortality [17]. If the density of the breast increases, the performance of the mammography decreases. This situation is inconvenient since breast cancer risk increases as the breast density increases [19].

2.4 Abnormal Mass Identification in Digital Mammography Images

Although mammography cannot detect every kind of breast cancer, it remains as the world widely used medical examination technique for breast cancer detection due to its inexpensiveness and low complexity. Statistics show that mammography detects around 80% to 90% of breast cancers [20]. Masses or abnormalities detection at early stage is quite possible with the usage of mammography. Mammography is use as a primary tool for detecting breast cancer [21]. In this method of detection, the breast area is extract as an image and processed before printing on the film for better visualization of size, location and angle of the mass. These optimized images are then observe by radiologist for detection of possible abnormalities. These observations are specific to patients and vary among them [22]. This analysis based on many existing methods [23]. The proposed technique identifies abnormal masses in mammography image. The algorithm will only identify abnormal masses to ease further investigation. The mammogram images used taken from mini (MIAS) database [24]. These images in the map are format of eight- bit having 256 grey levels. All types of breast tissues are considered. This research has also focused on calcification, circumscribed, speculated and other ill-defined masses.

2.5 Preprocessing

It well knows that mammogram prone images are difficult to interpret due to their complexity. Therefore, these images are process in a way so that they can be further use for segmentation. Preprocessing includes removal unwanted or irrelevant areas and to make prominent the area of interest by increasing the contrast. This done by setting a certain threshold value

2.6 Computer Aided Detection

Studies have concluded that correct mammographic detection of asymptomatic lesions is essential to discover early breast cancer phases, and thus increasing the treatment options and survival rate [25]. This is quite essential since significant efforts made to develop better understanding for the method of detecting breast cancer especially amongst women.

To properly detection mammogram lesions, radiologists may double read the exams as distinct readers miss different cancers [26]. However, less costly in man terms, would be the improvement of the performance of individual readers, as the double reading stops being required. Scientist and researched working in the field of breast cancer have taken advantageous by the development of using the computer to assist in the process of detecting the disease. In this process, software may be an important assistance [27].

Computer aided detection were focused over the recent years as an effective tool to make it better for correct detection of abnormalities in the breast. Computer aided diagnosis and computed aided detection, commonly shortened as CADx and CAD respectively, can be determined as disclosure and/or diagnosis made by the radiologist considering the outcome of a computing algorithm which identify lesions over automatic image anatomy [28], [29]. Rather than exchange, the human diagnosis. CAD systems as "second opinion" used to aid radiologists to set the lesions. This permits the variability reduction's in the radiologists' mammograms interpretation and the frequency of errors by affirm that doubtful regions are consult and increasing the impact of subtle markers, which may be discarded otherwise [30]. In short, it could understand as a method to filter through the results of a typical x-ray image so that the prime focus will be on detecting the damaged tissue or zones within the breast area.

In general, the use of CAD is supposed to follow the subsequent steps [31]: as in the following:

- Initial radiologist mammography reading, identifying suspicious areas;
- A CAD system scanning to detect any suspicious features;
- Radiologists' analysis of the prompts given by the CAD system and verification if the suspicious areas left unchecked in the first reading

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 [32]:

- 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.

Figure 2.6 shows example mammograms of each class (the mammograms extracted from the MIAS database). Note how the internal density of the breasts increases from BIRADS I (left) to BIRADS IV (right). It should note that besides density these BIRADS classes also included patterns that can be describe as various textures. As such, it seems appropriate to include both aspects in an automatic classification approach.

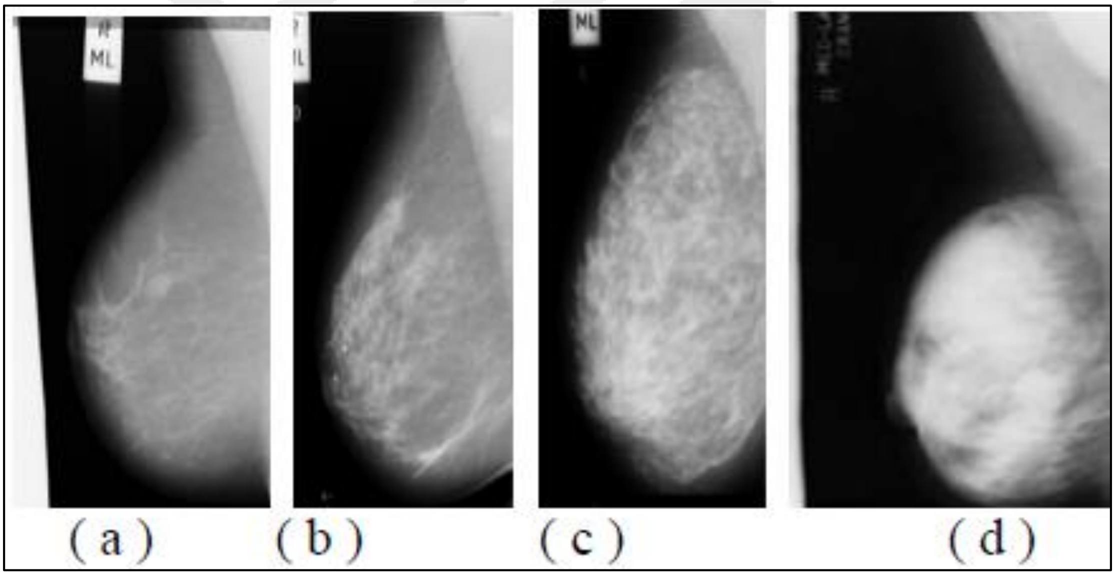


Figure 2.6 Example mammograms, where the breast density increases from (a) BIRADS I to (d) BIRADSIV [32]

CHAPTER 3

MATERIALS AND METHODS

The material include the dataset has been used in this research study is the mini Mammogram Images Analysis Society (MIAS) dataset. The hospital that provided this group of images is Royal Marsden Hospital/department of Physics by John Suckling in Fulham Road, London. SW36JJ, UK. In 19 January 1995, Email: j.suckling@rmh_lon.icr.ac.uk. The original MIAS Database (digitized at 50-micron pixel edge) has been curtailed to 200-micron pixel edge. and padded in order that each image is 1024 pixels x 1024 pixels. Moreover, it is done under supervision of expert radiologist and its gray scale type if images. The mini MIAS has 315 images. The description of the dataset is [33]:

- First column has the number of image reference.
- Second column has the type of tissue of the breast (Fatty – Fatty glandular – Dense glandular).
- Third column has existent of abnormal class:
    - CALC: is mean Calcification
    - CIRC: is mean circumscribed masses
    - SPIC: is mean Speculated masses
    - MISC: is mean Other, ill-defined masses
    - ARCH: is mean Architectural distortion
    - ASYM: is mean Asymmetry
    - NORM: is mean Normal
- The fourth column has type of abnormality (benign – malignant).

- The fifth and sixth column are for the x-axis and y-axis for the center of abnormalities.
- Seventh column represent approximate radius of a circle within the abnormality (in pixels) .

In addition, it can download from [10].

This research study has taken just the image with Mass, and the type of Masses are (CRIC – SPIC – MISC) and all implemented in MATLAB R 2013.

There are many methods to detect the mass region of breast cancer in mammography images. This research study proposed new method that combines different image processing techniques to extract mass regions. These techniques are preprocessing images using crop image depend on the information that given in the dataset, image enhancements by using two techniques (unsharpmask and adjust image), image segmentations using morphological operations, and feature extraction using regionprop technique.

3.1 Preprocessing Images

Depending on previous information of the dataset given by radiologist about the location and radius of the mass region cancer, we crop the image to avoid the effect of other tissue in the breast. For cropping, the image uses the standard function in Matlab (*imcrop).*

3.2 Image Enhancement (Unsharp Mask)

The technique was first use in Germany during the 1930s as a way of increasing the acutance, or apparent resolution, of photographic images. A well-known convolution mask is the un-sharp mask. When an image blurred by some unknown phenomenon, each pixel is composed of its own true value plus the fractional components of its neighbors. This technique uses this concept to reduce the blur and improve the image through the reduction of low frequency information and amplification of high frequency detail [34].

The image resultant from the unsharp masking can be obtained by subtracting a low-pass filtered image (†lp†) from the input image (†in), which corresponds to a high-pass filtered image (†hp†). This high-pass filtered image is weighted, C(x, y), and added to the input image, equations 2.1, 2.2 [35]. This operation allows the amplification of the details due to the high-pass filter, as reduces low-frequency information and amplifies high frequency details [35], [36].

$$\dagger\, out(x,y) = \dagger\, in(x,y) + C(x,y)\,[\dagger\, in(x,y) - \dagger\, lp \dagger (x,y)] \quad (2.1)$$

$$= \dagger\, in(x,y) + C(x,y)\ \dagger\, hp \dagger (x,y) \quad (2.2)$$

Typically, two settings control unsharp masking:

- Amount is listed as a percentage, and controls the magnitude of each overshoot (how much darker and how much lighter the edge borders become). This can also thought of as how much contrast added at the edges. It does not affect the width of the edge rims.
- Radius affects the size of the edges to enhance. On the other hand, how wide the edge rims become, so a smaller radius enhances smaller-scale detail. Higher Radius values can cause halos at the edges, a detectable faint light rim around objects. Fine detail needs a smaller Radius. Radius and Amount interact; reducing one allows more of the other.

It seems that Radius (rds) = 8 & Amount (amnt) = 4 is a good choice for enhancing. Figure 3.1 shows an example of the used of un-sharp filtering on a real mammographic image with chosen parameters that mention above.

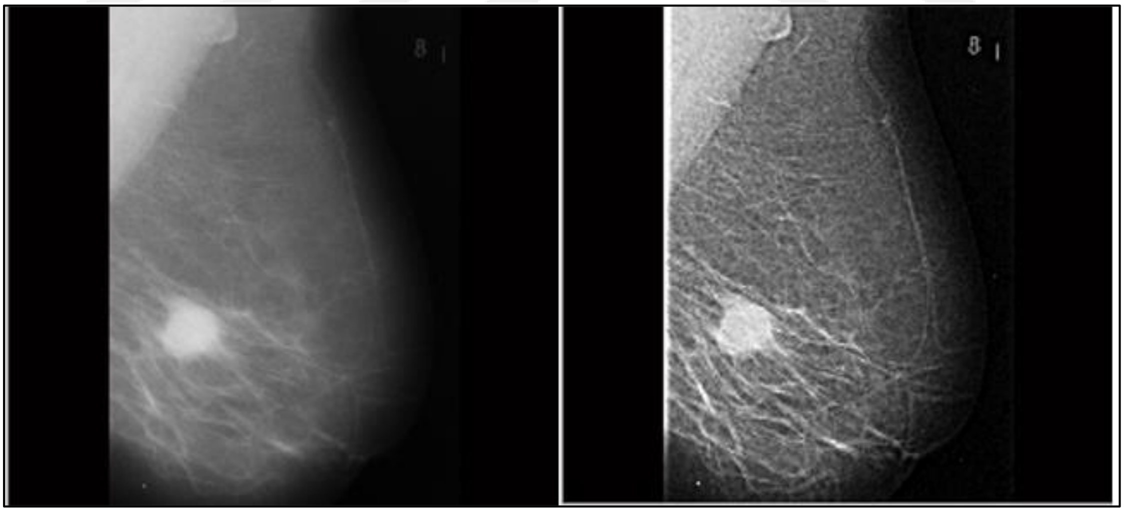


Figure 3.1– Result of un-sharp filtering

The next step in this part use adjust image to increase the contrast of the output image using matlab function (imadjust) [22]. After image has been, enhance. The following step is to convert the result image to logical (black-white, BW) image using thresholding. By trying different threshold values, at first the threshold value obtained using Otsu's method will be apply. Then the following percentage values of 50, 75 and 90th of pixel distribution will be implement. Results have shown that 75th percentile gave a good threshold to catch the "mass" region. Therefore, the experimental work focused towards using the last threshold image so segment the "mass" region.

3.3 Morphological Operations Based Segmentation

Binary images might be contained uncounted weakness. In several conditions, binary regions constructed by plain thresholding are buckle by noise and construction. Morphology is a vast range of image processing operations that modulate the images based on forms. It is consider one of the data processing procedures useful in image processing. It has many implementation such as texture analysis, noise abstraction, boundary obtaining etc. [37]. Working on morphological image chases the goal of eliminating all these weakness and protect the structure of image. Morphological operations are confident only on the correlating regulation of pixel values, rather than their numerical values, so they are concentrate more on binary images, however it can also exercise to grayscale images such that their light transfer functions are unknown and therefore their absolute pixel values are not possessed into account [38]. Morphological methods verify the image with a small template named structuring element. This structuring element applying for all possible positions of the input image and creates the same size output. At this method, the output image pixel values based on similar pixels of input image with its neighbors. This operation produces a new binary image in which if test is successful it will have non-zero pixel value at that location in the input image. There are multiple structuring element such as cross-shaped, square shaped, diamond shaped etc. The base of the morphological operation is dilation, erosion, opening, closing donated by logical AND, OR notation and described by group of analysis. Among them in this paper, two operations are applied dilation and erosion. Dilation adds pixels whereas erosion removes the pixels at borders of the items. This removal or adding of pixels depends on the structuring element used for processing the image.

3.3.1 Dilation

It is one of the mathematical morphology. It is possible to apply to a binary image and a grayscale image. When implementation dilation, the objects grown in size. The effect of dilation will increase the boundaries of foreground, therefore areas increase in size and holes [39]. Dilation depended on two parts as data. Firstly, the input image and second is

Structure Element (SE), which is helping to determine how much the image dilated, and is known as kernel. The definition of dilation in mathematical can be explain as follows [40]:

Suppose A be a set of input image coordinates, B be a set of SE coordinates, and Bx is a translation of B so that its origin is at x. Thus, dilation of A by B is set of all points of x such that intersection of Bx with A is not null. In terms of set operations, dilation of A by B as below:

$$A \oplus B = \{z \in E | (B^s)_z \cap A \neq \emptyset\} \quad (2.3)$$

Bx refers to the symmetric of B, that is,

$$Bx = \{x \in E | -x \in B\} \quad (2.4)$$

3.3.2 Filling the Region

Dilation operation makes the borders of the object wide so for segmenting the object the subsequently step is to fill the holes. The flood fills procedure most frequently known to fill the holes in the specified input image. For binary images, it makes changes for background pixels into foreground pixels while it touch the object borders and for grayscale images, it create the density level same i.e. it makes the dark areas surrounded by lighter areas to same density levels [20]. In binary images and gray-scale images, the boundaries of the objects need to be specifying by connectivity. In binary images, the starting point for filling can also specified. If we specify holes as an argument then it is no need to specify any starting points [20]. In this paper fill, operation used on binary image with arguments holes so it automatically fills the holes of different objects in image.

3.3.3 Erosion

Erosion is a mathematical morphology. In addition, it causes to shrink the objects or become thin in size, and it erodes the foreground boundaries, these lead areas of those objects shrink in size, and holes of those areas become larger [21]. Therefore, after dilution and filling the holes of object in some images the boundaries mixed up so to somewhat separate the boundaries erosion applied to make the boundaries of the objects thinner for better output. Erosion is depending on two parts. Firstly, the input image and secondly, is on Structure Element (SE), which is necessary to determining how much the image can eroded. The definition of erosion can be as below [40]: Suppose A be a set of input image coordinates and B be a set of SE coordinates and Bx is a translation of B so that its origin is at x. Thus, dilation of A by B is set of all points of x such that Bx is a subset of A. In terms of set operations, erosion of A by B as below:

$$A \ominus B = \{x \in E | B_x \subseteq A\} \quad (2.5)$$

B_x is the explanation of B through the vector x

Different morphological operations on thresholding image to segment the "mass" region can implemented, while removing line-like structures (i.e. streaks). As expected, imopen (first erosion, then dilation) provide better results than closing. Therefore, the focus made towards the "opening". By trying different Structuring Element (SE) sizes and use all these elements together (one after another).

Opening operator is the dilation of the erosion of a set A by a structuring element B[41]

$$A \circ B = (A \ominus B) \oplus B \quad (2.6)$$

In the equation above, $\ominus$ stands for erosion and $\oplus$ attributes to dilation.

Opening operation enthralls the features below [42]:

1. Opening of A via B is a subgroup of A.
2. As long as C is a subgroup of D. In that case, C ◦ B is a subgroup of D ◦ B.
3. (A ◦ B) ◦ B = A ◦ B.

If dilation followed by erosion resulting is closing operator. The closing of an image A by the structuring element B is expressed through $A \bullet B$ and as can be seen in the following equation:

$$A \bullet B = (A \oplus B) \ominus B \quad (2.7)$$

When there is no change in an image A by closing with B, it considered to be close with reference to B.

The closing operation enthralls the properties below:

1. A is a subgroup of closing of A via B.
2. As long as C is a subgroup of D. In that case, C • B is a subgroup of D • B.
3. (A • B) • B = A • B.

The third condition in two cases (opening and closing)that multiple openings or closing of a set have no effect after has been applied the operator [42].

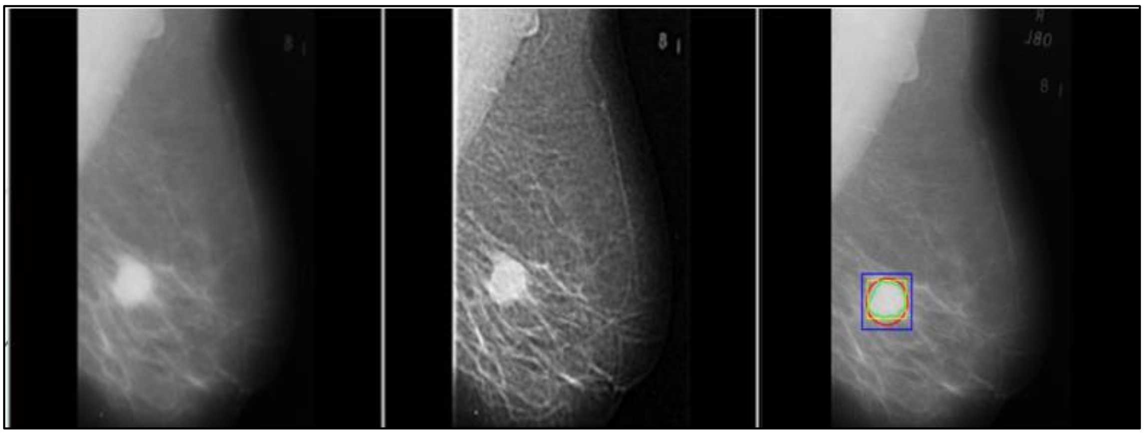
Opening and closing with an equivalent structuring component has been utilized so as to eradicate explicit details of the image less than the structuring element. Nonetheless, the general construction of the objects is not corrupted. Closing connects objects, not remote from each other, melt away minute holes, and smooth the object, which is defined by melting away narrow gulfs. The terms 'near', small and 'narrow' refer to the dimensions and the pattern of the structuring component [43]. Incorporating openings and closing with image subtraction leads to top-hat and bottom-hat transformations [41].The top-hat transformation of a gray-scale image (A) is shown as (A) minus it is opening.

$$T_{hat}(A) = A - (A \circ B) \tag{2.8}$$

The bottom-hat of a gray-scale image (A) is closing minus A:

$$B_{hat}(A) = (A \bullet B) - A \tag{2.9}$$

The principle application of these filters is removing undesirable objects from an image by using an structure element (SE) .The top-hat filter is exercised in order to discharge light objects on a dark background and correcting the effects of non-uniform illumination whereas the bottom-hat filter is used for dark objects on a light background. It seems that seize 15 is good choice. The following step is to clean the border and fill holes if there is any by using (imclearborder) and (imfill) and Remove small objects (less than 50 pixels) and Doing a little bit of smoothing on the border of the mass, using a disk shaped by Closing operation. As show Figure in 3.2.



a b c

Figure 3.2 a) Original image b) Enhancing image c) Segment image

3.4 Feature Extraction

In this part designation, the region props is stander command in Matlab to measure a set of properties for each connected component in the segment image. The properties that been used during this work are:

a- Area: The actual number of pixel in the region

b- Centered: Returns a 1-by-Q vector specifying the center of mass of the area.

c- MajorAxisLength: Returns a scalar that specifies the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

d- MinorAxisLength: Returns a scalar that specifies the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

e- BoundingBox: Returns the smallest rectangle containing the region, specified as a 1-by-Q*2 vector.

f- PixelIdxList: Returns a p-element vector that contains the linear indices of the pixels in the region.

The Figure 3.3 is portrays proposed method

```
Load images (I)
Info_ I = read x-axis, y-axis, and radius of (I)
Crop_ image = imcrop (I, Info_ I)
I_ Enhance1 = unsharpmask (Crop_ image) %for first enhance image
I_ Enhance2 = adjust (I_Enhance1) %for increasing the contrast of I_ Enhance1
I_ BW = Threshold (I_ Enhance2) %convert the image from gray scale to black and white
I_ open = Opening (I_ BW) %for remove small objects
I_ segment = Opening (I_ open) %for segmentation
I_ fillholes = imfill (I_ segment) %fill the holes
I_ clear = imclear (I_ fillholes) %for clear borders
I_ smooth = Closing (I_ clear) %for smoothing the border of Mass
Feature_ Extraction = regionprob (I_ smooth)
```

Figure 3.3-pseudo code of proposed method

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Results

The table 4.1 shows the comparison between the original results and the results of our proposed method. The legend for the table is as follows:

First column: MIAS database image name.

Second column shows background tissue's Character

F means Fatty

G means Fatty-glandular

D means Dense-glandular

Third column shows the abnormal existence class

CALC means Calcification

CIRC means Well-defined/circumscribed masses

SPIC means Spiculated masses

MISC means other, ill-defined masses

ARCH means Architectural distortion

ASYM means Asymmetry

Fourth column shows abnormality's riskiness (Diagnosis)

B means Benign

M means Malignant

Fifth, sixth columns: x, y represents the image-coordinates of center for abnormality (provided in the database)

Seventh column: represent approximate radius of a circle within the abnormality (in pixels).

Eighth, ninth columns: x, y represents the image-coordinates of center for abnormality (estimated by our approach)

Tenth column: Radius (in pixels) of a circle enclosing the abnormality (estimated by our approach)

Eleventh column: Percentage Error between Original (column 7) and Estimated Radii (column 10)

Percentage Error in Radius (%)

$$= 100 \times \frac{|Original\ Radius - Estimated\ Radius|}{Original\ Radius} \qquad (4.1)$$

Table 4.1 Comparison between the original data and results of our proposed method.

MIAS Database Image Name	Character of Background Tissue	Mass Type	Diagnosis	Original Mass Center and Size			Estimated Mass Center and Size			Percentage Error in Radius (%)
				X	Y	Radius	X	Y	Radius	
mdb001	G	CIRC	B	535	425	197	561	451	166	15.74
mdb002	G	CIRC	B	522	280	69	504	314	61	11.59
mdb005	F	CIRC	B	477	133	30	487	146	26	13.33
mdb005	F	CIRC	B	500	168	26	498	154	23	11.54
mdb010	F	CIRC	B	525	425	33	523	425	28	15.15
mdb012	F	CIRC	B	471	458	40	461	463	31	22.50
mdb013	G	MISC	B	667	365	31	675	364	24	22.58
mdb015	G	CIRC	B	595	864	68	599	862	53	22.06
mdb017	G	CIRC	B	547	573	48	544	552	42	12.50
mdb019	G	CIRC	B	653	477	49	654	484	42	14.29
mdb021	G	CIRC	B	493	125	49	499	134	40	18.37
mdb023	G	CIRC	M	538	681	29	541	677	24	17.24
mdb025	F	CIRC	B	674	443	79	684	437	66	16.46

Table 4.1 (cont'd)

mdb028	F	CIRC	M	338	314	56	339	317	46	17.86
mdb030	G	MISC	B	322	676	43	317	652	42	2.33
mdb032	G	MISC	B	388	742	66	386	729	56	15.15
mdb058	D	MISC	M	318	359	27	327	346	22	18.52
mdb063	D	MISC	B	546	463	33	551	453	27	18.18
mdb069	F	CIRC	B	462	406	44	478	392	38	13.64
mdb072	G	ASYM	M	266	517	28	274	506	26	7.14
mdb075	F	ASYM	M	468	717	23	471	712	21	8.70
mdb080	F	CIRC	B	432	149	20	428	158	17	15.00
mdb081	G	ASYM	B	492	473	131	506	462	109	16.79
mdb083	G	ASYM	B	544	194	38	552	196	34	10.53
mdb090	G	ASYM	M	510	547	49	501	540	50	2.04
mdb091	F	CIRC	B	680	494	20	681	491	19	5.00
mdb092	F	ASYM	M	423	662	43	429	655	37	13.95
mdb095	F	ASYM	M	466	517	29	467	511	26	10.34
mdb097	F	ASYM	B	612	297	34	613	293	30	11.76
mdb099	D	ASYM	B	714	340	23	709	328	19	17.39
mdb102	D	ASYM	M	415	460	38	415	453	32	15.79
mdb104	D	ASYM	B	357	365	50	356	353	43	14.00
mdb105	D	ASYM	M	516	279	98	535	294	89	9.18
mdb107	D	ASYM	B	600	621	111	557	595	103	7.21
mdb110	D	ASYM	M	190	427	51	214	433	45	11.76
mdb111	D	ASYM	M	505	575	107	512	565	89	16.82
mdb115	G	ARCH	M	461	532	117	462	464	102	12.82
mdb117	G	ARCH	M	480	576	84	525	576	81	3.57
mdb120	G	ARCH	M	423	262	79	446	306	68	13.92
mdb121	G	ARCH	B	492	434	87	548	444	83	4.60
mdb124	G	ARCH	M	366	620	33	360	606	23	30.30
mdb125	D	ARCH	M	700	552	60	654	538	59	1.67
mdb126	D	ARCH	B	191	549	23	202	541	21	8.70
mdb127	G	ARCH	B	523	551	48	555	532	44	8.33
mdb130	D	ARCH	M	220	552	28	225	542	26	7.14
mdb132	F	CIRC	B	252	788	52	239	781	47	9.62
mdb132	F	CIRC	B	335	766	18	333	763	14	22.22
mdb134	F	MISC	M	469	728	49	470	727	41	16.33
mdb141	F	CIRC	M	470	759	29	483	753	26	10.34
mdb142	F	CIRC	B	347	636	26	354	635	21	19.23
mdb144	F	MISC	B	233	994	29	236	1000	23	20.69
mdb144	F	MISC	M	313	540	27	317	543	23	14.81

Table 4.1 (cont'd)

mdb145	D	SPIC	B	669	543	49	648	523	41	16.33
mdb148	F	SPIC	M	326	607	174	418	572	167	4.02
mdb150	F	ARCH	B	351	661	62	344	701	54	12.90
mdb152	F	ARCH	B	675	486	48	689	493	43	10.42
mdb155	F	ARCH	M	448	480	95	456	497	78	17.89
mdb158	F	ARCH	M	540	565	88	489	567	84	4.55
mdb160	F	ARCH	B	536	519	61	521	529	56	8.20
mdb163	D	ARCH	B	391	365	50	426	402	31	38.00
mdb165	D	ARCH	B	537	490	42	514	472	39	7.14
mdb167	F	ARCH	B	574	657	35	573	655	30	14.29
mdb170	D	ARCH	M	489	480	82	528	455	71	13.41
mdb171	D	ARCH	M	462	627	62	486	589	55	11.29
mdb175	G	SPIC	B	592	670	33	590	653	30	9.09
mdb178	G	SPIC	M	492	600	70	491	576	59	15.71
mdb179	D	SPIC	M	600	514	67	568	544	57	14.93
mdb181	G	SPIC	M	519	362	54	531	374	50	7.41
mdb184	F	SPIC	M	352	624	114	352	626	90	21.05
mdb186	G	SPIC	M	403	524	47	416	512	29	38.30
mdb188	G	SPIC	B	406	617	61	444	600	52	14.75
mdb190	G	SPIC	B	512	621	31	516	619	27	12.90
mdb191	G	SPIC	B	594	516	41	603	517	35	14.63
mdb193	D	SPIC	B	399	563	132	393	470	87	34.09
mdb195	F	SPIC	B	725	129	26	719	141	26	0.00
mdb198	D	SPIC	B	568	612	93	540	588	81	12.90
mdb199	D	SPIC	B	641	177	31	642	183	27	12.90
mdb202	D	SPIC	M	557	772	37	557	770	30	18.92
mdb204	F	SPIC	B	336	399	21	337	395	17	19.05
mdb206	F	SPIC	M	368	200	17	372	201	14	17.65
mdb207	D	SPIC	B	571	564	19	570	555	17	10.53
mdb209	G	CALC	M	647	503	87	628	549	79	9.20
mdb211	G	CALC	M	680	327	13	681	317	13	0.00
mdb212	G	CALC	B	687	882	3	689	881	4	33.33
mdb213	G	CALC	M	547	520	45	550	499	41	8.89
mdb214	G	CALC	B	582	916	11	574	916	10	9.09
mdb218	G	CALC	B	519	629	8	516	622	7	12.50
mdb219	G	CALC	B	546	756	29	546	737	28	3.45
mdb222	D	CALC	B	398	427	17	399	437	16	5.88
mdb223	D	CALC	B	523	482	29	537	482	31	6.90
mdb223	D	CALC	B	591	529	6	590	528	6	0.00

Table 4.1 (cont'd)

mdb226	D	CALC	B	287	610	7	292	611	4	42.86
mdb226	D	CALC	B	329	550	25	340	569	18	28.00
mdb226	D	CALC	B	531	721	8	525	719	8	0.00
mdb227	G	CALC	B	504	467	9	511	467	9	0.00
mdb231	F	CALC	M	603	538	44	635	522	43	2.27
mdb236	D	CALC	B	276	824	14	274	818	13	7.14
mdb238	F	CALC	M	522	553	17	519	548	16	5.88
mdb239	D	CALC	M	645	755	40	629	729	35	12.50
mdb239	D	CALC	M	567	808	25	587	802	20	20.00
mdb240	D	CALC	B	643	614	23	628	604	23	0.00
mdb241	D	CALC	M	453	678	38	476	667	38	0.00
mdb244	D	CIRC	B	466	567	52	481	538	44	15.38
mdb248	F	CALC	B	378	601	10	376	598	9	10.00
mdb249	D	CALC	M	544	508	48	534	489	50	4.17
mdb249	D	CALC	M	575	639	64	610	650	69	7.81
mdb252	F	CALC	B	439	367	23	443	372	24	4.35
mdb253	D	CALC	M	733	564	28	724	545	24	14.29
mdb256	F	CALC	M	400	484	37	388	488	32	13.51
mdb264	G	MISC	M	596	431	36	579	418	30	16.67
mdb265	G	MISC	M	593	498	60	577	480	56	6.67
mdb267	F	MISC	M	793	481	56	794	478	50	10.71
mdb270	G	CIRC	M	356	945	72	309	965	68	5.56
mdb271	F	MISC	M	784	270	68	786	266	56	17.65
mdb274	F	MISC	M	127	505	123	106	503	126	2.44
mdb290	D	CIRC	B	337	353	45	356	355	38	15.56
mdb312	F	MISC	B	240	263	20	241	263	17	15.00
mdb314	F	MISC	B	518	191	39	509	196	37	5.13
mdb315	D	CIRC	B	516	447	93	541	468	91	2.15
Mean of Percentage Errors in Lesion Radius Estimation (%)										**12.65**

The total number of benign and malign cases in the table is 68 and 5, respectively. As a conclusion, we may want to see the distribution of the error according to the diagnosis type during next step. This information illustrated in Table 4.2. It observed that there is no significant difference between the average error rates (13.25 versus 11.84 %) for different diagnosis types.

Table 4.2 the distribution of error data/results according to diagnosis type (M - Malign, B - Benign).

Diagnosis	Number of Cases	Average Percentage Error (%)
B	68	13.25
M	51	11.84

Similarly, we may also want to see the distribution of the error as far as background tissue type is concerned. This information is illustrated in Table 4.3. Again, it is observed that there is no significant difference between the average error rates for different background tissue types.

Table 4.3 the distribution of error data/results in accordance with background tissue type (D - Dense-glandular, F - Fatty, G - Fatty-glandular).

Diagnosis	Number of Cases	Average Percentage Error (%)
D	40	12.99
F	41	12.33
G	38	12.63

Finally, we may also want to see the distribution of the error according to the type or shape of the abnormality, i.e. the lesion. This information is illustrated in Table 4.4. Again, it observed that there is no significant difference between the average error rates for different abnormality types or lesion shapes. All these tables indicate that even though there is considerable (i.e., approximately 10%) error compared to the ground truth, our proposed segmentation method is very robust as it works equally well across different cases (tissue types, lesion types, lesion shapes, etc.).

Table 4.4 The distribution of error data/results according to abnormality type.(ARCH - Architectural distortion, ASYM - Asymmetry, CALC - Calcification, CIRC - Circumscribed masses, MISC - Other, ill-defined masses, SPIC - Spiculated masses)

Diagnosis	**Number of cases**	**Average Percentage Error (%)**
ARCH	19	12.06

Table 4.4 (cont'd)

ASYM	15	11.56
CALC	27	9.70
CIRC	24	14.26
MISC	15	13.52
SPIC	19	15.53

4.2 Discussion

This research study works with the automatic enhancement and segmentation of mass in mammographic images. In addition, how breast cancer background affects the lives of many women is presented within this thesis. What is more, this dissertation describes the implementation and application of image preprocessing crop image and enhancement techniques such as unsharp masking and image adjust together with the evaluation of several different parameters. Image segmentation and smoothing techniques such as morphological operation image segmentation are outlined and applied. Afterwards, feature extract-using region props function of MATLAB's image processing toolbox. All experimental work and procedures performed in a set of mammograms with different breast densities from mini MIAS Database and implemented in MATLAB R2013.

Table 4.1 shows the result of error percentage that is obtained by applying proposed method in determining radius of mass region in breast cancer and comparing the results against original values provided in the database. This is the key point for the performance assessment of our proposed method.

The best images/cases where we have 0% error are mdb195, mdb211, mdb223, mdb226, mdb227, mdb240, mdb241. The image analysis steps and results are shown in Figures 4.1, 4.2, 4.3 4.4, 4.5, and 4.6 for these cases, respectively.

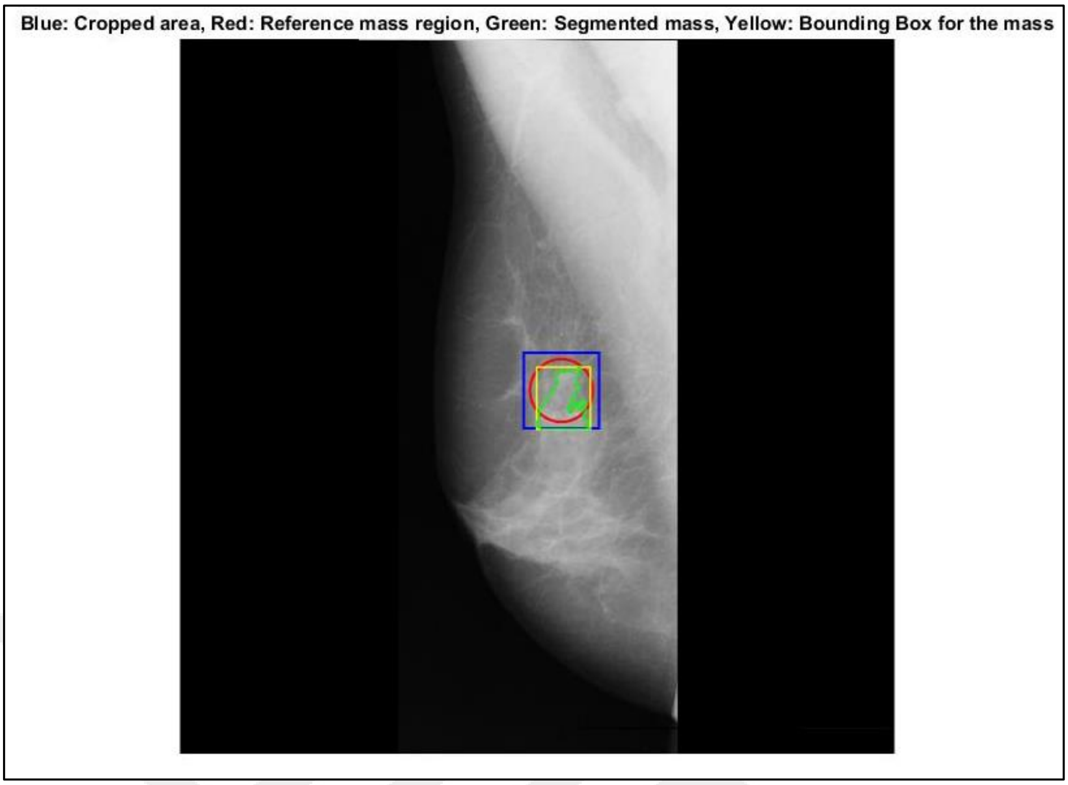


Figure 4.1 Malignant Mass Cancer (mdb211)

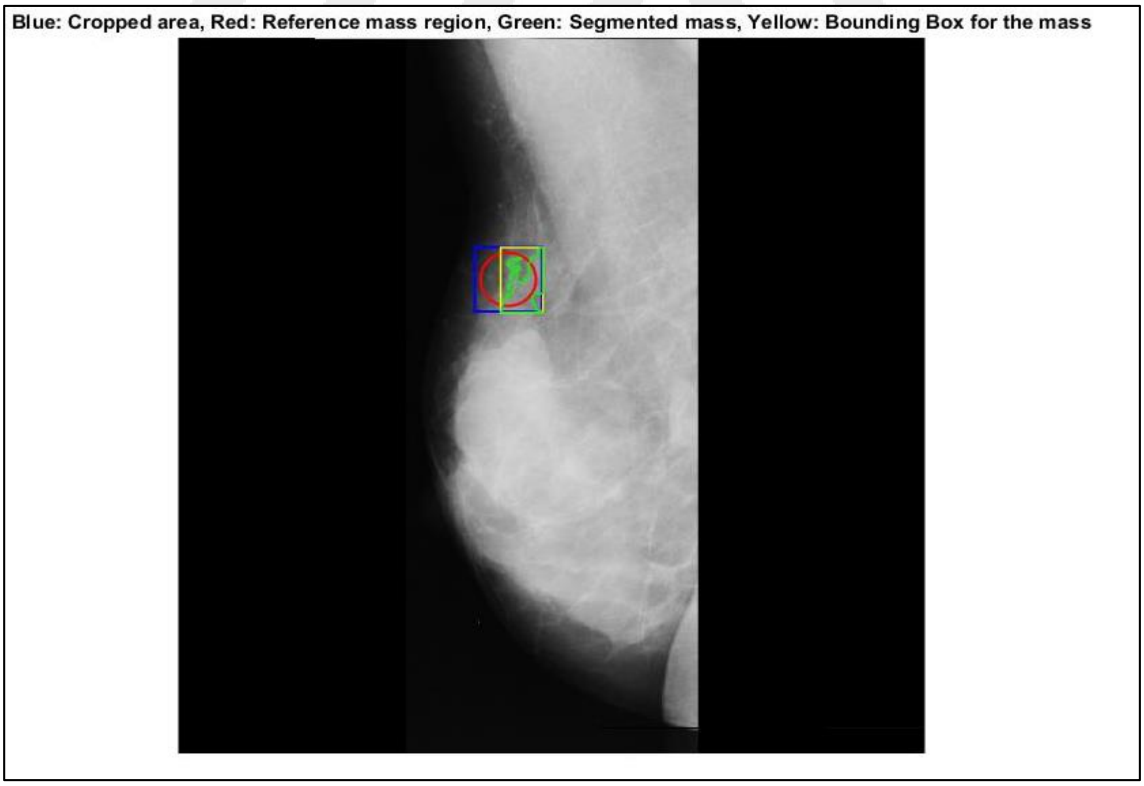


Figure 4.2 Malignant Mass Cancer (mdb241)

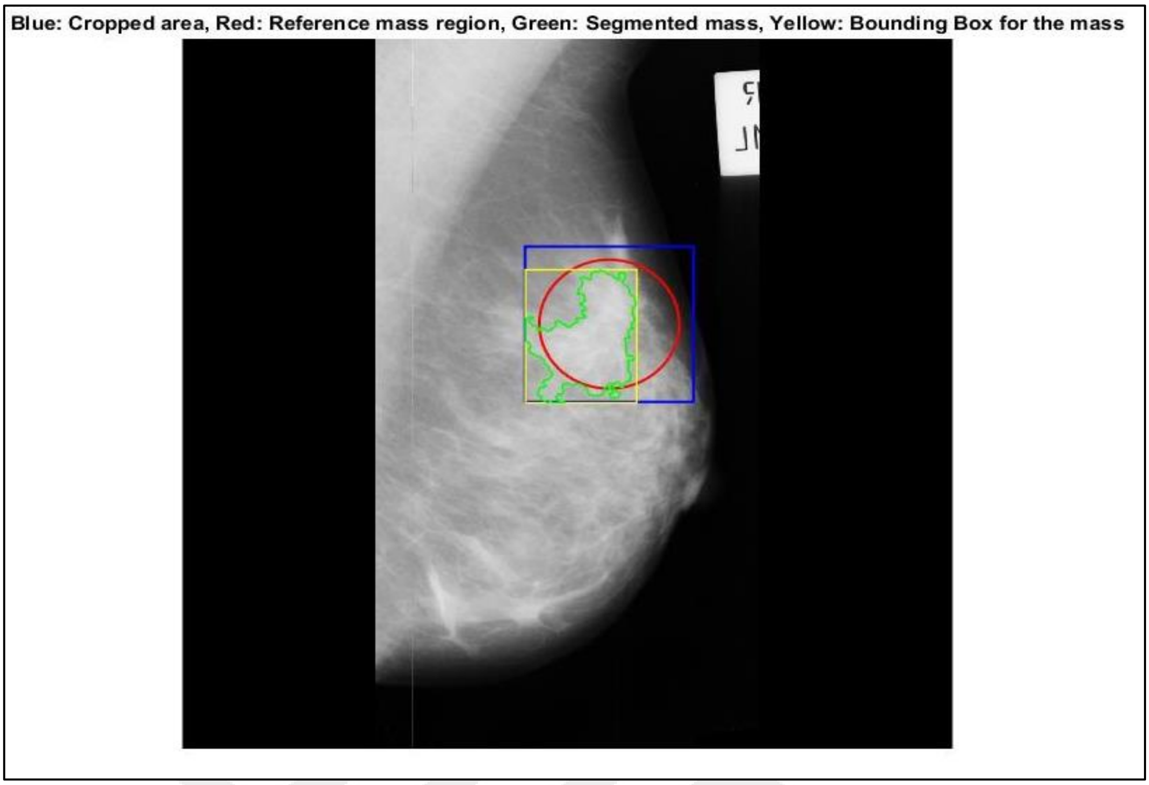


Figure 4.3 Benign Mass Cancer (mdb195)

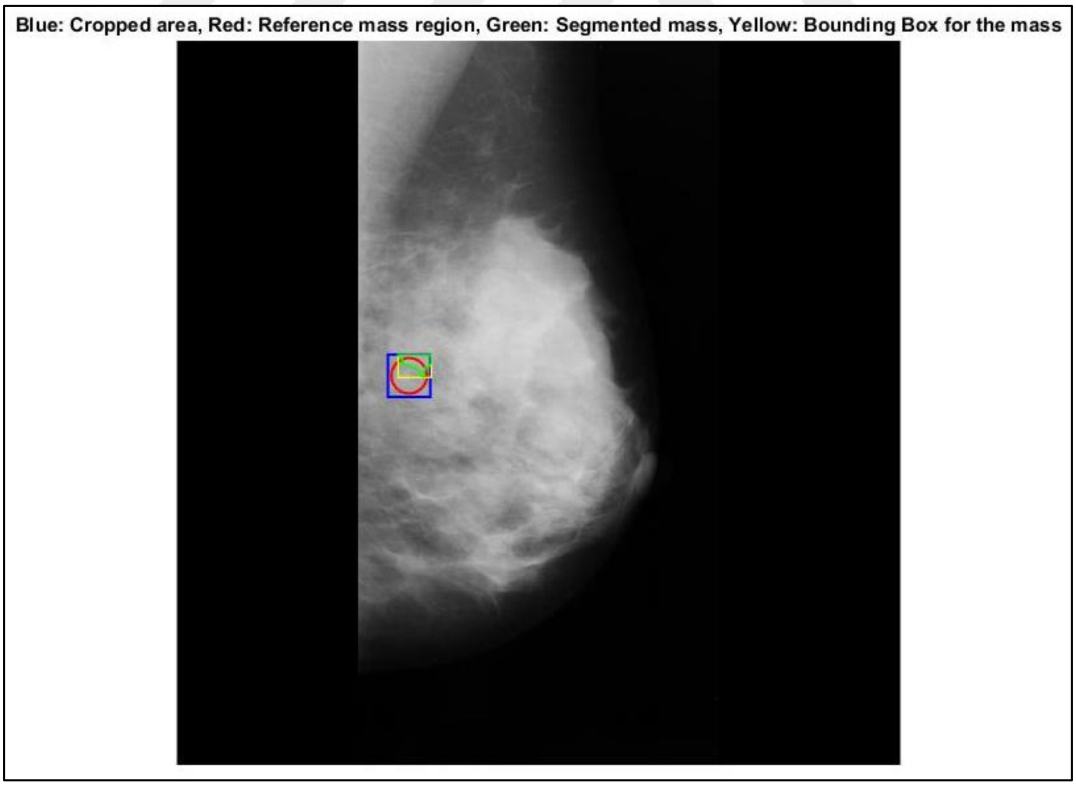


Figure 4.4 Benign Mass Cancer (mdb223)

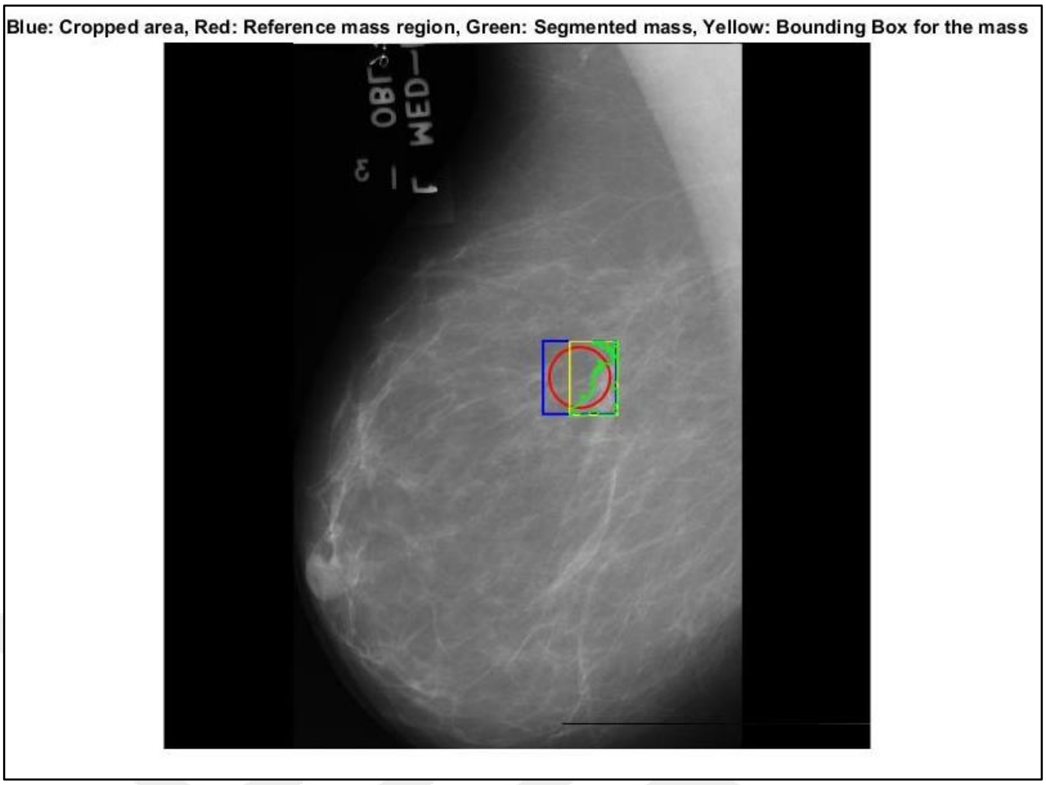


Figure 4.5 Benign Mass Cancer (mdb226)

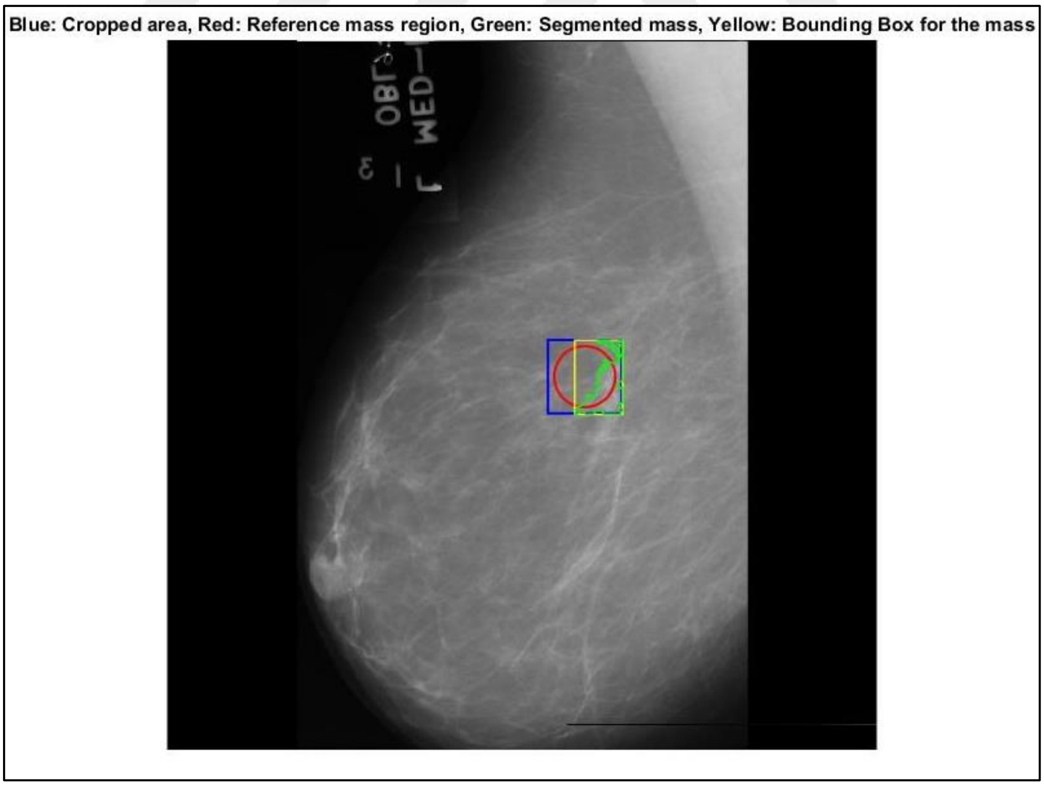


Figure 4.6 Benign Mass Cancer (mdb227)

The cases with the worst results were, mdb226 → 42%, mdb186, and mdb163 →38%, mdb212→ 33.33%, and mdb124→30.30%.The Figures 4.7, 7.8, and 4.9 respectively show these cases/results.

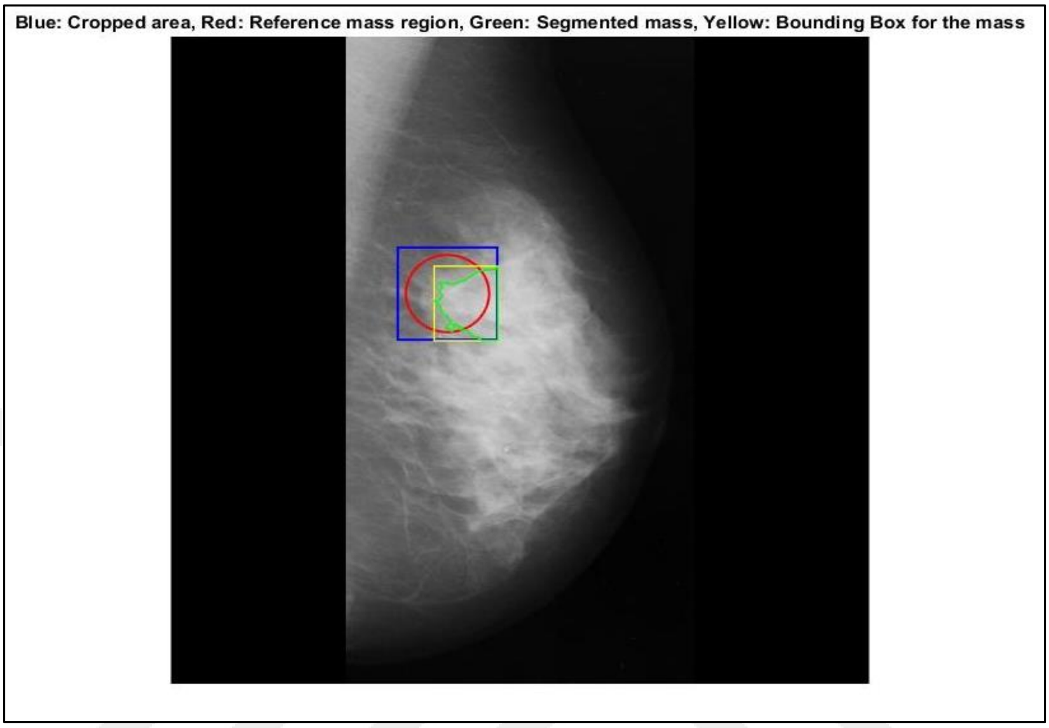


Figure 4.7 Malignant Mass Cancer (mdb186)

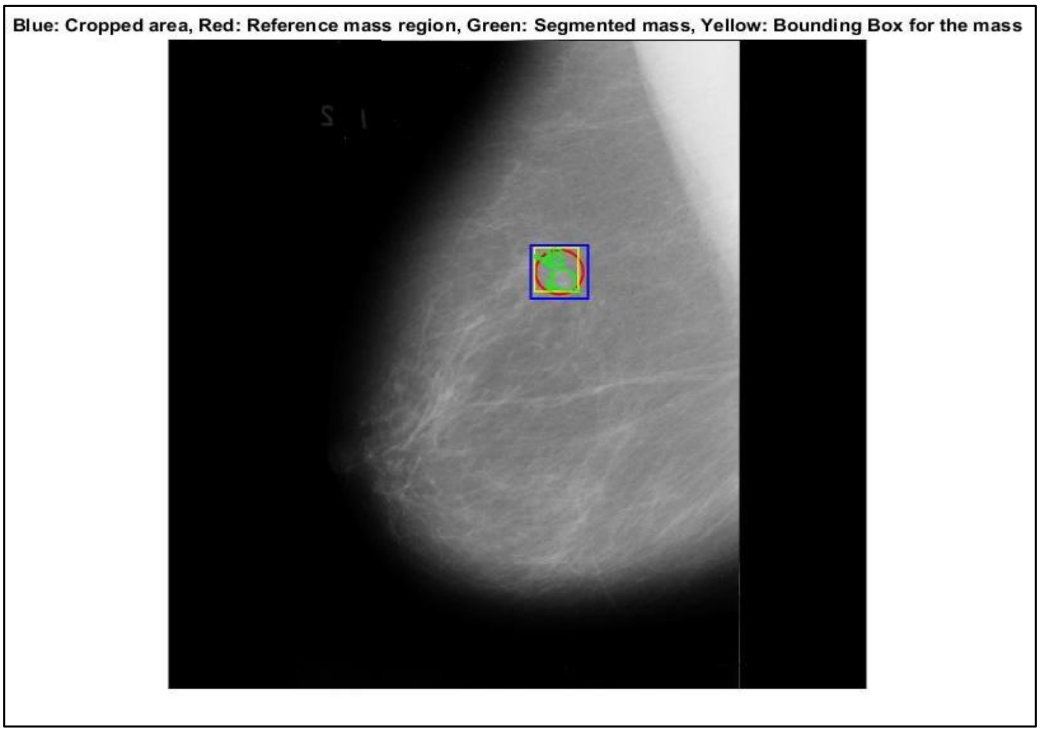


Figure 4.8 Benign Mass Cancer (mdb163)

Figure 4.9 Benign Mass Cancer (mdb226)

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Breast cancer is a type of disease and statistics suggest that breast cancer affects one in eight women throughout their lifetime. Reports also claim that there has been cases of breast cancer diagnosed in men. However, it seems that high percentage of this disease remains within the women population. It starts with a malignant tumor that initiated in cells of the breast and at later stages; it spreads to other tissues in the affected area. There are more than 50, 000 women diagnosed with breast cancer in the United Kingdom each year. Similar number also reported for the deaths caused by breast cancer in the United States.

Like any other disease, it is advantageous if symptoms of breast cancer are diagnosed and detected at the very early stages. Early diagnose means that right and appropriate medical treatment and help can be given straightaway and thus save the lives of tens of thousands of women every year. In recent years, information technology has applied as part of several methods to diagnose breast cancer. Medical Image Processing has developed and it is now widely used to help specialists in detecting the disease. People with breast cancer can take advantage of this technology by consulting their doctors as soon as they suspect unusual growth or any other symptoms of disease.

The disease is often explained by an un-controlled malignant growth. The growth typically starts in the cells within the breast tissues, which in a healthy and normal situation can governed and controlled to a specific order without causing any problem. In the case of breast cancer, the regeneration and growth of the cells becomes uncontrollable. Thus, the whole process of no reparation of the eventual mutations fails and this will lead to the creation of a cancer tumor.

It has agreed that key factor of controlling the spread of breast cancer disease is the early detection and diagnose as soon as the patient notices any unusual growth in the breast area. This will also help in drawing on the best and appropriate medical treatment to eradicate the disease and prevent it from becoming a terminal one and therefore, increasing the opportunities to save lives. Unfortunately, statistics shows that a growing number of patients are dying of breast cancer because of not taking medical help towards the early detection and diagnose. Recovery is very slim for those patients who tested positive for cancer after it stayed in their body un-detected. The earlier the disease is detected , the better chances of survival can be achieved. Research work conducted by Gunderman [12] has claimed that patients who are over 50 years old are more prone to having high risks of getting this type of cancer.

There are several characteristics associated with breast cancer. These include micro calcifications, masses, and architectural distortions. In addition, any asymmetry between breasts can also use as good indicator during the diagnoses of breast cancer disease. Moreover, these types are:

1- Micro calcifications: small size lesions, which can develop during the early stages of breast cancer. These lesions have a size typically in the range 0.05 to 1 mm.

2- Masses: appear as dense regions of different sizes and properties. They can be circular, oval, lobular or irregular/speculated and their margins can be:

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

   The ill-defined and speculated borders have higher probability of malignancy. A benign process is usually associated with the presence of circular or oval masses. Other studies have reported that the great variability of the mass appearance could be a huge challenge as well as an obstacle to a correct mammography analysis.

3- Architectural distortions refer to the derangement of the normal disposition of the parenchyma in a radiating or arbitrary pattern, without a visible center or mass. These are very variable and, consequently, very difficult to detect.

This research study focuses on the second type of this cancer called “Masses”. For the detection of the disease, X-rays are in use as diagnostic tool in mammography for the examination of human breast. These examinations are record as specialized images, which are then observe by radiologists for any possible abnormality.

There are many methods to detect the mass region of breast cancer in mammography images. This research study proposed new method that combines different image processing techniques to extract mass regions. These techniques are preprocessing images using crop image depending on the information given in the dataset, image enhancements by using two techniques (unsharpmask and adjust image), image segmentations using morphological operations, and feature extraction using regionprop technique. The error percentage for the result of this research study is 12% which is a good percentage compared to other results from different studies. Actually, sometimes the reference information that comes with the database is questionable as the data provider expressed in their information documents, the doctors marked the canters and the radii of the only in an approximate sense.

All experimental work and procedures were perform on a set of mammograms with different breast densities from mini MIAS Database, and implemented in MATLAB R2013.

5.2 Future Work

In this study, as an estimation of radii of mass regions in mammograms, we have obtained average percentage error rate of 12%, in comparison with the ground truth. This figure seems relatively high. However, as stated earlier, the radiologists may make mistakes while marking the regions or their marking result is only approximate. Therefore, there are some questions regarding to the integrity or validity of the ground truth information. In future studies, we are planning to test our methods on other data sets as well, so that we can shed light on some of these “ground truth” issues. We will then assess or observe the average of percentage error for those cases and if the values do not fall below 5%, we will revise and further develop our algorithms.

REFERENCES

[1] Jenefer, B.M. and Cyrilraj, V., (2014). "An Effictent Image Processing Methods for Mammogram Brest Cancer Detection", Journal of Theoretical and Applied Information Technology, 69(1): 32-39

[2] Basha, S., and Prasad, K., (2009). "Automatic Detection of Brest Cancer Mass in Mammograms Using Morphological Operators and Fuzzy C--Means Clustering", Journal of Theoretical and Applied Information Technology, 5(6): 704-709.

[3] Singh, N., Mohapatra, A., and Kanungo, G., (2011). "Breast cancer mass detection in mammograms using K-means and fuzzy C-means clustering. International Journal of Computer Applications, 22(2): 8887–0975.

[4] Yasmin, M., Sharif, M. and Mohsin, S., (2013). "Survey paper on diagnosis of breast cancer using image processing techniques", Research Journal of Recent Sciences ISSN, 2(10): 88-98.

[5] da Cruz, C., (2011). Automatic analysis of mammography images: enhancement and segmentation techniques, MSc thesis in Bioengineering, Porto University, Portogal.

[6] Cheng, H., (2003). "Computer-aided detection and classification of microcalcifications in mammograms", Pattern Recognition, 36(1): 2967 – 2991.

[7] Morton, M., Whaley, D., and Brandt, K., (2006). "Screening Mammograms: Interpretation with Computer-aided Detection—Prospective Evaluation", 239(2): 375-383.

[8] Bankman, I., Nizialek, T.,and Simon, I., Gatewood,B., (1997). "Segmentation algorithms for detecting microcalcifications in mammograms", IEEE Tran INF TechnolBiomed., 271-350.

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

[10] Dataset, http://peipa.essex.ac.uk/info/mias.html. 25 March 2016.

[11] 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.

[12] 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.

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

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

[15] 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.

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

[17] Eastman,W., and Wald, C., (2011). Getting started in clinical radiology: from image to diagnosis. Thieme. Berlin, Germany.

[18] Oliver, A., Lladó, X., and Pérez, E., (2010). "A statistical approach for breast density segmentation", Journal of digital imaging, 23(5): 527-537.

[19] Tcheslavski, V., (2010). "Morphological Image Processing: Gray-scale morphology." ELEN, 4304:5365.

[20] Yam, M., Tchou, J.,and English, R., (2001). "A mammographic Dilemma: Calcification or Haemosiderin as a Cause of Opacities? Validation of a New Digital Diagnostic Tool", The British journal of radiology, 74(887):1048-1051.

[21] Maitra, K., Nag, S., and Bandyopadhyay,K., (2011). "Identification of Abnormal Masses in Digital Mammography Images", International Journal of Computer Graphics, 2(1):17-30.

[22] Sterns, E., (1996). "Relation between Clinical and Mammographic Diagnosis of Breast Problems and the Cancer/Biopsy Rate", Canadian journal of surgery, 39(2): 128.

[23] Kekre,B., Sarode, T.K. and Gharge, M., (2009). "Tumor Detection in Mammography Images using Vector Quantization Technique", International Journal of Intelligent Information Technology Application, *2*(5): 237-242.

[24] Suckling, J., Parker, J.,and Dance, D., (1994). "The Mammographic Image Analysis Society Digital Mammogram Database", In Exerpta Medica. International Congress Series, (1069): 375-378.

[25] Lee, H., (2002). "Screening Mammography: Proven Benefit Continued Controversy", Radiologic Clinics of North America, 40(3): 395-407.

[26] Blanks, G., Wallis, G., and Moss, M., (1998). "A comparison of Cancer Detection Rates Achieved by Breast Cancer Screening Programmes by Number of Readers", Journal of Medical screening, 5(4):195-201.

[27] Astley, S., (2003). "Computer-Aided Detection for Screening Mammography", International Congress Series, 1256: 927-932.

[28] Masala, L., (2006). "Computer Aided Detection on Mammography", World Academy of Science, Engineering and Technology, 15(1): 1-6.

[29] Elmore, G., Wells, K.,and Lee, H., (1998). "What is New in Mammography", European Journal of Radiology, (27): 234-241.

[30] Rangayyan, R., Ayres, F., and Desautels, J., (2007). "A review of Computer-Aided Diagnosis of Breast Cancer: Toward the Detection of Subtle Signs", Journal of the Franklin Institute, (344): 312-348.

[31] Sheshadri,S., and Kandaswamy, A., (2005). "Detection of Breast Cancer by Mammogram Image Segmentation", Journal of Cancer Research and Therapeutics, 1(4): 232.

[32] 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.

[33] Sonka, M., Hlavac, V., and Boyle, R., (2014). Image processing, analysis, and machine vision. Cengage Learning. Stamford, New York, USA.

[34] Bae, U., Shamdasani, V., and Managuli, R., (2003). "Fast Adaptive Unsharp Masking with Programmable Mediaprocessors", Journal of Digital Imaging, 16(2):.230-239.

[35] Morrow, M., Paranjape,B., and Rangayyan, M., (1992). "Region-Based Contrast Enhancement of Mammograms", IEEE Transactions on Medical Imaging, 11(3):.392-406.

[36] Serra, J., (1982). Image analysis and mathematical morphology, v. 1. Academic press. London.

[37] Comer,L., and Delp,J., (1999). "Morphological Operations for Color Image Processing". Journal of Electronic Imaging, 8(3): 279-289.

[38] Baets, D., Kerre, E., and Gupta, M., (1995). "The Fundamentals of Fuzzy Mathematical Morphology Part 1: Basic Concepts", International Journal of General Systems, 23(2): 155-171.

[39] Rafael, C., Richard E., (2007), Digital Image Processing, House of Electronics Industry, China.

[40] Shapiro, L. and Haralick, R., (1992). Computer and Robot Vision, https://scholar.google.com.tr/scholar?hl=en&q=%5B39%5D%09R.+Haralick+and+L.+Shapiro+Computer+and+Robot+Vision%2C+Vol.+1%2C+Chap.+5%2C+AddisonWesley+Publishing+Company%2C+1992.&btnG=&as_sdt=1%2C5&as_sdtp=#, 12 July 2016 .

[41] Morpholog, thttps://en.wikipedia.org/wiki/Opening_(morphology). , 15 March 2016.

[42] Gleb, T., (2009). "Morphological Image Processing". ELEN 4304/5365.

.

CURRICULUM VITAE

PERSONAL INFORMATION

Name Surname : Safwat Alan ALDAWERI

Date of birth and place :19-3-1983/Iraq

Foreign Languages :English

E-mail : safwet.alan@yahoo.com

EDUCATION

Degree	Department	University	Date of Graduation
Master	Computer Engineering	YTU	2016
Undergraduate	Computer science	AL-MUSTANSIRYA UNVERSITY	28/06/2005

PUBLISHMENTS

Conference Papers

Al-Daweri, S., (2016). “Detection of Mass Region in Mammogram Image Using Segmentation Techniques", ICAS 2016” “International Conference on Advances Science, 31 August-2 September 2016, Istanbul, Turkey.