Sudan University of Science and Technology

Pierre and Marie Curie University

Detection of Breast Cancer in Mammogram Images Using Texture Analysis Methods

الكشف عن سرطان الثدي في صور الأشعة باستخدام التحليل النسجي

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biomedical Engineering

Ashgan Mohamed Omer

Supervisors:

Professor. **Alnazier O. Hamza (SUST)** **&** ***Professor.*** **Frédérique Frouin (UPMC)**

Dec. 2019

Dedication

For a Sudanese woman, who works for her family, tolerates the poorness, hardness, and life toughness.

Acknowledgments

I would like to acknowledge France Embassy in Sudan and Sudan University of Science and Technology, for financial support and for giving me the opportunity to develop this research project in the France and Sudan.

I would like to thank my supervisor at SUST, Prof. Alnazier Osman, for suggesting mammography as an area of radiography in which textural features would be interested and benefit. Also, I like to thank Dr. Mohamed Elfadil, for all valuable scientific materials in my research area.

I must express my deep gratitude to my supervisor at UPMC, Prof. Frédérique Frouin, for the strong motivation and enthusiasm, the nice dealing, caring when I am in France, patients, trust, strong supporting, following my work even on distance.

I would like to express my sincere gratitude to Mme. Claire Barakat, for her presence, continue supporting, nice dealing, and replying to all my emails without delaying.

Especial thanks to Mr. Pierre Muller, Mr. Roy Michel, and Mr. Abu-sufian Ali from France embassy for the gentle dealing, confident, and overcoming the difficulties in my PhD track.

My acknowledgment to all LiB staff, especially Mme. Frédérique team for attended all my presentations and for helpful discussions. Also, thanks to Jean Mac, my senior colleague for welcoming to help at any time.

I would like to say honestly and proudly, if not for the support of my family I would have done literally nothing. My grateful acknowledgements go to my mother for her unconditional support and love, and her continuousness looking after my children. My husband I would describe as a real honest Sudanese man, have supported me, and sacrificed all his rights of a stable situation with a settled wife. My great thanks to my children for tolerating a stress mother; hope they will be proud of me later.

Thanks to my beloved brother Haythem, who always pushes me to go further in pursuing the PhD, my dear sisters: Zeinab, Najlaa and Sara for continuous supporting with real love.

I am also grateful to my father's friend (Abbas Abdin), considered as real father, I would describe as honestly man, supported us financially since my father died when I was child till I graduated and get a job.

I am thankful to everyone taught me, helped me, and everyone has good willing for me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACR	American College of Radiology
Acc.	Accuracy
ANN	Artificial Neural Network
ASR	Age Standardized Rate
BIRADS	Breast Imaging Reporting and Data System
BMI	Body Mass Index
BSE	Breast Self-Examination
CAD	Computer-Aided Detection
CC-View	Craniocaudal-View
CIs	Confidence Intervals
CLAHE	Contrast Limited Adaptive Histogram equalization
CNR	Contrast to Noise Ratio
D	Dense
DICOM	Digital Imaging and Communication in Medicine
F	Fatty
$F_1, F_2, \ldots F_{21}$	SFTA features
FD	Fractal Dimension
FDA	Food and Drug Administration
FFDM	Full-Field Digital Mammography
FP	False Positive
FROC	Free-Response Operating Characteristic Curve
G	Glandular
GLCM	Grey Level Co-occurrence Matrix
GUI	Graphical User Interface
HE	Histogram Equalization
$H_1, H_2, \ldots H_{13}$	Haralick features
ICD-O	International Classification of Diseases for Oncology
JPEG	Joint Pictures Expert Group

MIAS	Mammographic Image Analysis Society
MLO-View	Mediolateral-View
MQSA	Mammography Quality Standards Act
N/A	Not Applicable
NaN	Not a Number
NCR	National population based Cancer Registry
PNG	Portable Network Graphics
PSNR	Peak Signal-to-Noise Ratio
RG	Region Growing
RGB	Red Green Blue
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RR	Relative Risk
SDF	Scatter Degradation Factor
SEER	Surveillance, Epidemiology and End Results
SFM	Screen Film Mammography
SFTA	Segmentation based-Fractal Texture Analysis
SNR	Signal to Noise Ratio
SR	Structured Reporting
SVM	Support Vector Machine
TNM	Tumor Node Metastasis
UICC	Union for International Cancer Control

Abstract

The early detection of breast cancer is of the outmost importance in order to increase survival rates. Mammography is considered as an effective tool to detect breast cancer, but due to the complexity of breast tissue texture and the observer variability effect, misdiagnosis of breast cancer frequently occurs. This research develops a simple Computer-aided detection (CAD) system in order to improve the accuracy and efficiency of mammogram interpretation. The first step of the proposed algorithm is mammographic image preprocessing which removes unwanted regions from the breast image and enhances image contrast without losing the image information. A region of interest (ROI) was then segmented using thresholding, and quantization based techniques. In mammography, the breast tissue type influences the performance of the detection. Thus, the proposed algorithm automatically characterizes the tissue type as fatty, dense or glandular. The last step of the proposed CAD system, aims at distinguishing normal from abnormal breast tissue, and subsequently differentiates benign from malignant lesions. In this thesis work, the classification tasks were based on texture features and supported vector machine (SVM). Features were extracted by using conventional well-known features (Haralick features) as well as new, less-known features (segmented fractal texture analysis SFTA). To obtain good classification performances, optimal features were selected and redundant features were removed. With the aim of showing the robustness of our approach, tests were performed using the well-known Mammographic Image Analysis Society (MIAS) database which contains annotations provided by radiologists. For tissue characterization, three approaches of features extractions were investigated; the best accuracy for distinguishing fatty tissue from non-fatty was obtained using combined Haralick and SFTA features (87% accuracy). Moreover, the best accuracy for differentiating between dense and glandular tissue was obtained using SFTA features (78 % accuracy). The SFTA features proved their superiority for the classification of normal vs abnormal lesion, and benign vs malignant lesion, with accuracies of 75.5% and 74.5% respectively.

Keywords- Breast cancer, Mammography, Computer-Aided Detection (CAD), Image processing, Tissue characterization, Segmentation, Texture features, Support vector machine (SVM)

المستخلص

ان سرطان الثدى من أحد أنواع السرطانات الشائعة عند النساء؛ وقد ثبت علمياً أن التشخيص المبكر لهذا الداء العُضال يزيد من معدلات فرص الشفاء. يعتبر التصوير الشعاعى للثدى (الماموجرام) أداة أساسية و فعّالة للكشف عن الأورام فى الثدى. لكن هنالك اسباب تتعلق بشكل ظهور الورم كصغر حجمه ،واسباب أُخرى تتعلق بأخصائى التشخيص كتعرضه للاجهاد البصرى ،تزيد من صعوبة تفسيرصورة الماموجرام ذات التدرجات الرمادية المتباينة، مما يؤدى فى كثير من الأحيان للتشخيص الخاطئ.

ان الهدف الرئيس من هذا البحث هو تصميم برنامج حاسوب يعمل على تشخيص صور أشعة الثدى تلقائياً بغرض تقديم "رأى مساعد" الى اخصائى الأشعة. يتكون البرنامج المطروح من عدة خوارزميات وتعتباراللبنة الأولى للبرنامج هى ازالة المناطق الغير مرغوب فيها من الصورة والابقاء على منطقة الثدى فقط مع تحسين الصورة دون التأثير على البيانات المراد دراستها فى الصورة.

هذه الأطروحة تعمل على تحقيق عدة أهداف ، أهمها تصنيف نوع نسيج الثدى الى دهنى، غُدى، أو كثيف. يليه تصنيف الثدى الى مُعافى أو مريض ومن ثم تصنيف آفة الثدى الى ورم حميد أو خبيث. وتعتمد هذه الدراسة على استخدام نوعين من الميزات النسجية، أحدهما النوع المعروف (Haralick) والآخر الحديث (SFTA). وللتصنيف المعتمد فى هذه الدراسة هو المصنف نوع (SVM).

وبهدف اظهار متانة التحليل و المقاربة ، أجريت معالجة الصور على قاعدة البيانات المعروفة (MIAS)، والتى تتضمن تشخيص مسبّـــق للصور من قِبـل أخصائيى الأشعـــة.

ان أفضل النتائج التى تم الحصول عليها هى دقة بنسبة 87% لتمييز النسيج الدهنى من غير الدهنى ،و دقة بدرجة 78% لتصنيف النسيج الغير دهنى الى غـدى أو كثيـف .وقد أظهر المميز النسجى نوع SFTA كفاه اعلى من نظيره Haralick على مستوى تصنيف الثدى الى سليم وغير سليم وتصنيف الورم الى حميد أو خبيث، حيث حقق درجات دقة 75.5 % و 74.5% على التوالى.

Résume

La détection précoce du cancer du sein est de la plus haute importance pour augmenter les taux de survie. La mammographie est considérée comme une technique efficace pour la détection du cancer du sein mais, en raison de la texture du tissu mammaire et de la variabilité inter observateur, les erreurs de diagnostic du cancer du sein sont fréquentes. Notre recherche consiste à développer un système simple de détection assistée par ordinateur (CAD) afin d'améliorer la précision et l'efficacité de l'interprétation des mammographies. La première étape de l'algorithme proposé est le prétraitement des images de mammographie, afin d'éliminer les régions indésirables de l'image mammaire et de rehausser l'image sans perdre les informations. Pour la suite du processus, la région d'intérêt (ROI) segmentée par seuillage et quantification. En mammographie, le type de tissu mammaire influe sur les performances de la détection. Ainsi, l'algorithme proposé caractérise automatiquement le type de tissu, soit gras, dense ou glandulaire. La dernière étape du système de CAD proposé consiste à distinguer les seins normaux des seins anormaux, puis à différencier les lésions bénignes des lésions malignes. Dans cette thèse, les tâches de classification sont basées sur les caractéristiques de texture et sur les machines à vecteurs supportées (SVM). Des caractéristiques « anciennes » bien connues (caractéristiques de Haralick) ont été extraites ainsi que de nouvelles caractéristiques moins connues (analyse fractale segmentée de texture SFTA). Pour obtenir de bonnes performances de classification, les caractéristiques optimales ont été sélectionnées et les caractéristiques redondantes supprimées. Dans le but de montrer la robustesse de notre approche, des tests ont été réalisées à l'aide de la base de données bien connue Mammographic Image Analysis Society (MIAS), qui comporte des annotations fournies par des radiologues. Pour la caractérisation des tissus, trois approches d'extraction de caractéristiques ont été étudiées; la meilleure précision obtenue est de 87% pour distinguer les tissus adipeux des tissus non gras en utilisant des caractéristiques combinées de Haralick et de SFTA. De plus, la meilleure précision de 78% a été obtenue pour différentier les tissus denses et glandulaires à l'aide de fonctions SFTA. Les caractéristiques de la SFTA ont prouvé leur supériorité pour classifier les lésions normales et anormales et les lésions bénignes et malignes, avec des précisions de 75,5% et 74,5% respectivement.

1. Introduction

This thesis addresses the problem of detection and classification of breast cancer by the application of computer programming tools for augmenting radiologists on their demanding job of diagnostic breast abnormalities using data from mammography. The focus, research objectives of the thesis, and strategies of research are described in this chapter, followed by the thesis statement, the main contribution and thesis organization.

1.1. Thesis focus

Breast cancer is the utmost usual cancer among the women and the most common form of cancer death. The early detection of breast cancer is of prime importance in order to increase survival rates. Although, X-ray mammography is the most common imaging tool for breast screening and for early breast cancer detection (Karellas and Vedantham, 2008), but mammograms are considered among the most difficult medical images to interpret according to: the overlap of tissues, the differences in the tissues types and their low contrasts. Also, it is hard to detect a breast cancer at an early stage, because the tumor is too small to perceive by the eye. Moreover, many researches (Kim *et al.*, 2019), (Elmore *et al.*, 2009), (Ciccone *et al.*, 1992) approved that there is an inter-observer and intra-observer variability of mammogram interpretation due to several factors concerning the radiologist as example: years of experience, fellowship training in breast imaging, and visual fatigue.

So, mammograms interpretation can produce false negatives and false positives. The goal of this research is to develop a CAD system that search abnormal areas and highlights these areas, alerting the radiologist to carefully assess the ROI, providing a valuable “second opinion” to a radiologist based on an objective method.

1.2. Research objectives

The main objective is to develop a reliable CAD system for detecting and localizing the breast cancer in the mammogram image based on texture features. The work can be divided into four specific objectives, corresponding to the main objective:

1. To develop an algorithm that classifies the breast tissue type, the normal/abnormal image, and delineates the breast cancer on the mammogram.
2. To choose the best features that estimate breast texture.
3. To assess the accuracy, sensitivity and specificity of the classifier.

1.3. Thesis statement

This thesis proposes an approach for detecting a breast cancer in mammogram images, the thesis states:

In mammogram image, breast tissue shows different texture appearance according to the tissue density type, normality and pathologically tissue. Estimation of statistical and structural features can be used to develop a CAD system for breast cancer detection.

1.4. Research strategy

This section briefly shows a flowchart that demonstrates the main algorithm criteria of thesis methodology [Figure (1-1)]. The mammogram image preprocess by removing the unwanted regions, noises, and enhancing the image contrast. Two segmentation processes applied, breast region segmentation and ROI segmentation. The algorithm used texture analysis, and selected the best subset of features that describe breast tissue. These textural indices were tested to assess their potential interest into different types of problems: the characterization of different tissue types, the classification between normal and abnormal mammograms, and the differentiation between benign and malignant lesions. To achieve classification tasks, machine learning method was applied, and the performance of the classifiers was tested on an independent dataset.

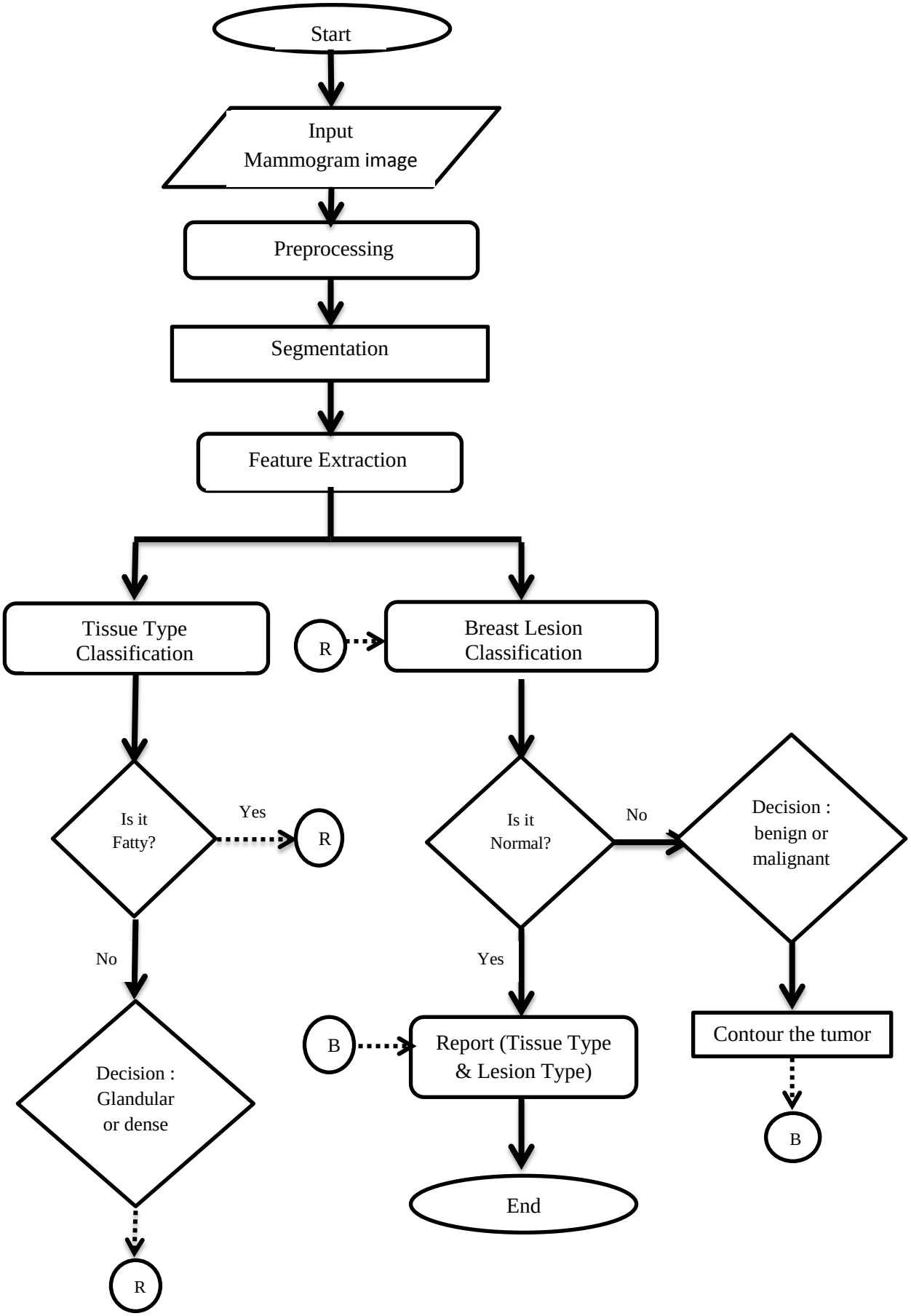


Figure (1-1): Flowchart represents the general methodology of thesis algorithm. R and B are abbreviations refer to (Returned and Back) respectively

1.5. Thesis organization

This manuscript is divided into five chapters, which are organized as follows:

Chapter 1: Introduction

Introduce a brief introduction to the thesis, including the thesis focus, objectives, thesis statement, research strategies, and thesis organization.

Chapter 2: Medical context and instrumentation viewpoint

Present an overview of the medical context and the breast cancer imaging modalities, focusing on mammography and available CAD systems. It consists of four sections. The first section gives a basic overview of the breast anatomy. The second section presents an overview of breast cancer disease, including some elements to stablish its diagnosis, epidemiological data and disease management. A specific focus is given on African and Sudan regions. The third section introduced the X-ray mammography, which is considered as the primary imaging tool for breast cancer screening; it mentions mammography equipment, the different types of mammography, textures in mammogram, and standard mammography reports with examples. The last section introduces the usage of CAD in mammography and highlights the most frequently used mammogram databases.

Chapter 3: Review and performance evaluation of CAD system approaches for mammographic image analysis

Is dedicate to the concept behind the research, including an overview of image processing methods that are applied to X-ray mammogram as well as CAD system approaches. This chapter describes some popular methods, which are used for X-ray mammograms processing. The four sections correspond to preprocessing, lesion segmentation, image feature extraction, and a specific classification tasks.

Chapter 4: Materials and Methods

Present the materials and methods that were used for the current work. The different sections correspond to the database that was used, the proposed processing pipeline, which consists of preprocessing stage, segmentation, feature extraction and classification. Texture analysis can be considers as the central part of this research. Finally, three types of classification tasks were investigated: breast tissue characterization, classification between normal and abnormal breast, and classifications between benign and malignant lesions. The datasets and the procedures, which are used for each task, are finally presented.

Chapter 5: Results and discussion

Presents the results that were obtained from the different steps previously described, including preprocessing, segmentation, and classifications. Finally, is a general discussion and compares the results obtained by other studies related to our approach with our results.

Chapter6: Conclusion and prospective work

The main achievements are summarized for the future work.

2. Medical Context and Instrumentation Viewpoint

2.1. Breast anatomy

Breasts are made up of a mixture of lobules, glandular ducts, fatty and fibrous connective tissues [Figure (2-1)]. Lobules produce milk; ducts are the tiny tubes that carry milk from the lobules to the nipple. The breast lobules are called glandular tissue; the fibrous connective tissue and the fatty tissue give breasts their sizes and shapes and hold the glandular tissue in place. Fibroglandular breast tissue is made up of epithelial glandular component, including terminal ductal lobular units and ducts, as well as stromal components, including the supportive fibrous connective tissue within the inter and intra-lobular stroma (Seely *et al.*, 2004).

The adult breast sites atop the pectoralis muscle and the ribcage. The breast tissue extends horizontally from the edge of sternum out to the midaxillary line. The breast tissue is encircled by a thin layer of connective tissue called fascia, the deep layer of fascia sites on the top of the pectoralis muscle while the superficial layer sits just under the skin (Netter, 2006).

The blood supply from the breast comes primarily from the internal mammary artery, which runs underneath the main breast tissue. The lymphatic vessels flow in the opposite direction of the blood supply and drain into lymph nodes. Most lymphatic vessels flow to axillary lymph nodes, while a small number of lymphatic vessels flow to internal mammary lymph nodes located deep on the breast. When a breast cancer metastasizes, it usually involves the first lymph node in the chain of lymph nodes (Blackburn *et al.*, 2012).

The three major hormones affecting the breast are: estrogen, progesterone, and prolactin. During the menstrual cycle and pregnancy, many changes happen in the women's breast tissue due to hormones, progesterone and estrogens. In a woman who is not pregnant or suckling, the alveoli are very small and solid, but during the pregnancy enlarge, and the cells undergo rapid multiplication. In a woman who has given birth more than twice the breast becomes large and pendulous, while in elderly women the breast becomes small because of the decrease in fat and of the glandular tissue atrophy (Moore *et al.*, 2010).

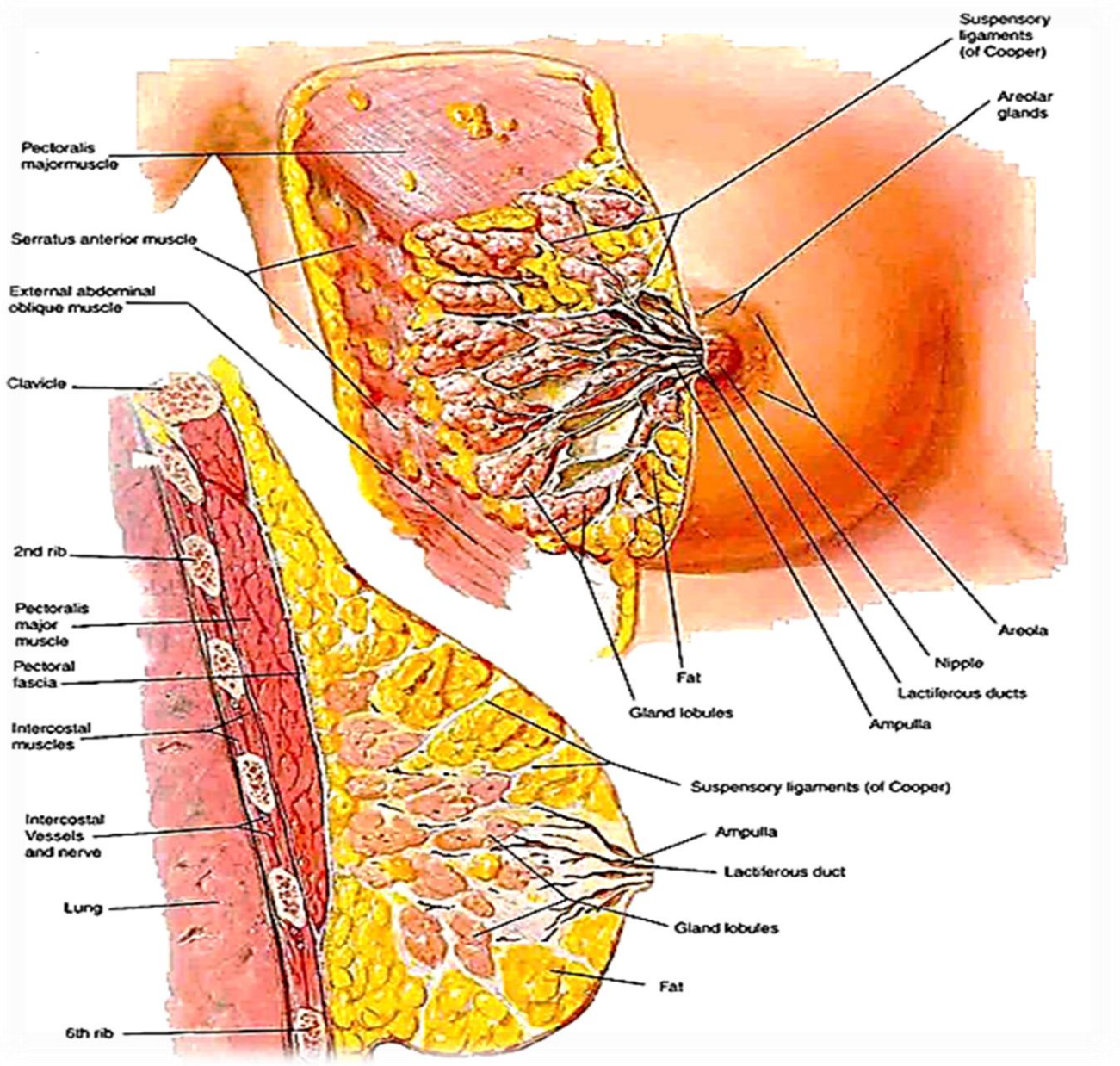


Figure (2-1): Mammary gland, anterolateral dissection and sagittal section, (Netter, 2006)

2.2. Breast cancer

Breast cancer is associated with an image of severe gravity because it affects a body part full of symbolism in motherhood and femininity. It is much more common in women, it can also affect men; breast cancer is the second common leading cancer type, exceeded only by lung cancer [Figure (2-2)]. Breast cancer is a heterogeneous disease, some clinical characteristics being correlated with risk predictors (Garcia *et al.*, 2006).

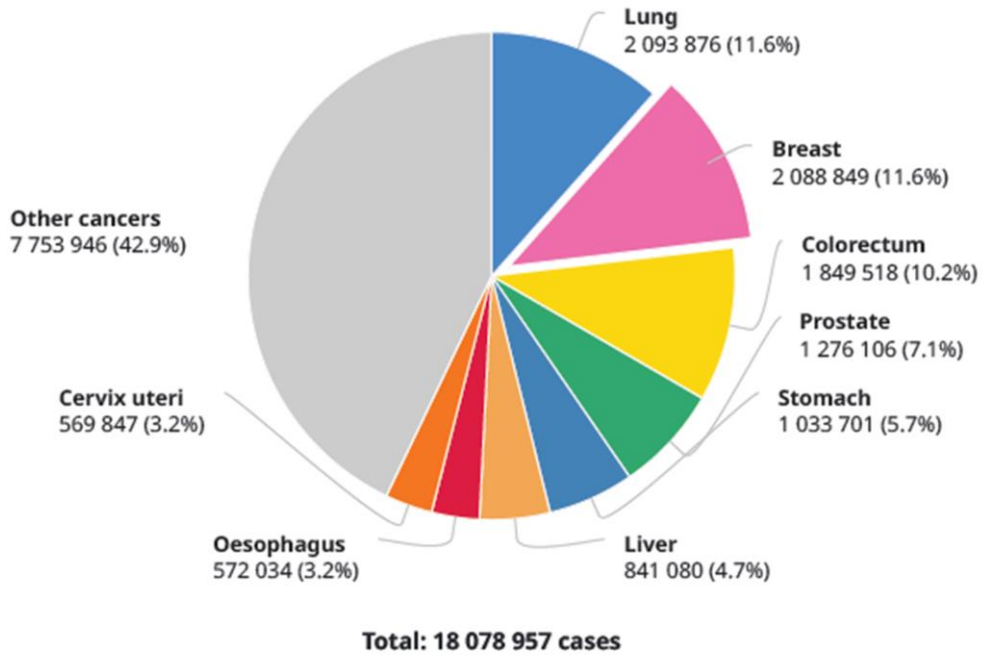


Figure (2-2): Number of cancers' new cases in 2018, both sexes, all ages, (GLOBOCAN 2018)

2.2.1. Breast cancer epidemiology

Breast cancer is the most frequently occurring cancer in women around the world. The incidence of breast cancer varies greatly around the world. According to 2012 GLOBOCAN statistics, nearly 1.7 *million* women were diagnosed with breast cancer with an incidence increase of 18% when compared to 2008 (Ghoncheh *et al.*, 2016). Breast cancer incidence in developed countries is higher (Vanderpuye *et al.*, 2017), but the relative mortality is greater in economically disadvantaged countries, as shown in [Figure (2-3)] and [Figure (2-4)].

According to (GLOBOCAN 2012) statistics, the age standardized incidence rate (ASR) of breast cancer per 100,000 women is: 84.8 in North America, 56.8 , and 38.3 in South and Central America respectively, 92.6in Western Europe, 73 in Northern Europe, 56 in Southern Europe, 49 in Eastern Europe, 26 in South Eastern Asia, 22 in South Central Asia while the incidence range is between 13.5 and 30 in sub-Saharan Africa. In sub-Saharan Africa, the majority incidence estimation of breast cancer is 38.9 and 38.6 cases in Southern and Western Africa, 30.4 in Eastern Africa, and 26.8 in Central Africa (Ferlay *et al.*, 2015). Study in (Chokunonga *et al.*, 2013) estimated the incidence of breast cancer in Africa is to double of (GLOBOCAN 2012) statistics by 2050. [Figure (2-5)] shows the (ASR) of breast cancer per 100,000 women as given by GLOBOCAN 2018:

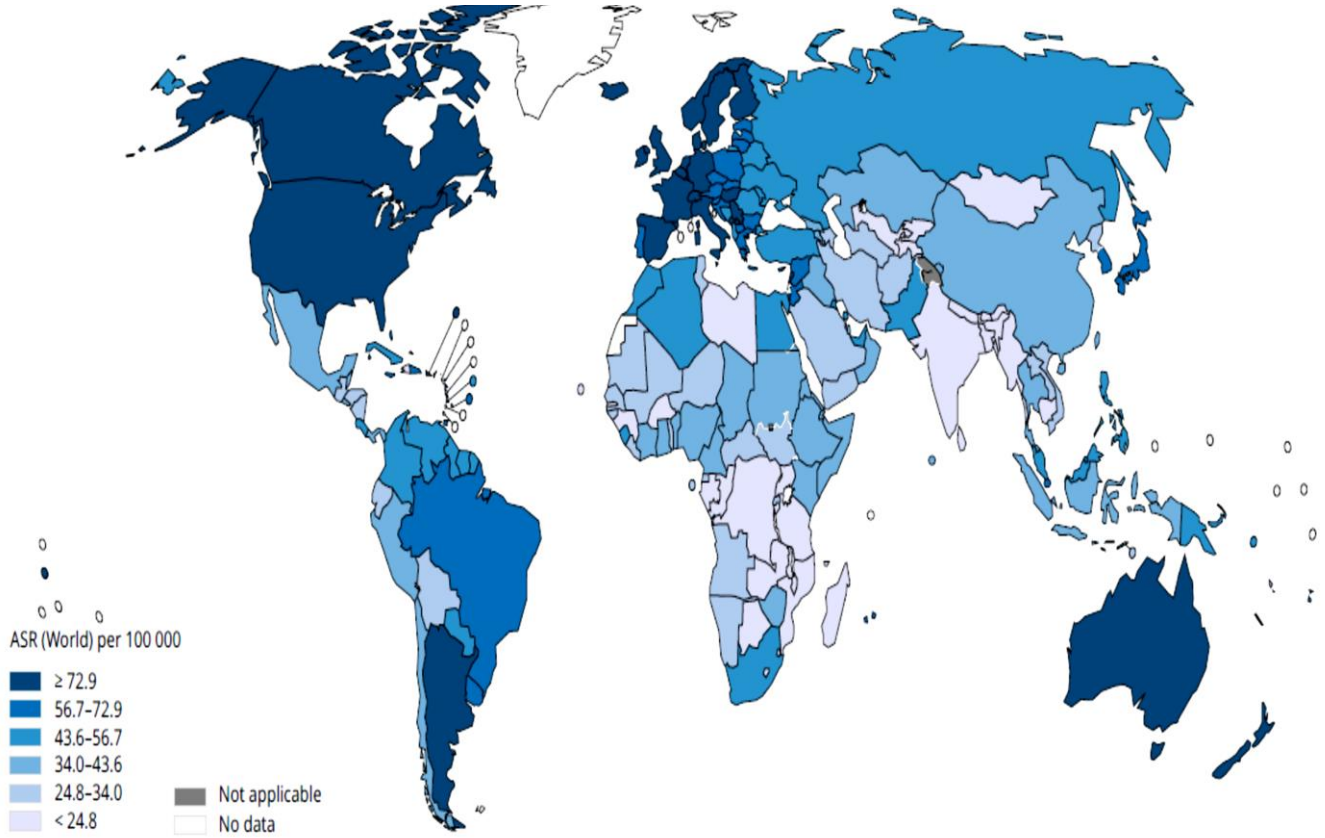


Figure (2-3): Distribution of standardized incidence rate of breast cancer in world (GLOBOCAN, 2018)

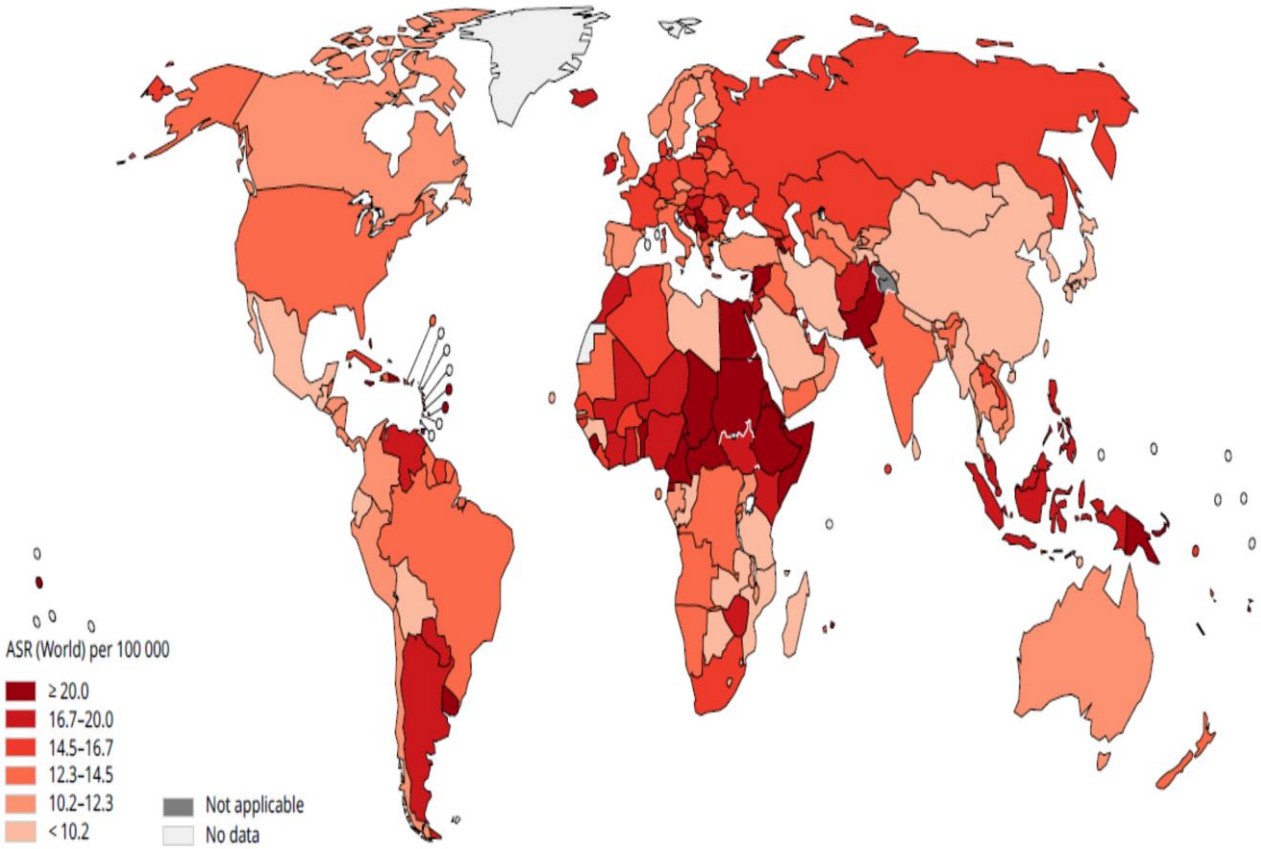


Figure (2-4): Distribution of standardized mortality rate of breast cancer in world (GLOBOCAN, 2018)

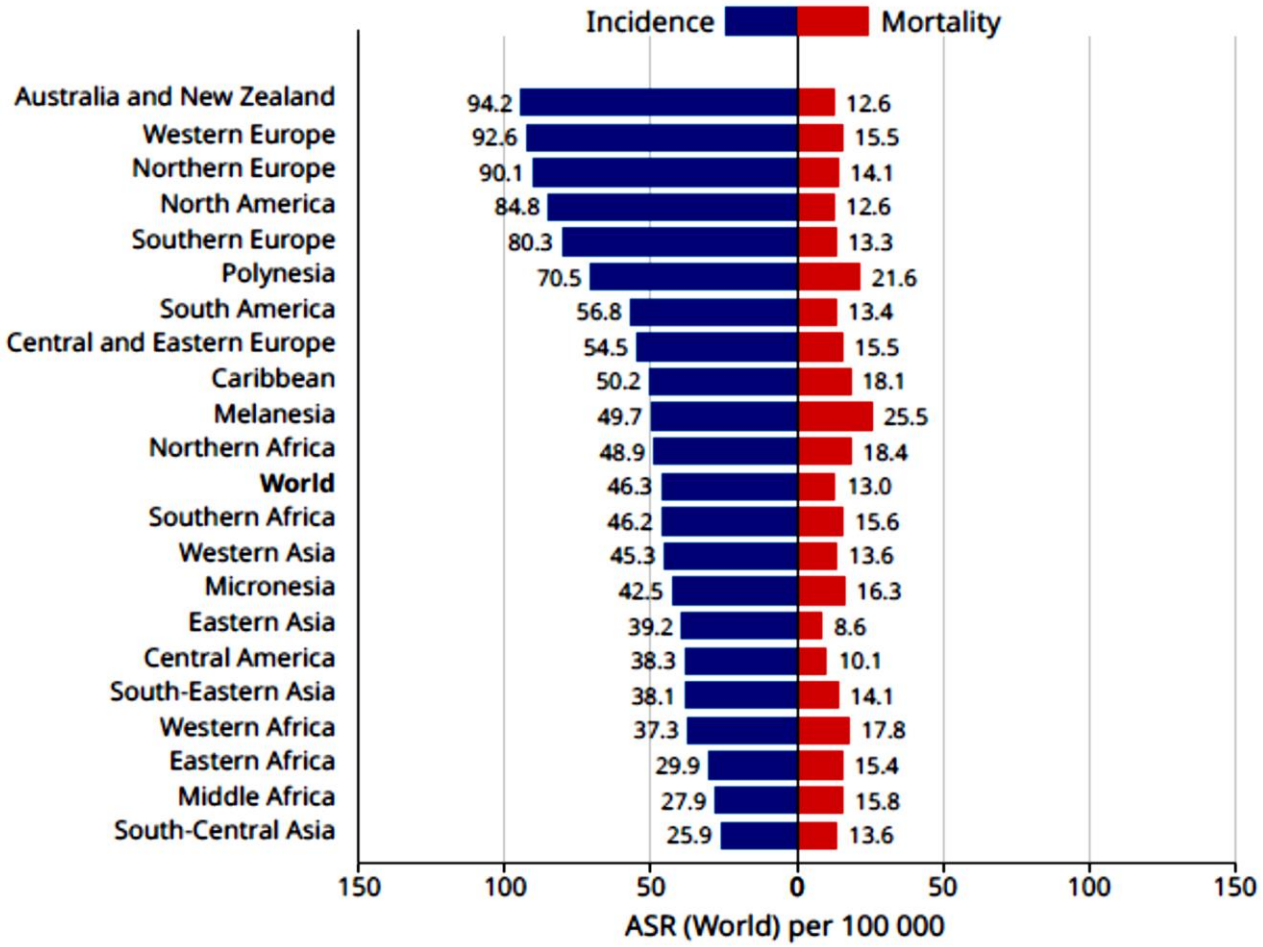


Figure (2-5): Age standardized incidence and mortality rates of breast cancer for female worldwide, (GLOBOCAN 2018)

2.2.2. Breast cancer classifications

Breast cancer can be classified into biological and clinical subgroups according to its histological grade and its histological type. The purpose of this classification is to select the best treatment Grade is an assessment of the degree of differentiation and proliferative activity of a tumor, and histological type refers to the types of tissue involved in the tumor (Elston and Ellis, 1991).

There are two major regions in the breast where cancer can arise, the lobules and the ducts. Ductal carcinomas are the most common type of breast cancer, followed by lobular carcinoma, and other rare cancer types include tubular, medullary, papillary, and mucinous carcinoma. Breast cancer is also classified as invasive or noninvasive. Invasive cancers are those spread to breast tissue outside of the tissue of origin, and noninvasive cancers, which are called in situ carcinoma, i.e. tumors that not spread outside of origin tissue.

2.2.3. Breast cancer staging

The Tumor-Node-Metastasis (TNM) classification system is the most important coding for staging malignant tumors, published by the Union for International Cancer Control (UICC). In addition, the TNM classification system is important in cancer research for a correct description and classification of the anatomical extent of a tumor (Boeker *et al.*, 2016). In TNM classification, (T) describes the size of the tumor; (N) describes whether the cancer has spread to the lymph nodes, and (M) describes whether the cancer has spread to a different part of the body. [Table (2-1)] provides a complete view of this system.

Table (2-1): TNM staging, (Boeker *et al.*, 2016)

Descriptor	Values	Meaning
T	0-4, is, X	Extent of the primary tumour
N	0-3, X	Extent of metastasis in regional lymph nodes
M	0-1	Existence of distant metastasis
Prefix to T, N, M	p, c	Clinical (pre-therapeutical) or pathological (post-surgical assessment)
Suffix to pNn	(mi)	Micrometastasis (< 0.2 cm)
Suffix to pNn	(sn)	Sentinel lymph node metastasis
Suffix to pN0 or pM0	(i+), (mol+)	Isolated tumour cells, positive findings
G	X, 1-4	Histopathological grading
Suffix to T	(m)	Multiple primary tumours at a single side
Prefix to c/ p	y	Assessment during multimodal therapy
Prefix to c/ p	r	Recurrent tumour
Prefix to c/ p	a	Assessment during autopsy
L	X, 0-1	Lymphatic invasion
V	X, 0-2	Venous invasion
Pn	X, 0-1	Perineural invasion
C	1-5	Validity of the assessment, can follow each of T, N, M
R	X, 0-2	Residual tumour

2.2.4. Breast cancer size and grading

Breast tumor varies in shape and size; it is usually expressed in millimeters or centimeters, ranges from $2\ to\ 50\ mm$ (Narod *et al.*, 2013). The smallest lesion that can be felt by hand is approximately$1.5\ to\ 2\ cm$. Although, small cancers have generally a better outcome, size alone doesn't always give the prognosis. For instance, a small cancer can be fast growing while a larger cancer can be slow growing. In case of breast cancer with multiples localization, multi-centric means there is more than one area of breast cancer in different quarters of the breast and multi-focal means more than one area of cancer in one-quarter of the breast. The importance, the size and localization of the cancer in relation to breast size affect the surgery that can be offered; for small cancer, lumpectomy (cancer and margin removed) is recommended, and for larger cancer mastectomy is recommended (Gottlieb, 2000).

Grade is the assessment of how closely cancer cells resemble the cells of the organ in which the cancer originated. As the grade number increases, the resemblance of cancer cells to those of the organ of origin decreases. For grade-1, cancer cells look most like normal breast cells and slow-growing; for grade-2, cells look less like normal cells and are growing faster; for grade-3 cells look different to normal breast cells and are fast-growing. People with grade-3 cancer are more likely offered chemotherapy to help destroy any cancer cells that may have spread as a result of cancer fast growing (Morrow and Krontiras, 2001).

2.2.5. Breast clock and quadrants

Physicians often describe the location of a breast tumor using clock positions. [Figure (2-6)] shows the right and left breasts from physicians' viewpoint. Each breast is divided into horizontal halves, upper and lower, and vertical halves, inner and outer. Each quadrant is then divided into three subparts, mimicking a clock division. Each part is then coded according to the International Classification of Diseases for Oncology (ICD-O). (Fritz *et al.*, 2013)

The (ICD-O) code is a dual classification, taking into account both topography and morphology. Topography code describes the anatomical site of origin of the neoplasm. The code has a prefix "C" followed by three digits that indicate the site (two digits) and the subsite (one digit), separated by a decimal point. Morphology code describes the characteristics of the tumor, including its cell type and biological activity. The code is

composed of four digits that indicate the cell type and one digit that indicates the behavior, separated by "/". The behavior digit can be 0 (benign), 1 (uncertain behavior), 2 (carcinoma in situ), 3 (malignant, primary site), 6 (malignant, metastatic site), or 9 (malignant, uncertain whether primary or metastatic site).

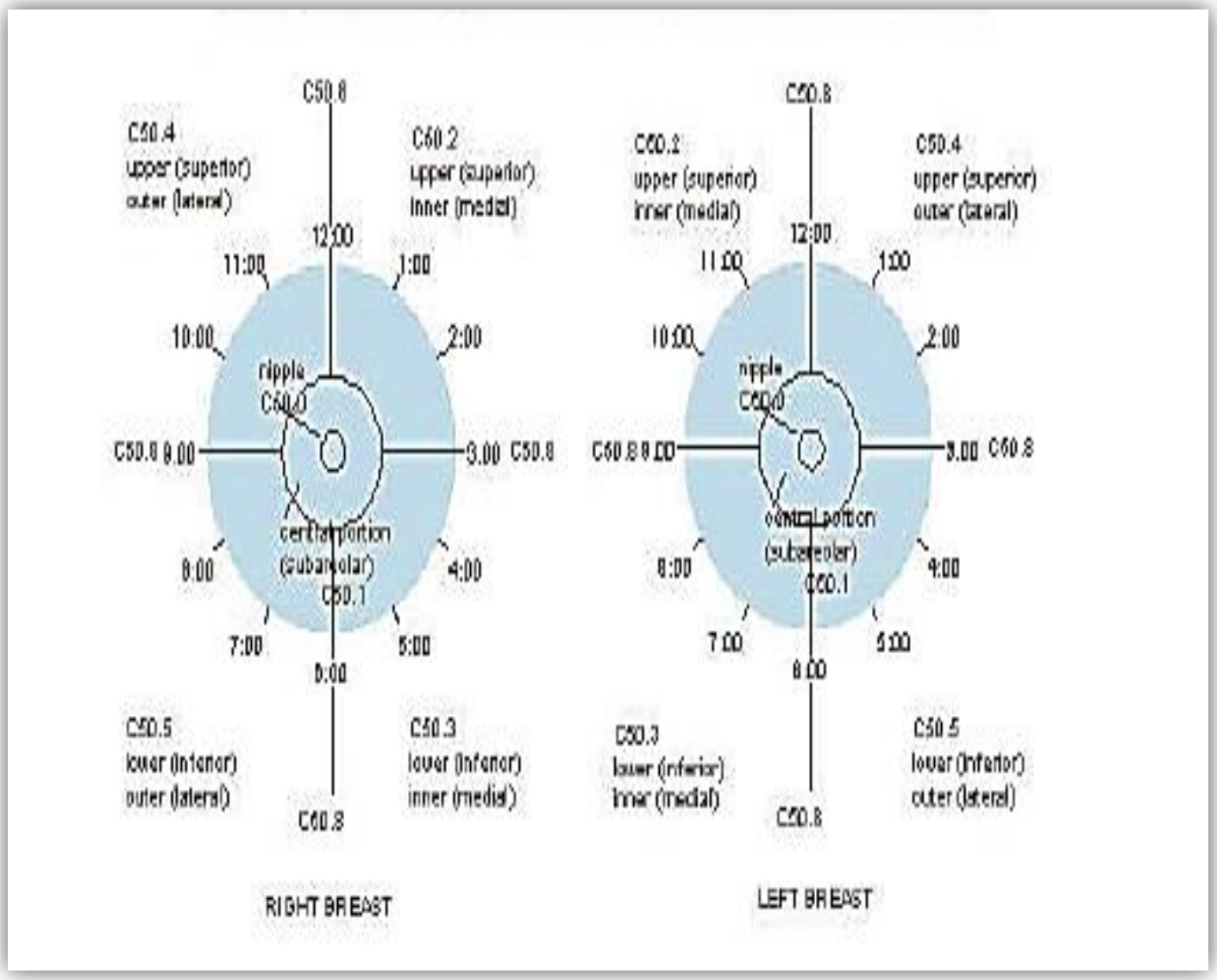


Figure (2-6): "Clock" positions Quadrants and ICD-O codes of the breast, (Fritz *et al.,* 2013)

2.2.6. Breast cancer in lymph nodes

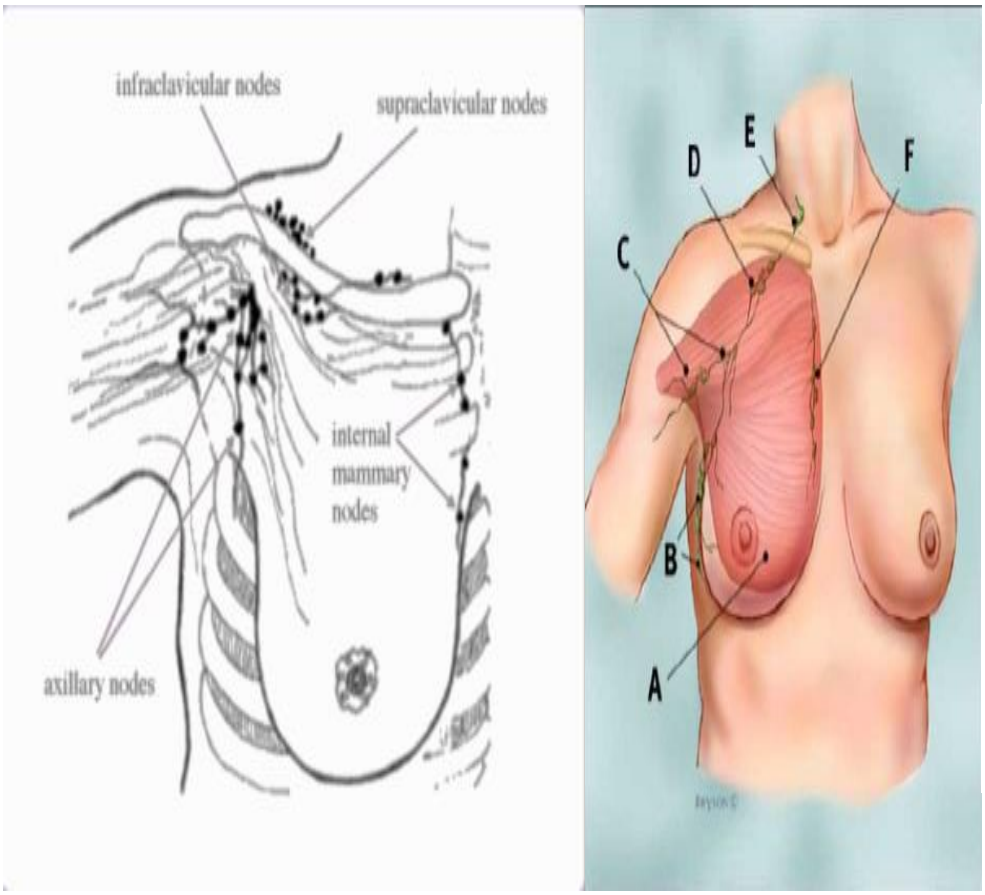
Breast tissue is drained by lymphatic vessels that lead to axillary nodes (which lie on the axilla) and internal mammary nodes (which lie along each side of the breast bone). If breast cancer spreads, the lymph nodes (axillary lymph nodes) are the first place where the spreading occurs, [Figure (2-7)].

Lymph node status is highly related to prognosis, the more lymph nodes are affected the poorer prognosis tends to be:

(1) Lymph node-negative means the axillary lymph nodes do not contain cancer

(2) Lymph node-positive means the axillary lymph nodes contain cancer.

Furthermore, the larger the tumor is, the more likely breast cancer has spread to lymph nodes. Thus, lymph node status is partially related to tumor size.



B, C, D are axillary lymph nodes level I, II, III respectively. (levels according to breast distance).

A: Pectoralis major muscle;
E: Supraclavicular nodes,
F: Internal mammary lymph nodes.

Figure (2-7): Lymph nodes for breast, (SEER, 2001)

2.2.7. Breast cancer risk factors

The biggest risk factors for breast cancer are age and genes (family history). The two most common genes involved in familial breast cancer are BRCA1 and BRCA2. A mutation of one these two genes is the cause of most familial disease cases. For women before age45 $years$, who carry a BRCA mutated gene; have a risk of developing breast cancer

between 60% and 80% (Malone *et al.*, 1998). Moreover, high breast density on an X-ray mammogram is linked to an increased risk of breast cancer (Freer, 2015), extremely dense breasts and first degree relatives with breast cancer were each associated with a more than $2-fold$ increased breast cancer risk in women with age between $40-49$ (Nelson *et al.*, 2012).

Researchers have also found a relation between an overweight and breast cancer. Obese women after menopause have higher risk of breast cancer, due to highly estrogen levels that are produced by fatty tissues, and highly insulin levels. Moreover, obese women–before menopause with body mass index greater than $30kg/m2$ have a 42% higher risk of developing triple-negative breast cancers compared with non-obese women (Pierobon and Frankenfeld, 2013). To assess the correlation between body mass index (BMI), menopausal status, and breast cancer, a meta-analysis of 34 cohort studies including over $2.5\,million$ women was conducted and concluded that each $5\,kg/m2$ increase in body mass index was associated with a 12% increased relative risk of postmenopausal breast cancer, whereas each $5\,kg/m2$ increase in body mass index was associated with an 8% decreased relative risk of premenopausal breast cancer (Renehan *et al.*, 2008)).

Lacey in (Lacey *et al.*, 2009), evaluated risk factors for breast cancer ($N = 2085$) among $70{,}575$ women who were randomized in the prostate, lung, colorectal and ovarian cancer screening trial. The adjusted relative risks (RRs, with 95% confidence intervals (CIs) were calculated for lifestyle and reproductive factors during an average of five years. Result showed, increasing age, nulliparity, positive family history of breast cancer and use of menopausal hormone therapy were strongly positively correlated with breast cancer. In addition, younger ages of menarche or menopause older were weakly associated with breast cancer due to prolong hormone exposure. Finally, there is week positive association between breast cancer and taller heights or heavier weight. [Table (2-2)] extracted from this study summarizes the different risk factors.

Globally, more than $210\,million$ women use either hormonal contraceptive pills or injectable contraceptives (Urban *et al.*, 2012), and there is several researchers studied whether hormonal contraception is associated with an increased risk of breast cancer. In 1996 the Collaborative Group on Hormonal Factors in Breast Cancer meta-analysis pooled a total of 54 epidemiological studies from 25 countries and included $53{,}297$ women with breast cancer and $100{,}239$ controls (Bjelic-Radisic and Petru, 2010). Study concluded, women

using hormonal contraceptives had an elevated risk for breast cancer ($relative\ risk = RR\ 1.24$). Women who had used contraceptives before the age of20, had an elevated risk for breast cancer over the subsequent years ($RR\ 1.95$ until the $30th$ year of age, $RR\ 1.54$ between 30 and 34 years, and $RR\ 1.27$ between the age of$35\ and\ 40\ years$, respectively) compared to those who started to use contraceptives after 20 years of age.

In South Africa, an injectable hormonal contraceptive has been used more commonly than anywhere else in the world (Bailie *et al.*, 1997). In (Shapiro *et al.*, 2000), a case-control study was conducted, in which cases of breast cancer ($n = 419$) and controls ($n = 1{,}625$) hospitalized for conditions unrelated to contraceptive use were interviewed from 1994 to 1997 in hospitals in Cape Town, South Africa. The women were aged $20 - 54$ years. The results showed, the relative risk for exposure to injectable contraceptives was 0.9 without associations of age categories. For exposure to oral contraceptives, the overall relative risk was 1.2. The findings suggested that injectable contraceptives do not increase the risk of breast cancer, and that oral contraceptives may increase the risk among women below age 35 years.

A comparison study between North Africa and Western countries (Corbex *et al.*, 2014), suggested that women in North Africa are more protected from breast cancer due to many factors such as younger mean age at first pregnancy, higher mean number of children, longer mean duration of breastfeeding, lower mean age at menopause, lower prevalence of contraceptive use and lower alcohol consumption.

Table (2-2): Breast cancer and risk factors by (Lacey *et al.*, 2009)

	Breast Cancers	Woman-Years	Crude Rate	RR	95% CI
Attained Age					
55–59	274	63,088	434.31	1.00	Ref.
60–64	608	108,640	559.65	1.36	1.18–1.57
65–69	582	104,433	557.30	1.44	1.24–1.67
70–74	434	79,134	548.44	1.48	1.26–1.73
75+	187	34,419	543.30	1.47	1.21–1.79
Race					
White	1,854	345,926	535.95	1.00	Ref.
African-American	89	19,765	450.29	1.05	0.84–1.32
Asian/Pacific Islander	111	17,230	644.23	1.14	0.86–1.51
Other/unknown	31	6,793	456.35	0.92	0.64–1.32
Family history of breast cancer					
No	1696	336,410	504.15	1.00	Ref.
Yes	389	53,304	729.78	1.44	1.29–1.60
Number of Live Births					
0	240	35,545	675.20	1.00	Ref.
1	160	28,296	565.45	0.70	0.55–0.89
2	524	88,510	592.02	0.76	0.62–0.92
3	526	96,163	546.99	0.75	0.62–0.91
4	331	66,673	496.45	0.72	0.59–0.88
5+	301	73,888	407.37	0.65	0.53–0.80
Age at First Birth					
None	240	35,545	675.20	1.00	Ref.
< 20	267	62,165	429.50	0.68	0.53–0.86
20–24	901	181,709	495.85	0.74	0.59–0.91
25–29	471	80,523	584.93	0.83	0.66–1.03
30–34	148	20,822	710.79	1.03	0.81–1.31
35+	47	6,894	681.75	1.02	0.74–1.41
Age at Menarche					
< 12 years	457	76,987	593.61	1.00	Ref.
12–13 years	1099	209,082	525.63	0.86	0.77–0.97
14–15 years	439	84,595	518.94	0.85	0.74–0.97
≥ 16 years	88	17,874	492.34	0.81	0.65–1.02
Age at Menopause					
< 45 years	123	25,254	487.05	1.00	Ref.
45–49 years	340	64,113	530.31	1.07	0.87–1.31
50–54 years	700	126,080	555.20	1.12	0.92–1.35
55+ years	211	32,192	655.44	1.29	1.03–1.62
Surgical menopause	574	122,386	469.01	0.84	0.69–1.02
Radiation/medication	94	12,061	779.37	1.32	1.01–1.74
Menopausal hormone therapy use					
Never	571	134,329	425.08	1.00	Ref.
Former	280	64,773	432.28	1.02	0.88–1.18
Current, < 5 years	372	61,931	600.67	1.44	1.26–1.66
Current, ≥ 5 years	847	125,006	677.57	1.67	1.49–1.87
Weight (kg)					
< 60	437	82,961	526.75	1.00	Ref.
60 – 64.9	310	59,068	524.82	1.03	0.86–1.19
65 – 69.9	311	56,091	554.46	1.18	0.97–1.30
70 – 74.9	257	49,325	521.03	1.08	0.92–1.26
75 – 79.9	267	47,072	567.22	1.21	1.03–1.42
≥ 80	486	91,892	528.88	1.20	1.04–1.38
Height (m)					
< 1.60	527	101,460	519.42	1.00	Ref.
1.60 – 1.64	605	117,236	516.05	1.01	0.89–1.13
1.65 – 1.69	561	102,205	548.90	1.06	0.94–1.20
≥ 1.70	385	66,707	577.15	1.11	0.97–1.27
BMI (kg/m^2)					
< 18.5	25	5,135	486.85	0.88	0.59–1.32
18.5 – 24.9	848	154,447	549.06	1.00	Ref.
25 – 29	712	133,799	532.14	1.06	0.95–1.17
30 – 34.9	305	59,109	516.00	1.10	0.97–1.26
≥ 35	173	32,478	532.67	1.21	1.02–1.43

2.2.8. Breast cancer incidence and epidemiology in Sudan

Sudan is the third largest country in Africa, located in the Northeast of Africa with more than 40 $million$ people; Nile River is the most noticeable geographical feature (Wikipedia, 2012). Sudan has diverse ethnic groups including Afro, Arab and Afro-Arab tribes. Sudan's cancer burden is not well documented because the health system focuses on communicable or infectious diseases such as malaria, tuberculosis and human immunodeficiency virus (HIV).

In 2009, the first National Population based Cancer Registry (NCR) was established in Sudan. The first statistic data from the NCR ($number\ of\ cases\ =\ 6771$) for the capital of Sudan (Khartoum State, [Figure (2-8)] for the period 2009-2010 indicate that breast cancer is the most commonly cancer in women with an incidence rate of $25.1\ per\ 100{,}000$, [Figure (2-9)] (Intisar *et al.*, 2014)

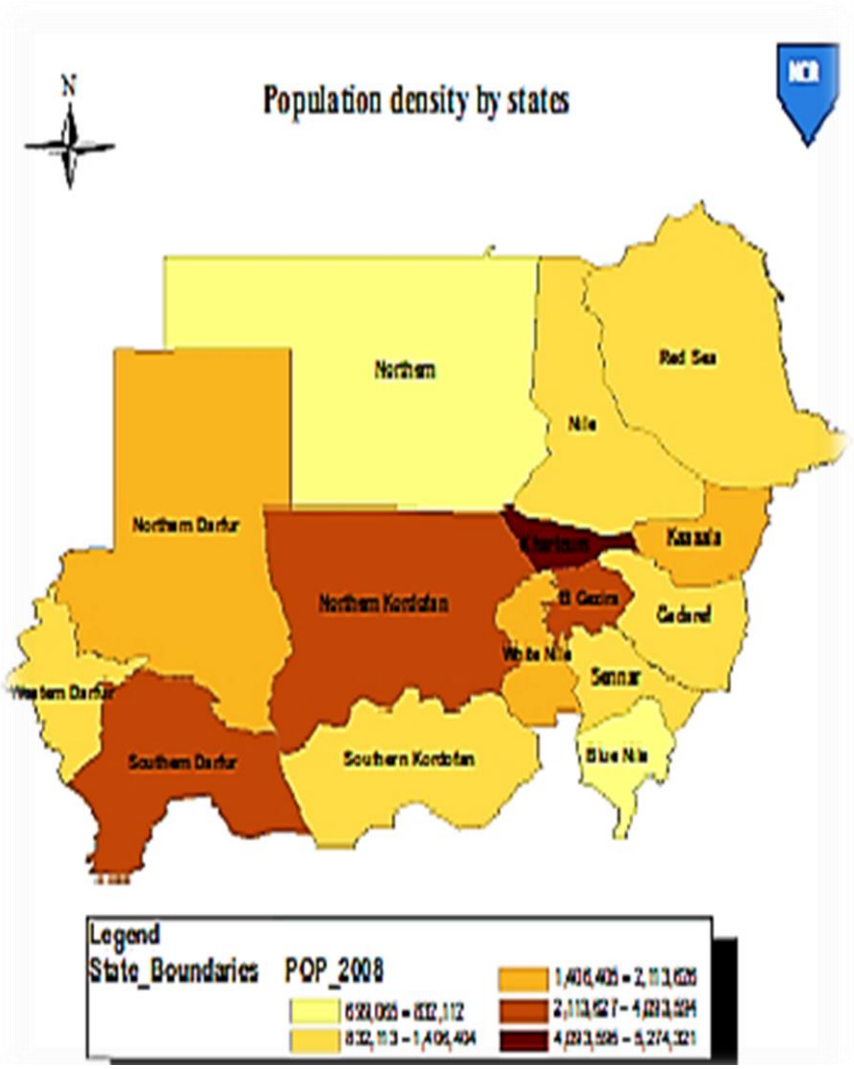


Figure (2-8): Sudan States population distribution. (NCR: 2009-2010). (Intisar *et al.*, 2014)

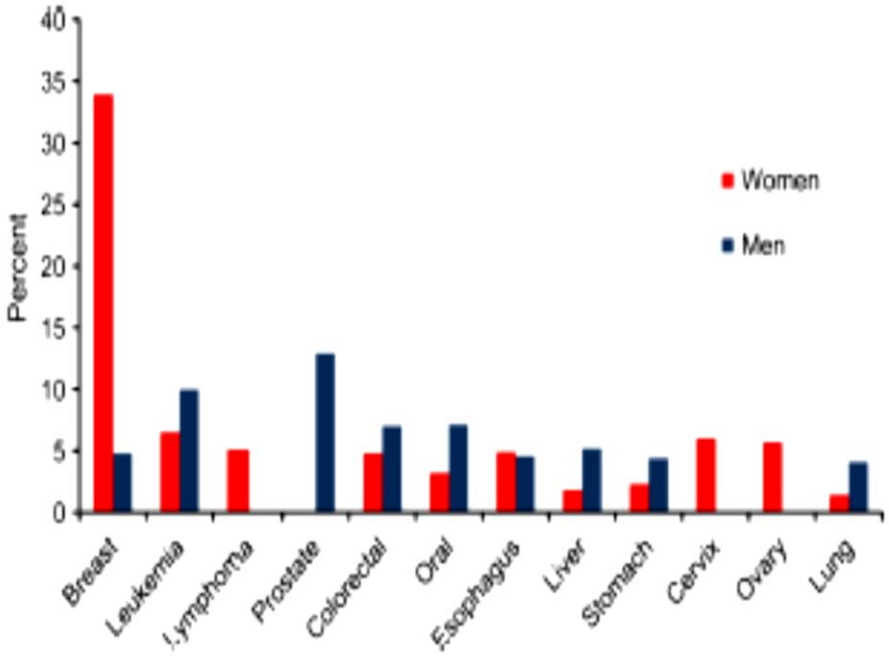


Figure (2-9): Most common cancer by gender in Khartoum, Sudan, (NCR: 2009-2010). (Intisar *et al.*, 2014)

A descriptive study (Elgili *et al.*, 2010) focused on the type, stage and age distribution of breast cancer in women living in central Sudan including al Gezira, Blue Nile, White Nile, and Sennar States. This study was conducted on 1255 women by the Institute of Nuclear Medicine, Molecular, Biology and Oncology, from 1999 to 2006. Results showed that 74% of the women were less than age $50\ years\ old$ or premenopausal, the most common pathology was invasive ductal carcinoma(82%). The majority of women presented stage III or higher stages tumors that had already metastasized, while ductal carcinoma in situ was the least prevalent(0.5%), [Figure (2-10)].

The variation and clinical significance of hormone receptors, estrogen (ER), progesterone receptors (PR), and human epidermal growth factor receptors-2 (HER2) varies from region to region. To evaluate the ER, PR, and HER2 biomarkers, a study were conducted in North-eastern Africa (Sudan and Eritrea) (Sengal *et al.*, 2017). Data were assessed in consecutive female who had been diagnosed with invasive breast cancer from 2011 to 2015 in Gezira University Pathology Laboratory (Sudan) and National Health Laboratory (Eritrea). The study included 678 women, a total of 562 Sudanese cases, and 116 Eritrean women. Results

found that more than 54% of women with breast cancer were ER negative and 62% were PR negative. These percentages are remarkably lower than percentage reported from studies coming from West Africa, Tunisia, and Uganda, but much higher than the ones reported for Egypt, Ethiopia, South Africa and Caucasian women in the west. Also, younger women ($age \leq 50\, years$) are more likely to develop ER negative breast cancer ($p = 0.039, OR = 1.5, 95\%\, CI\ 1.02 - 2.15$) than older and poorly differentiated ($grade\ III$) tumors are more strongly associated with ER negative breast cancer than $I\ or\ II$ ($p = .000, OR = 2.2, 95\%\, CI\ 1.5 - 3.2$).

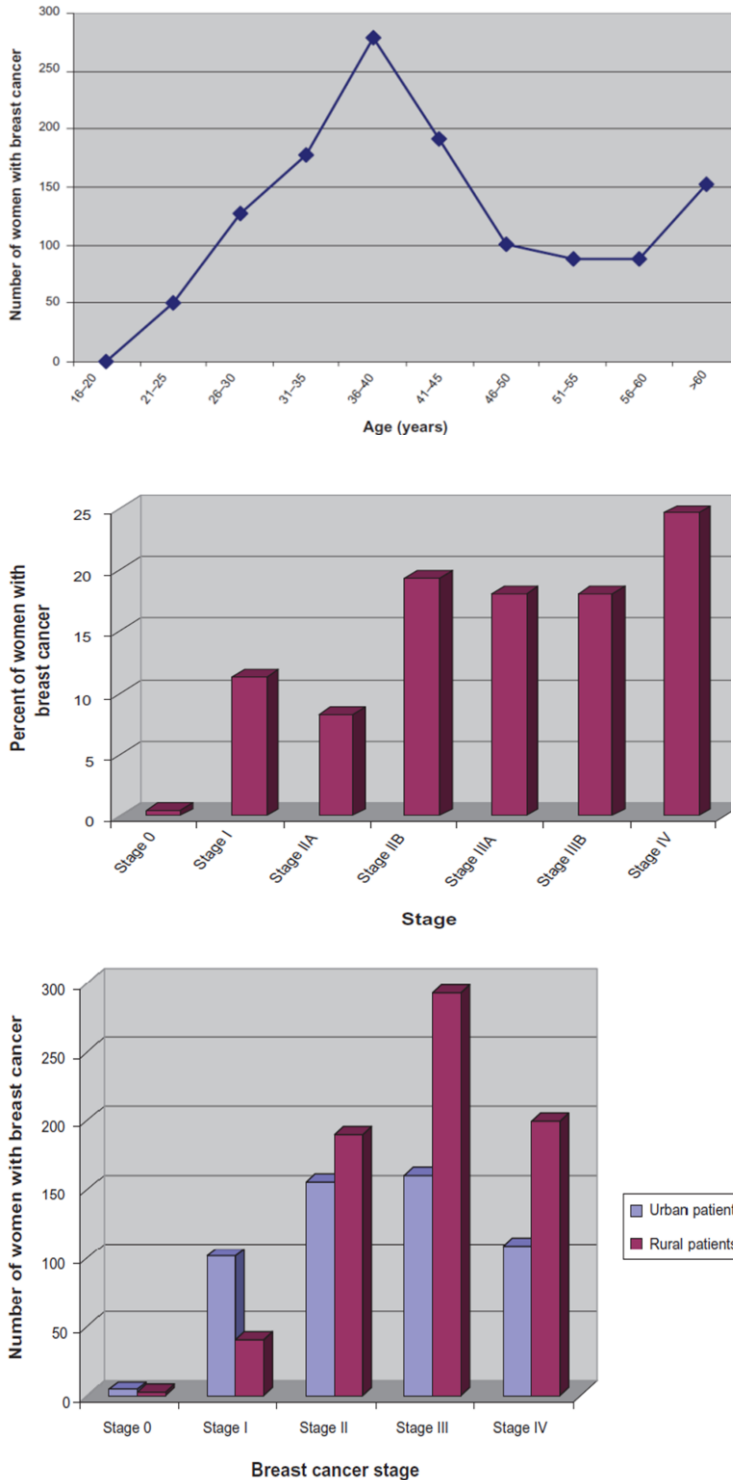


Figure (2-10): Breast cancer in central Sudan (N=1255), according to age distribution, stages, and stage by residency. (Elgili *et al.*, 2010)

2.2.9. Breast cancer management and prevention

In developed countries, women with breast cancer are diagnosed at an early stage because of disease because of established screening and awareness programs, while women in Africa are diagnosed at later stages. Most often they require radical mastectomy to control the disease. Mastectomy rate is 30% in Europe, while it increases to 85% in Africa (Bhikoo *et al.*, 2011).

For breast cancer management, a recent study published in (Vanderpuye *et al.*, 2017) was devoted to research papers related to breast cancer in Africa from 1999 to 2016. This work was based on searches using electronic databases (PubMed, Medline, and Journal online). Among 1320 articles, the majority publications originated from North and West Africa(60%), East Africa(10%), Central and South Africa(12%). Comparing with developing countries, Africa showed a slow progression of improved outcomes for breast cancer due to the inadequacy of health care infrastructure, poverty, limited expenditure of health budget on cancer, late diagnosis, lack of continued educations and awareness program. This study recommended the creation of an African alliance to improve breast cancer care.

Education is a very important factor of breast cancer prevention, and study in (Okobia *et al.*, 2006) showed that participants with higher education were 3.6 $times$ more likely to practice Breast Self-Examination (BSE) program than women with lower education levels. Socio- economic issues are also important, so that all women can access to medical care from screening to advanced treatment (Tao *et al.*, 2015).

2.2.10. Breast cancer detection modalities

In complement to palpation, many imaging techniques can be used to diagnose breast cancer as early as possible. Although mammography is the gold screening modality for breast cancer detection (Berry *et al.*, 2005) , but there is another commonly used breast cancer detection modalities as, breast ultrasound (US), computed tomography (CT-scan), magnetic resonance imaging (MRI), breast thermography, optical imaging, positron emission tomography (PET), scintimammography, and electrical impedance based imaging.

Breast ultrasound

Ultrasound is a safe and widely available method for breast imaging, is used to detect breast cancer and the location of suspicious lesion. Ultrasound imaging based on the application of sound waves, the ultrasound transducer directs high frequency sound waves into the breast tissue and detects the reflected sound waves, and reflected echoes forwarded as electronic signal to a computer system that finally generate the 2D image (Ziskin, 1993Advancement in US technology includes 3D breast US that transduces the sound wave into 3D images (Leproux *et al.*, 2010). Automated whole-breast ultrasound (AWBU) approved by the Food and Drug Administration (FDA) in 2009 is a potential option for providing breast US screening on a widespread basis (Kaplan, 2014). Doppler US (Cura *et al.*, 2005), and elastography (Raza *et al.*, 2010) are especial tools that may be available on US machine.

CT- scan

A CT scan is an X-ray technique, which gives information of internal organs by capturing image slices of the examined body parts. To increase the contrast of CT images, iodinated contrast media is injected. The injected media enhances the visualization of tumors (Sree *et al.*, 2011).

Magnetic resonance imaging

Breast MRI is a widely used imaging modality for the early detection of breast cancer. Moreover, it is used for screening young women with known increased risk of breast cancer due to gene mutations. MRI uses a combination of a magnetic field and radio waves to change the alignment of hydrogen nuclei which is abundant in water and fat, in order to create high contrast images of the breast. MRI has better resolution and less operator dependence than ultrasound, but MRI has high cost than mammogram and ultrasound (Elmore *et al.*, 2005).

Positron emission tomography

Positron emission tomography considers as one of a nuclear medicine technique. Nuclear imaging produces functional images based on molecular properties; this technique shows promise in detecting breast cancer especially in high risk patients (Nover *et al*., 2009). PET uses glucose metabolism to detect cancer. Although many types of PET radiotracers have been developed, ^{18}F-fluorodeoxyglucose (^{18}F-FDG) is the most widely used and the only FDA approved PET tracer. FDA approved PET radiotracer (Schuster, 2015).

Scintimammography

Scintimammography is a nuclear medicine imaging technique uses a radioisotope to visualize lesions of the breast; it is suitable for dense breast, breast with implants, large and palpable abnormalities, and recommended when multiple tumors are suspected (Munshi, 2008).

Breast thermography

Breast thermography is a non-invasive, painless, with no radiation involved, low cost. Thermography is a sensitive screening tool because it is able to diagnose breast cancer at least ten years in advance (Sree *et al.*, 2011). Thermography technique is based on the principle that cancerous and pre-cancerous tissues have a higher metabolic rate resulting in growth of new blood vessels supplying nutritious to the fast growing cancer cells (EYK, 2009).

Optical imaging

An optical imaging technique based on the optical properties variation between the normal and diseased tissues. Optical imaging provides a functional imaging approach with decent spatial resolution and contrast (Godavarty *et al.*, 2015). It uses near-infrared light as a continuous-wave, frequency, or pulse signals. Optical imaging has been utilized in detecting the abnormal tissue, and distinguishing between benign and malignant tumors (Erickson *et al.*, 2011)

Electrical impedance based imaging

This technique based on the tissue impedance to the flow of electric current, studies have shown that malignant tumors have lower electrical impedance than surrounding normal tissues. Electrical impedance tomography (EIT) and electrical impedance mapping (EIM) are the two types of electrical impedance based imaging techniques (Zou *et al.*, 2003).

2.3. X-ray mammography

X-ray mammography is a specific imaging system that uses a low dose X-ray system for the examination of the breast. Mammography is considered as the best available method for the early detection of breast cancer, particularly in the case of small or non-palpable lesions. The first peer-reviewed paper of use mammography to diagnose breast cancer was written by Egan in 1960 (Egan, 1960) and he reported results in imaging the breasts of 1000 patients. In 1964, Wolfe obtained a used xeroradiography system and started to experiments of breast imaging, Wolfe made the first presentation at the Fifth Mammography Conference at Emory University, Atlanta, in 1966 (Kalaf, 2014). In 1980 a large number of clinical researches published on screening mammography performance, including reports of the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) program. These researches led to the program initiated by the American college of Radiology (ACR), which proposed the Mammography Quality Standards Act in 1992(Burnside, 2009) .

2.3.1. Conventional mammography equipment

Mammography is uses low amplitude and high current X-rays. X-ray is an electromagnetic radiation with high energy, wavelength in the range of 10^{-12} and $10^{-8}m$ corresponding to frequencies between 10^{16} and10^{20}Hz, these characteristics allow their penetration in objects and bodies (Bronzino, 2000).

Mammography unit consists of an X-ray tube and an image receptor mounted on opposite sides of a mechanical assembly. For different views, the assembly can be rotated about a horizontal axis, and assembly elevation can be adjusted for different heights of patients, as shown in [Figure (2-11)].

From the X-ray tube, the radiation passes through a metallic spectral shaping filter, a beam that define aperture and a plastic plate, which compresses the breast on to the breast support platform. X-rays transmitted through the breast and the breast support is incident on an anti-scatter grid, then on the image receptor, where they interact and deposit most of their energy locally. In screen film and cassette based digital mammography systems, a fraction of the X-rays passes through the receptor without interaction and these impinge upon the sensors of the automatic exposure control (AEC) mechanism of the mammography unit, any remaining primary X-rays being attenuated by a primary beam stop (Webster, 2006).

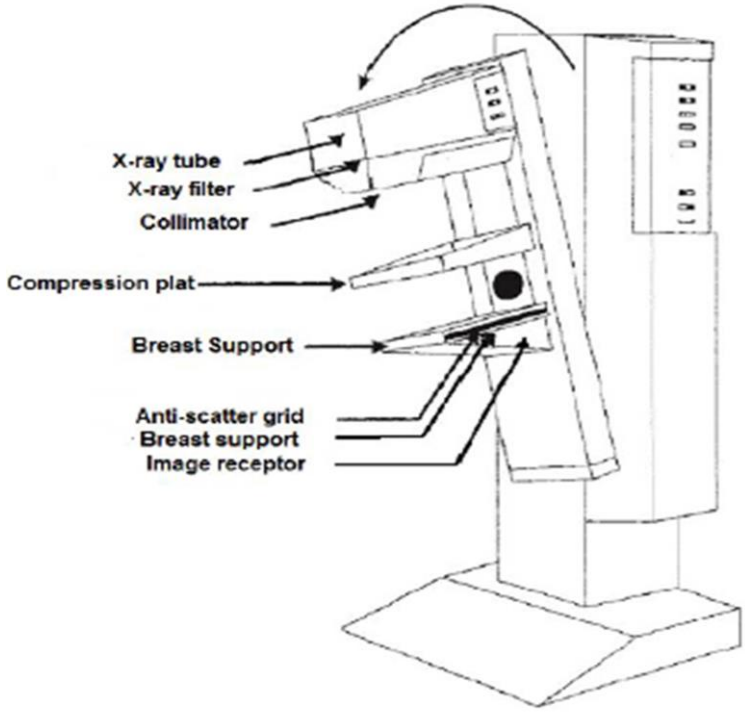


Figure (2-11): Schematic of a mammography unit, (Bronzino, 2000)

2.3.2. Screening and diagnostic mammography

There are two types of mammographic examination, screening and diagnostic mammography. Screening mammography refers to examination of asymptomatic women to detect clinical breast cancer; it is applied to a large mass of women while the diagnostic mammography is applied to an individual patient who has symptoms or abnormal findings. Screening mammographic examinations should be performed annually on women with genetic mutations or significant family history of breast cancer. The recommended age of mass screening in the United States is 40 (Lee *et al.*, 2010), it is between 40 and 50 in

Europe. Diagnostic mammography is generally performed using multi views images, at different magnification levels and with varying compression techniques, and includes online review of the images and physical examination (Barlow *et al.*, 2002).

2.3.3. Full-field digital mammography and screen-film mammography

Full-field digital mammography (FFDM) and screen-film mammography (SFM) are the currently two available technologies for breast cancer screening. In January 2000, FFDM received FDA approval for screening and diagnosis of breast cancer for the same clinical indications as film based mammography (Gold *et al.*, 2012). FFDM is based on digital detectors instead of X-ray films to improve some of the inherent weaknesses of SFM, such as the lake of contrast. Several researches were conducted to compare radiologists' interpretation on FFDMs and SFMs, globally there was no significant difference between FFDM and SFM neither in the detection nor in the classification of the breast cancer, however, FFDM is more accurate in women with dense breasts (Lewin *et al.*, 2001) (Lewin *et al.*, 2002) (Pisano *et al.*, 2005) (Souza *et al.*, 2013).

2.3.4. Types of views

Breast positioning is critical, improper positioning may lead to exclusion of parts of breast from the field of view, so risking the non-detection of a cancer. To reduce the impact of the breast thickness during imaging procedure, the breast is compressed. This ensures better penetration, reduction of scatter, and reduction of motion artifacts. Routine evaluation includes two views of each breast, a craniocaudal (CC) view, and a medio-lateral oblique (MLO) view [Figure (2-12)]. In the CC projection, compression is applied from the top of the breast toward the caudal surface, and in the MLO projection, a compression is applied sidewise from the center of chest wall toward the outer surface of the breast position at an angle varying between$45\ and\ 60\ degrees$. The high (bright) values in the image represent high absorption of X-rays and the low (dark) values represent low absorption of X-rays (Bushberg *et al.*, 2002).

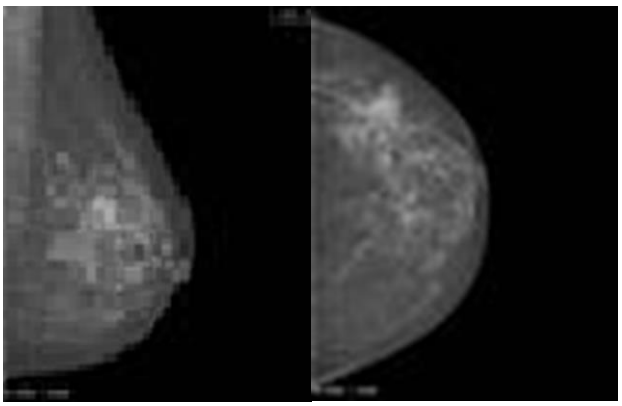


Figure (2-12): Mammography standard views, MLO view and CC-view (from *left to right*), (Moreira *et al.*, 2012)

2.3.5. Mammogram image quality

The image formation depends on the structures densities that are penetrated with the X-rays, since their absorption depends on the structures densities. Mammography provides an excellent spatial resolution ($0.1\ mm/pixel$) and good contrast sensitivity to allow visibility of breast abnormalities (Karssemeijer *et al.*, 1993). The latitude of the image must be adequate to record information over a wide range of intensities, suitable to thin regions near the skin as well as to thicker regions near the chest wall. Because the breast is one of the more radiosensitive regions in the body, the breast examination done with low radiation dose is compatible with the required image quality. The total dose for a screening mammogram with two views of each breast is about $3.7\ mGy$ for FFDM digital and $4.7\ mGy$ for SFM (Hendrick, 2010). A comparative study (Beldalli *et al.*, 2010), examined the dose impact of breast screening using different FFDM models: Philips MicroDose Essential, GE Senographe Essential and Hologic Selenia systems. The study found that the Philips mammography system has the lowest average dose $0.89\ mGy$, average doses were $1.455\ mGy$ and $1.4\ mGy$ for GE Senographe and Hologic Selenia respectively.

The quality of image processing is important for all of radiography but is critical for mammography. The quality of the images relies on the design and performance of the radiographic unit, image receptor and to the current use of that equipment to acquire the mammogram (Kundel and Revesz, 1976). The most effective method for evaluating the relation between physical properties and image quality is the receiver operating curve (ROC) methodology (Birdwell *et al.*, 2001).

2.3.5.1. Contrast in X-ray mammogram

Image contrast is the difference in display signal between the object of interest and its surround and is a product of two components, equation (2-1):

$$Image\ contrast = display\ contrast \times subject\ contrast \quad (2\text{-}1)$$

Display contrast is a function of the display, for conventional screen film is film density, while in digital images; the display contrast adjusts so that the anatomy of interest spans the desired video-monitor gray-scale range.

Subject contrast is commonly defined as in equation (2-2):

$$Subject\ contrast = \log_e(I_2/I_1) \quad (2\text{-}2)$$

where $I_2$the transmitted X-ray intensity is associated with the object of interest and $I_1$is transmitted intensity of its adjacent surrounding.

$$Subject\ contrast = \log(1 + \Delta P/P) \quad (2\text{-}3)$$

$$\cong\ \Delta P/P\ (when\ \Delta P/P < 0.1) \quad (2\text{-}4)$$

where P is the primary intensity of the surround and ΔP is the difference in primary intensity between the object and its surround. That is $I_1 \rightarrow P, I_2 \rightarrow P + \Delta P$

2.3.5.2. Physical factors affecting mammogram quality

Many Physical factors affect the contrast of the mammographic images, including scatter radiation, attenuation, noise, and compression of the breast.

1. **Scattered radiation:** Scatter radiation degrades the image contrast of mammogram. Scatter acts as a slowly varying background or out-of-focus radiation level superimposed on the image (Yaffe, 1990). When scattered radiations (S) reach the image receptor, they contribute to additional intensity in the points of the image. When scatter radiation are taken into account, $I_1 \rightarrow P + S, I_2 \rightarrow P + \Delta P + S$, then equation (2-2) becomes:

$$Subject\ contrast = \log[1 + \Delta P/(P + S)] \quad (2\text{-}5)$$

$$\cong (\Delta P/P)(1 + S/P) \quad (2\text{-}6)$$

$$\cong (\Delta P/P)SDF \quad (2\text{-}7)$$

where SDF is known as the scatter degradation factor. Numerically, the factor takes on value from one (no scatter) to zero (infinite scatter).

Grids are used to reduce the scattered radiations, thus the image obtained with a grid improved its contrast when compared to an image obtained without a grid (Chen *et al.*, 2015). A grid consists of an array of radiopaque foil strips separated by strips of radiolucent spacing material; the grid is positioned between the patient and image receptors (Sorenson *et al.*, 1980).

2. **Attenuation:** as the X-ray beam passes through tissue, photons get absorbed so there is less energy; this is known as attenuation. Variations in tissue composition give rise to differences in attenuation, which in turn give rise to variations in image contrast.

The formation of contrast in the X-ray beam transmitted through tissue occurs as a result of differential attenuation along different rays in the beam. The local transmission of an X-ray, I passing a distance dx through an object with a linear attenuation coefficient μ is given by:

$$I = I_o e^{-\mu dx} \quad (2\text{-}8)$$

where I_0 is the incident intensity.

If scatter is disregarded, the ratio of two transmissions X-rays (I_1, I_2) passing through an identical thickness (dx) of surrounding tissue, with a linear attenuations (μ_1, μ_2 respectively) is given by:

$$I_1/I_2 = 1 - [(\mu_1 - \mu_2)dx] \quad (2\text{-}9)$$

where $[(\mu_1 - \mu_2)dx]$ describes the scatter-free contrast in the transmitted X-ray beam. So, contrast relies on the object thickness and the difference between the two attenuation coefficients (μ_1 and μ_2).

3. **Noise:** is the result from statistical fluctuations of the whole acquisition process (going from the X-ray tube to the receptors). The contrast-to-noise ratio CNR is the key determinant of the visibility of a given lesion (Huda and Abrahams, 2015). Improving the image quality requires the CNR to be increased; this can be achieved either by increasing the contrast or reducing the amount of noise. Quantum noise can be reduced by increasing the number of X-ray photons used to make the image at the cost of increased radiation, and using image receptors with high quantum efficiency (Yaffe *et al.*, 2009).

4. **Compression:** compression of the breast to a uniform thickness has an important influence on both radiation exposure and image quality. Firm compression spreads out the anatomic structures, minimizing the superposition of shadows, reducing scatter radiations and thereby improving the image quality (Holland *et al.*, 2017). Contrast is improved when dealing with a "thinner" breast due to reduced scattering. A comparison study (Helvie *et al.*, 1994) conducted on 250 paired MLO-view and CC-view mammograms, to compare the thickness of the compressed breast between MLO and CC mammograms and to relate these differences in thickness to image quality and radiation dose. Results found the mean thickness of the compressed breast on the CC view was less than the mean thickness on the MLO view (4.4 versus 4.8 cm) despite the greater force used to compress the breast for MLO than for CC views (93 $versus$ 86 $newton$). A 5% and 12 % loss of contrast was noted when a 4.4 cm was compared to 4.8 cm thick breast. Mean glandular dose at 4.4 and 4.8 cm was 1.4 and 1.7 respectively. The study concluded that, although the compressed breast is 8% thicker on MLO than on CC mammograms but this difference results in loss of spatial and contrast resolution, these image quality differences may explain the better visualization of carcinoma on the CC-view than in MLO view.

2.3.6. Digital image and texture in mammography

Mammogram image is a gray-scale image. The background of the mammogram image is black (low values), and the breast will display grays and whites (higher values). The variation of gray levels is due to various X-ray attenuation that occurs according to tissue type, [Figure (2-13)]. The darker areas correspond to fatty tissue; the lighter areas to denser tissue which can contain ducts, lobes, and other structures. Light areas of mammogram reveal glandular breast tissue or breast tumor (Bassett *and* Gold, 1987).

Image texture can be defined as the spatial variation of gray-level pixels intensities, at a local scale. Texture is considered as an important characteristic that can be used for the analysis of radiological images. The spatial arrangement of the differing intensities within a particular region of an image may be more or less regular, random, linear, structural or probabilistic dependency of one upon another (Nailon, 2010).

In mammography the radiologist evaluate image texture within an area of the image by looking at the fine details and the overall spatial organization of these details. Any changes in either the linear or coalescent densities permit the radiologist to differentiate between normal and abnormal areas.

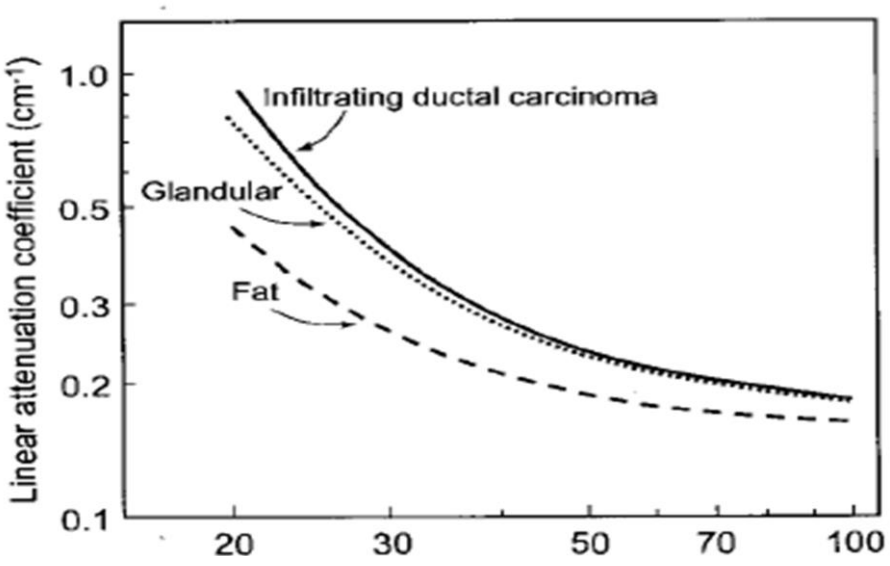


Figure (2-13): Attenuation of breast tissue as a function of energy, (Yaffe, 1994)

2.3.7. Texture of normal /abnormal mammogram

As abnormal mammogram images have certain signs and characteristic for evaluation, the normal mammograms have different appearance.

2.3.7.1. Normal mammogram

The shadows in the mammograms that appear as diffuse amorphous clouds of density with indistinct borders are caused by glandular tissue in the breast; these clouds raise the

local average brightness of image. [Figure (2-14)] shows normal mammograms from three types of tissue density. In general, two observations can be made of normal mammograms (Evans *et al.*, 2003);

- Unequivocally normal areas have lower overall density, with no spikes, spots or large bright areas.
- Normal areas have linear markings which represent the shadows of ducts and connective tissue elements

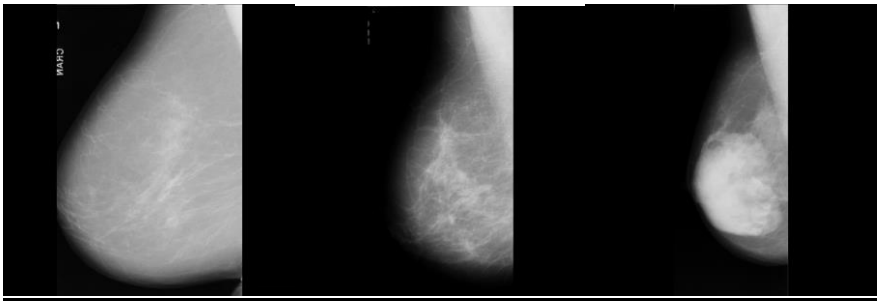


Figure (2-14): Normal Mammograms for different tissue types: fatty, glandular and dense (from *left* to *right*), (MIAS database)

2.3.7.2. Abnormal mammogram

There are two types of breast cancer tumors, one of them are non-cancerous or 'benign', the others are cancerous or 'malignant'. Benign tumor cells grow only locally cannot spread by invasion or metastasis and are often surrounded by a protective 'sac' that segregates it from the rest of body enable it to easily remove. Malignant cells invade neighboring tissues, enter blood vessels and metastasize to different sites. However, differentiating between benign and malignant finding in mammogram images is difficult (Wedegartner *et al.*, 2001).

Many different signs indicate for mammographic abnormalities which may benign or malignant. Among these signs are the presence of micro-calcifications, speculated lesions, circumscribed (well-defined) masses, ill-defined (irregular) masses, architectural distortions and asymmetry. The following paragraphs provide a brief description of these abnormalities.

1. **Calcification:** is a tiny calcium deposit that has accumulated in the breast .tissue and it appears as small bright spots on the mammogram. A cluster is defined to be at least 3 to 5 micro-calcifications within a (1 $cm2$) region. The calcifications vary in size from ($0.1mm - 5mm$) in diameter. Calcifications can be categorized into benign, intermediate and malignant types (Nalawade, 2009) according to their morphology and distribution as in [Figure (2-15)]. [Figure (2-16)] show example of variety shapes of calcifications.

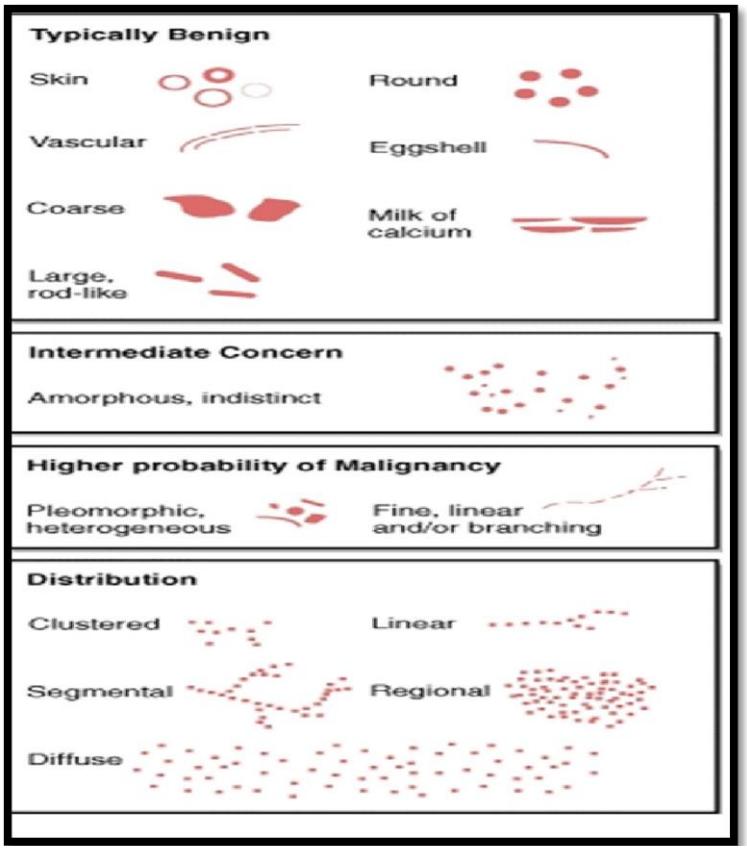


Figure (2-15) : Morphology of calcifications, (Lawrence *et al.*, 2003)

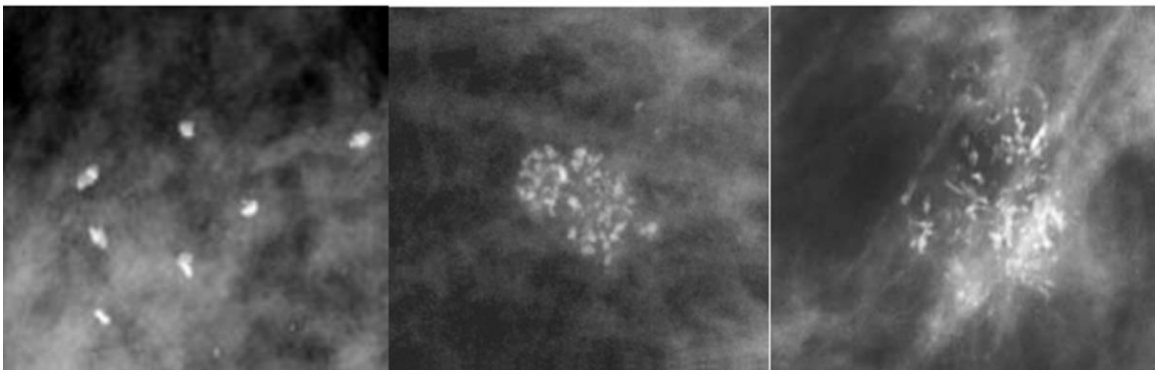


Figure (2-16): Example of different shapes of micro-calcifications, (extracted from DDSM database)

2. **Masses:** appear as dense regions of different sizes and properties. Generally there is a circumscribed mass, which is circular in shape and has a distinct border. More difficult to detect are ill-defined masses which have an irregular shape and less distinctive border. The most significant features that differentiate between benign and malignant masses are its shape and margins (Elston and Ellis, 1991). [Figure (2-17)] shows different shapes and margins for masses. [Figure (2-18)] shows example of masses

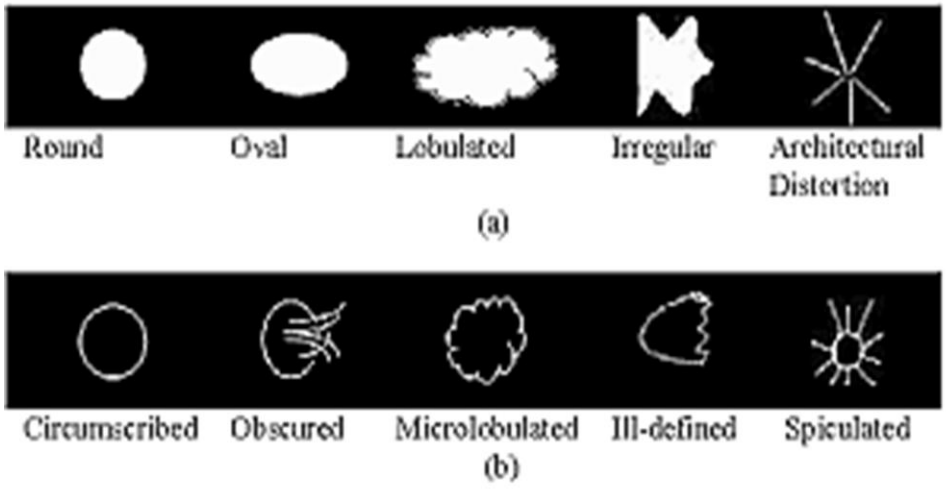


Figure (2-17): Various well-known masses (a) shapes (b) margins; (Surendiran and Vadivel, 2012)

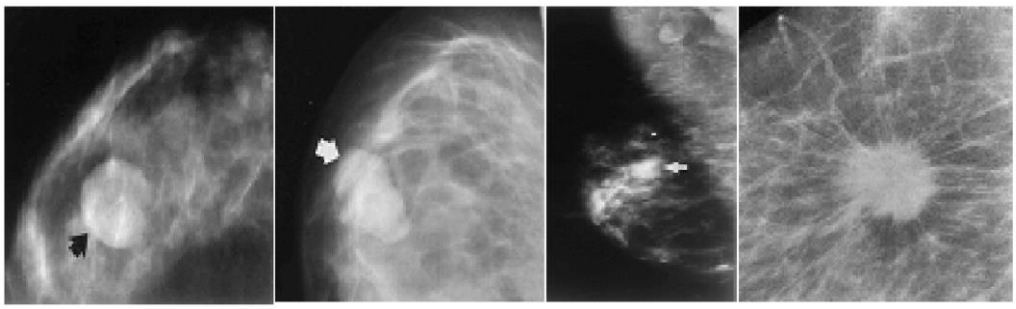


Figure (2-18): Masses examples; (circumscribed shape, lobular shape, ill-defined margin, and irregular shape) respectively, (Lawrence *et al.*, 2003)

3. **Architectural distortion** is defined by the Breast Imaging Reporting and Data System (BI-RADS) as an appearance in which the normal architecture of the breast is distorted with no definite mass visible. Architectural distortion representing nearly

6% of abnormalities detected on screening mammography and it is difficult to diagnose because it can be subtle and variable in presentation (Gaur *et al.*, 2013).

4. **A spiculated lesion** is defined as a mass or an architectural distortion characterized by thin lines radiating from its margin. Spiculated lesions has a star shape with blurred borders, Star shaped caused by radially oriented spicules extending from the tumor center into the surrounding breast tissue, [Figure (2-19)]. The density of the tumor center and the radiating structure are important factor for spiculated lesion analyzing (Franquet *et al.*, 1993).

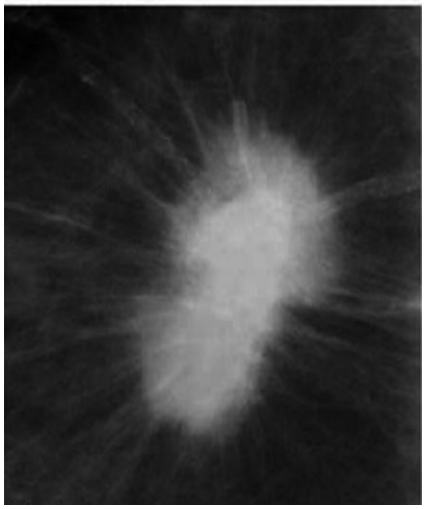


Figure (2-19): Spiculated lesion example, (Sampat *et al.* , 2008)

5. **An asymmetry:** means there is more tissue, or white stuff on the mammogram in one area than on the opposite side. Symmetry is a mammographic abnormality visible on at least two different mammographic projections, the observations of asymmetry taken by comparing the mammogram with previous examinations (Price *et al.*, 2015).

2.3.8. Standardized mammography report

The Breast Imaging-Reporting and Data system (BIRADS) was published by the American College of Radiology (ACR). It relies on a quality assurance tool designed for mammography and aims at defining accurate breast imaging reporting to reduce confusion in

breast imaging interpretations. It contains a lexicon to standardize terminology for mammography, breast US and breast MRI (Bassett *et al.*, 2003).

Standardized Mammography report should include the reasons for breast examination, breast tissue composition, a description of mammographic findings using a standardized lexicon [Table (2- 3)] (D'Orsi *et al.*, 2013), and a final assessment with a management recommendation, as it is detailed below:

1. **Breast composition:** In the BI-RADS edition of 2003, the assignment of the breast composition was based on the overall density of fibro-glandular tissue resulting in four categories: category 1 ($< 25\%$ fibroglandular tissue), category 2 ($25 - 50\%$), category 3 ($50 - 75\%$), and category 4 ($> 75\%$). In 2013, BI-RADS discouraged the use of percentage; assignment of the breast composition is changed into a, b, c, and d-categories
2. **Important findings:** The part of the report should describe any significant finding using standardized terminology. Masses and micro-calcifications are the most common morphological descriptors. Descriptors may have associated features as skin retraction, nipple retraction, and skin thickness.
3. **Final assessment category:** Standardized report should concludes with one of the six assessment categories (BI-RADS categories 1, 2, 3, 4, 5, and 6) [Table (2-4)] (D'Orsi *et al.*, 2013). An incomplete (category 0) assessment is usually given for screening examinations when additional imaging evaluation is recommended before it is appropriate to render a final assessment. The Mammography Quality Standards Act (MQSA) requires that the assessment category be included in the mammography report.

Table (2-3): Standardized mammogram lexicon, (BIRADS Atlas, 5th ed.)

ACR BI-RADS® Atlas Fifth Edition
QUICK REFERENCE

MAMMOGRAPHY

Breast composition	a. The breasts are almost entirely fatty b. There are scattered areas of fibroglandular density c. The breasts are heterogeneously dense, which may obscure small masses d. The breasts are extremely dense, which lowers the sensitivity of mammography	
Masses	Shape	Oval
		Round
		Irregular
	Margin	Circumscribed
		Obscured
		Microlobulated
		Indistinct
		Spiculated
	Density	High density
		Equal density
		Low density
		Fat-containing
Calcifications	Typically benign	Skin
		Vascular
		Coarse or "popcorn-like"
		Large rod-like
		Round
		Rim
		Dystrophic
		Milk of calcium
		Suture
	Suspicious morphology	Amorphous
		Coarse heterogeneous
		Fine pleomorphic
		Fine linear or fine-linear branching
	Distribution	Diffuse
		Regional
		Grouped
		Linear
		Segmental
Architectural distortion		
Asymmetries	Asymmetry	
	Global asymmetry	
	Focal asymmetry	
	Developing asymmetry	
Intramammary lymph node		
Skin lesion		
Solitary dilated duct		
Associated features	Skin retraction	
	Nipple retraction	
	Skin thickening	
	Trabecular thickening	
	Axillary adenopathy	
	Architectural distortion	
	Calcifications	
Location of lesion	Laterality	
	Quadrant and clock face	
	Depth	
	Distance from the nipple	

Table (2-4): BIRADS assessment (BIRADS Atlas, 5th ed)

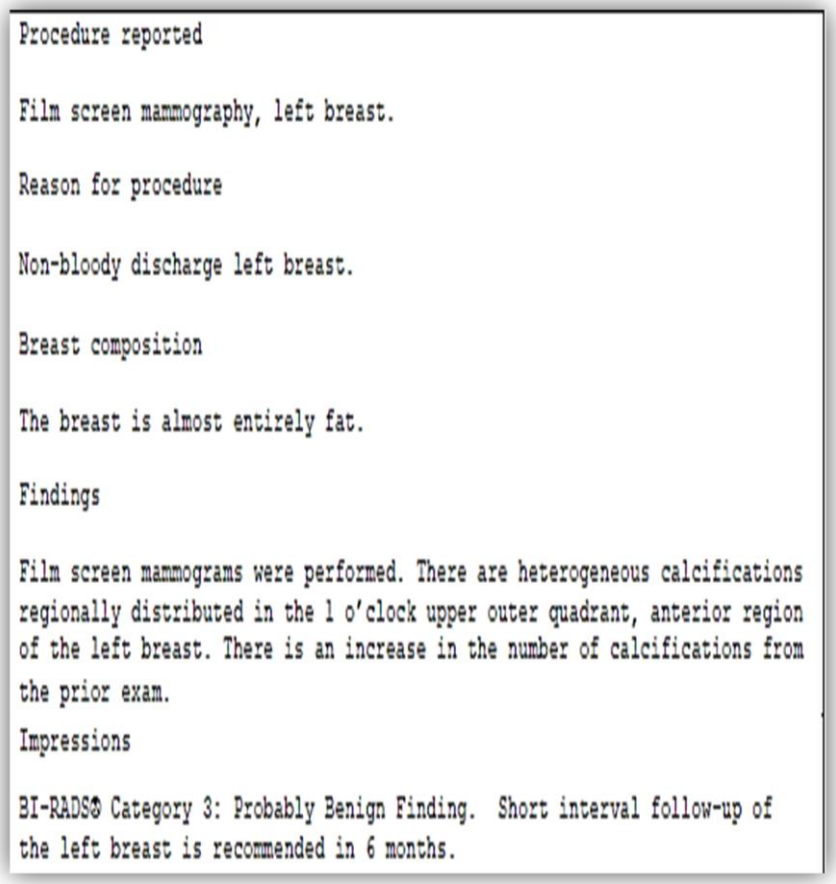
BI-RADS® ASSESSMENT CATEGORIES		
Category 0:	**Mammography:** Incomplete – Need Additional Imaging Evaluation and/or Prior Mammograms for Comparison **Ultrasound & MRI:** Incomplete – Need Additional Imaging Evaluation	
Category 1:	Negative	
Category 2:	Benign	
Category 3:	Probably Benign	
Category 4:	Suspicious	Mammography & Ultrasound: Category 4A: Low suspicion for malignancy Category 4B: Moderate suspicion for malignancy Category 4C: High suspicion for malignancy
Category 5:	Highly Suggestive of Malignancy	
Category 6:	Known Biopsy-Proven Malignancy	

The Digital Imaging and Communication in Medicine (DICOM) standard allows different structured reporting (SR), called SR templates. There are different DICOM working groups specialties for different (SR), DICOM Working Group 15 provides template for mammography in supplement 50, Likewise, DICOM breast imaging templates (DICOM Supplement 79) link the breast imaging report with BIRADS findings within the impression section of the structure report (Hussein *et al.*, 2004).

The two following DICOM reports are illustrations of encoding mammography procedure based on Breast Imaging Report; the first one is for the screening procedure [Figure(2-20)], the second one for diagnostic procedure [Figure(2-21)].

```
Procedure reported

Film screen mammography, both breasts.

Reason for procedure

Screening

Comparison to previous exams

Comparison was made to exam from 11/14/2001.

Breast composition

The breasts are heterogeneously dense. This may lower the sensitivity of
mammography.

Findings

No significant masses, calcifications, or other abnormalities are present.
There is no significant change from the prior exam.

Impressions

BI-RADS® Category 1: Negative. Recommend normal interval follow-up in 12
months.

Overall Assessment

Negative
```

Figure (2-20): Report sample of mammogram with negative findings, (DICOM Standards Committee, 2004)



Procedure reported

Film screen mammography, left breast.

Reason for procedure

Non-bloody discharge left breast.

Breast composition

The breast is almost entirely fat.

Findings

Film screen mammograms were performed. There are heterogeneous calcifications regionally distributed in the 1 o'clock upper outer quadrant, anterior region of the left breast. There is an increase in the number of calcifications from the prior exam.

Impressions

BI-RADS® Category 3: Probably Benign Finding. Short interval follow-up of the left breast is recommended in 6 months.

Figure (2-21): Report sample of mammogram with positive findings, (DICOM Standards Committee, 2004)

2.3.9. Emerging technologies in mammography

When compared with conventional X-ray mammography, exam sensitivity can be improved with the digital computer assisted techniques as digital tomosynthesis, mammography with synchrotron radiation, Phase contrast mammography, stereoscopic mammography, and computed tomography laser mammography (Gur *et al.*, 2009).

Digital breast tomosynthesis is a mammography based technique, Tomosynthesis acquisition involves acquiring multiple image of a stationary compressed breast at different angles during short scan it permits to study single slice of the breast without tissue overlapping, and this technique is useful in dense breast and fits the current mammographic system easily (Baker *et al.*, 2011). In 2012, tomosynthesis obtained the FDA approval for commercial system (Kopans, 2014).

Mammography with synchrotron radiation is an innovative X-ray imaging technique. Synchrotrons generate X-ray beams that are practically monochromatic, tunable, laminar, and have sufficient spatial coherence. It exhibits high resolution and contrast, depicting structures and details missed by conventional mammography (Castelli *et al.*, 2011).

Phase-contrast mammography is an X-ray based technology that has been shown to provide enhanced soft tissue contrast and improved visualization of cancerous structures. The X-ray energies used in this modality are typically low$(20 - 30\ kv)$, at these energies, small differences in the reactive index have a higher impact on the phase than on the absorption, this lead to improve the contrast of the image (Auweter *et al.*, 2014).

Stereoscopic mammography is providing direct in depth views of the internal structure of the breast and increases the screening accuracy. The lesion seen as separate from normal tissue, aligned with it but at different depth in the breast volume. Stereoscopic mammography provides more accurate diagnosis particularly for micro-calcifications and architectural distortion (D'Orsi *et al.*, 2013).

Computed tomography laser mammography (CTLM) is a supplementary method of the basic examination, which is used for visualization of vascular structures for both physiological blood vessels and neovascularization. CTLM uses laser beam of the wavelength equal to 808 nanometers, which is absorbed in blood pigments of blood vessels and is able to display their distribution. CTLM use for breast disease diagnosis and is able to distinguish malignant tumor and benign lesion (Bilkova *et al.*, 2010).

The combination of multimodal imaging techniques could provide a more accurate analysis and an early diagnosis of breast cancer. There are strong recommendations for developing a full breast 3D ultrasound system and digital mammography system, in which the ultrasound scanning will be performed in the same compression as the digital mammogram so that the lesions can be easily correlated geometrically (Carson *et al*., 2004).

2.4. CAD for mammography

Computer-Aided Diagnosis (CAD) for a mammography represents a set of methods that are used for the detection and the characterization of cancer. The first paper dealing with the interest of computer-aided methods in mammography was published in 1967 (Winsberg *et*

al., 1967). Although this paper reported that computer analysis could play a role in early detection of mammographic abnormalities, but these early attempts were not sufficiently successful due to limited computing power (Doi, 2007). The improvement of computing power and of computer vision techniques leads to CAD improvement. In 1995, at the University of Chicago, a CAD system was first introduced in clinical practice for analyzing masses and calcifications (Giger *et al.*, 2008).

CAD systems have become important tools to assist radiologists in the mammographic interpretation process (Hadjiiski *et al.*, 2006). There are two types of CAD systems: the first type of systems is computer-aided detection (CADe) methods, which improve radiologists' accuracy in the detection of breast cancer. The second type of CAD systems is computer-aided diagnosis (CADx) methods, which classify some detected suspicious regions into malignant or benign categories to help the radiologists and clinicians define patient management.

CAD systems have been developed for SFMs and FFDMs. Studies conducted to compare the performances between the CAD systems for FFDMs and for SFMs (Wei *et al.*, 2007) (Ge *et al.*, 2007). However, CAD system based on FFDMs provide the gains of having higher signal to noise ratio (SNR), wide dynamic range, and higher contrast sensitivity than SFMs.

Once CAD has proposed an assessment, radiologists have to further scrutinize the mammograms, especially the region, which has been detected by the CAD system and finally check whether the ROIs containing the CAD detected anomaly are pathological lesions or benign lesion or normal area.

Indeed, CAD systems are intended to assist radiologists but not to replace them; the radiologists should be the final judges in determining the final assessment (Muralidhar *et al.*, 2008). [Figure (2-22)]. Also, there is a clear benefit to the use of CAD by less experienced radiologists (Dromain *et al.*, 2013).

The overall scheme of CAD systems generally consisted of several steps. The breast region boundaries are first segmented from the mammogram, the mammogram can be preprocessed to enhance the suspicious regions. Suspicious regions are then segmented from the breast image based on their gray level contrast, gradient orientation, or shape information. Features descriptors are then extracted from the segmented objects. Finally a classification

step is applied, using either rule-based classifier, linear, nonlinear, or neural network classifiers (Giger *et al.*, 2008).

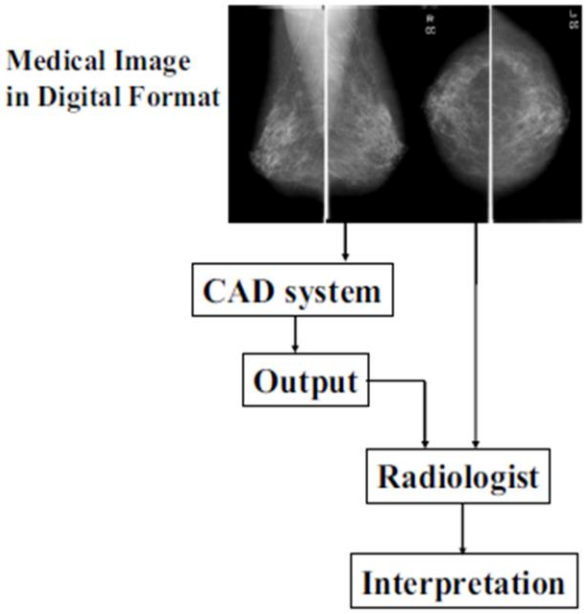


Figure (2-22): Schematic diagram of CAD for mammogram interpretation, (Giger *et al.*, 2008)

2.4.1. CAD performance

As CAD systems mark some ROIs on images, as being potential anomalies; some marked locations are true positives while others are false positives. Many studies conducted to evaluate the impact of CAD systems on the performance of radiologists. Few studies concluded, the using of CAD systems had no statistically significant effect on radiologists' performance (Fenton *et al.*, 2007) (Biggelaar *et al.*, 2010). But, many studies (Luo *et al.*, 2005), (Gilbert *et al.*, 2006), (Bolivar *et al.*, 2010), (Scaranelo *et al.*, 2010) (Giger, 2000) (Birdwell *et al.*, 2001) demonstrated that radiologists obtained a statistically significant improvement in performance when they used the CAD versus when they did not; thereby CAD helped them decide whether women need a biopsy, and so reduce the number of unnecessary biopsies.. Currently most CAD systems are calibrated in order to detect more false positive cases than false negative cases in order not to miss some lesions.

For CADe systems evaluation, the free-response operating characteristic (FROC) curve employed in assessing performance of detection algorithm (Wei *et al.*, 2011), while CADx systems are evaluated using the receiver operating characteristic (ROC) curve analysis method (Wagner *et al.*, 1998).

In the ROC model the observer assigns a single rating to each image, and the location of the perceived abnormality. The ROC curve is created by plotting the true-positive rate $(sensitivity)$ (2-10) against the $false-positive\,rate$ $(1-specificity)$ (2-11). In the FROC model, the observer is mark and rates any suspicious region in the image. The FROC curve is created by plotting the $sensitivity$, versus the number of $false-positives\,(FP)\,per\,image$ (2-12) for different thresholds of detection (Chakraborty, 2013). Most CAD systems report good sensitivity but high false positive rates. Sensitivity generally decreases as the recall rate and false positive rate decreases (Yankaskas *et al.*, 2001). Recall rate was defined as the percentage of screening studies for which further workup was recommended by the radiologist.

$$Sensitivity = \frac{Number\ of\ True\ Positives}{(Number\ of\ True\ positives + Numbber\ of\ False\ Negatives)} \quad (2\text{-}10)$$

$$Specificity = \frac{Number\ of\ True\ Negatives}{(Number\ of\ True\ Negatives + Numbber\ of\ False\ Positives)} \quad (2\text{-}11)$$

$$FP/im = \frac{Number\ of\ False\ positive\ Marks}{Total\ number\ of\ Images} \quad (2\text{-}12)$$

In 1990, a group of researchers, reported the first observer study showing the improved of radiologist's performance when using CAD in detection of micro-calcifications [Figure (2-23)] (Chan *et al.*, 1990). CAD systems have in generally good reliability especially for the detection of micro-calcifications and masses. Detection scores of 99% for micro-calcifications (Burhenne *et al*., 2000) and from 75% to 89% for masses (Houssami *et al.*, 2009).

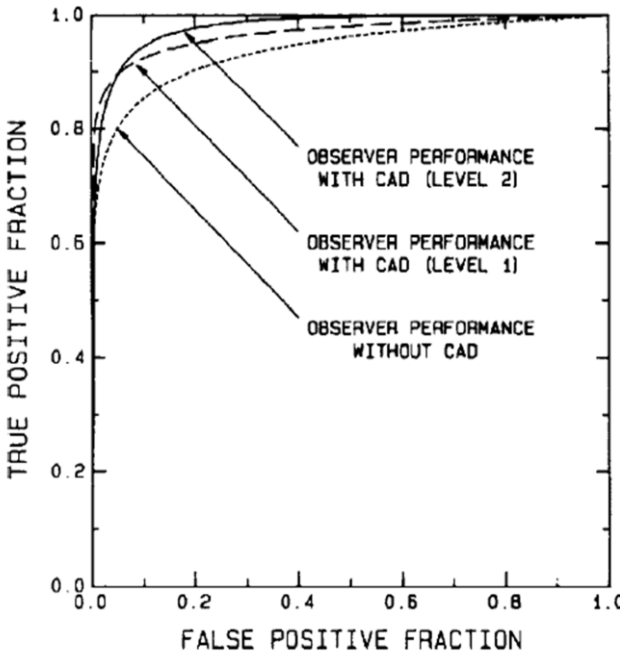


Figure (2-23): ROC curve illustrating statistically significant improvement in radiologists' detection of microcalcification cluster when CAD is used. Level 1 corresponds to use of CAD having a performance level of 87% true-positive rate and an average of four false-positive clusters per image. Level 2 corresponds to use of CAD having a performance level of 87% true-positive rate and a simulated average false-positive cluster rate of one false-positive cluster per image (Chan *et al.*, 1990)

2.4.2. Commercial CAD for mammography

At present, commercial mammographic CAD systems are being widely used in medical centers; nevertheless, their performances very often used as reference to private databases (Cascio *et al.*, 2014). Many reports have been published on the usage of some of the commercial systems in clinical practice (Birdwell *et al.*, 2005) (Morton *et al.*, 2006). The results showed that the cancer detection rate increased with an escorted increase in the recall rate. Some big companies such as Siemens, Hewlett Packard, General Electric (GE), Sterling Diagnostic Imaging, and Hologic rushed to develop CAD systems in mammography.

The first commercial CAD system for the detection of breast cancer in mammography, ImageChecker, was developed by R2 Technology. It was based on the licensing of CAD technologies from the University of Chicago, and the first FDA approval of a CAD system for clinical use was obtained in 1998 (Li and Nishikawa, 2015).

A study was conducted to test the ability of this CAD (R2 Technology, version 2.0) to detect cancers, which were missed at screening due to the density of breast tissue or to

distracting lesions. Considering 115 $cases$, CAD marked most 77% cancers missed at screening mammography; the CAD identified 86% of missed calcifications and 73% of missed masses (Birdwell *et al.*, 2001). .

In 2006, an outpatient imaging center was plan to purchase software for CAD to assist in analysis of digital mammography, the preparation of purchase leads to a comparison of several specifications. So, two commercial CAD systems are compared by researchers, a study conducted on [R2 ImageChecker (version 8.3.17) and iCAD SecondLook (version 7.2-H)] from screening mammograms obtained with the General Electric Senographe DS unit. The comparison study applied on 94 mammography images, results showed no important differences between the two systems, either in terms of the number of false positives or in the placement of the markers in appropriate location. However, the R2 marked fewer false positive masses than iCAD system, and the iCAD marked fewer false positive calcifications than R2 system (Leon *et al.*, 2009).

2.4.3. Mammogram databases used in CAD system

Mammographic databases play an important role in the development of algorithms for CAD systems. Moreover, public databases allow the comparison of results that were obtained by different studies (Antoniou *et al.*, 2009).

Several databases have commonly used as tester for the performance of the proposed CAD algorithms. There are public, online-freely access databases as Mammographic Image Analysis Society (MIAS) (Suckling *et al.*, 1994) database and Digital Database for Screening Mammography (DDSM) (Heath *et al.*, 1998), MIAS and DDSM consider as the most commonly used databased. Besides, there are currently new projects developing mammographic databases as well as several old projects.

MIAS: The MIAS is an organization of UK research groups interested in the mammography and has initiated a database of digital mammograms. MIAS database consider as the most publicly available, easily accesses, and therefore the most commonly used database. Mammographic images are available via the Pilot European Image Processing Archive (PEIPA) at the University of Essex. In this research, we using MIAS database, so, further information has been presented in the later chapters.

DDSM: The DDSM database was published in 1998; it comes from USA and contains 2,479 studies. Each study includes two images of each breast in (CC) and (MLO) views, these results of 9,916 mammography images. Spatial resolution ranges between $50 - 42\ microns$. Images were coded using the lossless Joint Pictures Expert Group (JPEG) standard. For all cases, there are plain text files with information on the type of digitizer and a staging of the tissue density according to ACR, appropriate information about the lesions types and a chain code with the localization and delineation of lesions.

LLNL/UCSF: this database (Chandrasekhar and Attiziouzel, 1998) was created by the Lawrence Livermore National Laboratories and Radiology Department at the University of California at San Francisco

Computer-Assisted Library for Mammography database (CALMa): is an Italian collaboration among some institutes departments and some medical centers of mammographic screening (Marzulli, 1999). It consists of 3000 images with contrast resolution$12\ bits/pixels$, and $85\mu m$ spatial resolution .Its goal is to collect a set of digital mammographic images and to work out a suitable CADx system (Amendolia *et al.*, 2001).

IRMA: Integrated to the Image Retrieval in Medical Applications (IRMA) project (Oliveira *et al.*, 2008) contains 10,509 reference images collected from the combination of: MIAS, DDSM; LLNL/UCFS, and images from the Rheinisch-Westfalische Technische Hochschule (RWTH) Aachen. Using the IRMA code, standardized coding of tissue type, tumor staging, and lesion description was developed according to the ACR and BI-RADS.

African-American breast cancer mammography database: this database (Ross *et al.*, 2008) outcome of collaboration between the Howard University Electrical and Computer Engineering department, the Howard University Hospital Radiology department, and the Georgetown University Radiology Department. This database contains 5,000 digitized mammography images; mammography films taken from 260 patients, all of these patients were diagnosed with breast cancer between 1994 and 2004 and were between the ages of 24 and 88. Each patient having approximately 20 to 40 $images$ due to several regular screening mammograms, each exam would contain four typical views (Left-MLO, Right-MLO, Left-CC, and Right-CC). The digitized image size is 40 or $60MB$,

depending on the size of the mammogram films (8×10 or 10×12 $in.$, respectively.). A viewing system, D-Viewer was developed to display the digitized mammograms. This viewer is coded in Microsoft Visual C#, the view system uses graphical user interface (GUI) and can open by DICOM. This database is accessible via announcements on the Howard University website and freely access for basic database, but with fees for advanced functions as content-based retrieval operations.

BancoWeb: BancoWeb LAPIMO , is a Brazilian database (Matheus and Schiabel, 2011), it was published in 2010, and has a total of 1400 $images$ from around 320 patients. Most of the mammogram images performed by women from 40 to 60 $years\ old$. The database has images of $12 - bit$ in gray scale contrast with spatial resolutions between $0.075mm$ (32%) and $0.150mm$ (68%). Statistical profiles are available, images are in TIFF format, and can open by DICOM. BancoWeb database is free online access, but the users need a registration.

INbreast: The INbreast database (Moreira *et al.*, 2012) has a total of 115 $cases$ (410 $images$), in which 90 cases are from women with both breasts (four images per case) and 25 cases are from mastectomy patients (two images per case). Images are in (MLO) and (CC) views. Several types of lesions were included. Accurate lesion contours made by specialists are also provided in XML format. Images were saved in the DICOM format. For each image in INbreast database, its density in ACR standard scale and the BI-RADS classifications is available. This database is free online access, but the users need a registration.

The Cancer Imaging Archive (TCIA): the National Cancer Institute (NCI) contracted with Washington University to create the TCIA (Clark *et al.*, 2013). TCIA is an open access information resource to support researches related to medical imaging of cancer. In the first year of operation, 2011-2012, TCIA accumulated $3{,}268{,}644$ $images$ from different imaging modalities including the mammography, and from different anatomical sites including the breast.

Optimam Mammography Imaging Database (OMI-DB): this database (Patel *et al.*, 2017) collecting images from multiple sites throughout the UK, the images extracted from the National Breast Screening System (NBSS). The database contains both processed and unprocessed mammographic images, associated data, annotations and

expert-determined ground truths along with computational image feature extraction. OMI-DB contains statistics of 5437 $clients$, with 680 normal cases, 450 benign cases, and 4307 malignant cases. The database is view by MedXViewer software application. Furthermore, mammography images open by DICOM.

The National Mammography Database (**NMD)**: the American College of Radiology's National Mammography Database (NMD) is leverages data that collected under MQSA federal mandate to create reports that benchmark facilities and physician performance and exceed the FDA's audit data collection requirements. In 2016, study conducted to analyze screening mammography data submitted to the NMD (Lee *et al.*, 2016), results showed that the successful of NMD to collected and analyzed data for 3,181,437 screening mammograms performed between January 2008 and December 2012. Mean values for outcomes were cancer detection rate of 3.43 per 1000 (95% $Cl, 3.2 - 3.7$), recall rate of 10% (95% $Cl, 9.3 - 10.7$), positive predictive values for biopsy recommended of 18.5% (95% $Cl, 16.7 - 20.2\%$), and biopsy performed of 29%

CBIS-DDSM: Curated Breast Imaging Subset of DDSM (Lee *et al.*, 2017), is an updated and standardized version of the DDSM. The CBIS-DDSM collection includes a subset of the DDSM data selected and curated by a trained radiologist, updated mass segmentation and bounding boxes, and pathologic diagnosis for training data. The images have been decompressed and converted to DICOM format. The database contains 753 calcification cases and 891 mass cases.

3. Review and Performance Evaluation of CAD System Approaches for Mammographic Image Analysis

A CAD system defines software algorithms to look for abnormal patterns in mammogram images, such as masses. The algorithms are trained to identify abnormal patterns using data from a defined sample of mammogram images including normal cases and abnormal cases; this sample is called the training set. Once trained or optimized, the CAD device can be tested using new mammogram images. For these images, the selected patterns observed are compared to the pattern discovered in the training set and are further classified. CAD algorithms are not adaptive; indeed they are freeze, and they only change with new software revisions (Giger *et al.*, 2008).

Most of the CAD algorithms include a succession of preprocessing steps to overcome some limitations of the acquisition. Then the main step is the segmentation of the region of interest, it is followed by feature extraction in order to characterize previously delineated objects. The last step of CAD algorithms corresponds to the classification step, which is based on the extracted features (Sampat *et al.*, 2005). [Figure (3-1)] illustrates the main steps of CAD systems.

Figure (3-1): Main stages of CAD system for breast cancer detection (Sharma and Khanna, 2015)

1.1. Preprocessing of mammogram

Preprocessing is a very important stage. It aims at limiting the search for abnormalities on only breast region, in order to get rid of the effect of the background. Preprocessing contributes to the reduction of mammogram size, and also improves the quality of the images to make the feature extraction phase more reliable.

As it will be detailed in the following subsections, numerous papers introduced different approaches for preprocessing mammogram images. Various methods were proposed to enhance mammograms (section 3.1.1), and segment the breast region in order to eliminate the background and pectoral muscles (section 3.1.2).

1.1.1. Mammogram enhancement techniques

Noise in mammography can be due to the capture and transmission processes. Noise is either Gaussian, or impulse. The impulse noise appears as light and dark pixels under a random spatial distribution in the image. The noise affects both the image processing and image visualization. Therefore, removing the noise has an important impact on the mammogram image processing. The Median filter is a rank-order filter and it has widely used in digital image processing. Its noise reducing effects depend on the size and shape of the filtering mask; and its algorithmic complexity mainly depends on how to get the median value (Zhu and Huang, 2012). Compared to linear filters, median filters are well adapted to remove impulsive noise, while both types of filters can reduce Gaussian noise.

Image enhancement can improve the quality of mammogram image for accurate diagnosis. Image enhancement include techniques such as contrast and intensity manipulation, noise reduction, background removal, edge sharpening and filtering. However, some image enhancement techniques lead to misdiagnosis due to under-enhancement which can cause false negative or over-enhancement which can cause false positive (Cheng *et al.*, 2003). Most common enhancement techniques applied to mammograms are the following:

1.1.1.1.Histogram modeling

Histogram modeling techniques modifying the dynamic range, and contrast of an image by altering the histogram of the image. A histogram of an image is the intensity distribution

of pixels in the image, or the probability of occurrence of a specific gray-level in the image. The histogram of a digital image with gray levels in the range $[0, L-1]$ corresponds to a discrete function:

$$h(r_k) = n_k \tag{3-1}$$

where r_k is the kth gray level and n_k is the number of pixels having gray level r_k, a normalized histogram (Rafael and Gonzalez, 2002) is expressed in the term of probability of occurrence of gray-level as :

$$P(r_k) = \frac{n_k}{n} \tag{3-2}$$

where n is the total number of pixels and $k = 0,1, \ldots . . L-1$

There are different enhancement methods based on histogram modeling, such as: histogram equalization, adaptive histogram equalization, contrast limited adaptive histogram equalization

Histogram equalization (HE)

Is a method of contrast adjustment using the image's histogram. HE used to redistribute the intensity values of pixels via replace every pixel by integral of the histogram of the image in that pixel (Ponraj *et al.,* 2011) so as to obtain a histogram as uniform as possible. From equations (3-1) and (3-2), HE transformation function (s) is given by (3-3). This technique was used in (Langarizadeh *et al.,* 2011).

$$s = \sum_{j=0}^{k} P(r_j) = \sum_{j=0}^{k} \frac{n_j}{n} \tag{3-3}$$

where $k = 0,1,2, \ldots L-1$ and $P(r)$ corresponds to the probability of occurrence of gray level r_k in the image.

Adaptive histogram equalization (AHE)

In this technique, the peak of the histogram stretch through the intensity levels (Pizer *et al.,* 1987). In the AHE each pixel is modified based on the pixels that are in a region surrounding that pixel. This region is called 'contextual region'. Unlike the HE method, the AHE computes several histograms from distinct section of the image, and uses them to reassign the lightness values of the image. AHE has been shown to enhance contrast in mammogram images, which in general have a large global dynamic range, but small local feature gray-level variations (Zimmerman *et al.,* 1988).

Contrast limited adaptive histogram equalization (CLAHE)

This technique was initially implemented for enhancing biomedical images as mammograms. The CLAHE can overcome the problem of over-enhancement which is a major problem of AHE (Pizer *et al.,* 1987). CLAHE method will present in more details in the following chapter. However, some papers presented other enhancement techniques based on CLAHE method. Histogram modified contrast limited adaptive histogram equalization (HM-CLAHE) is proposed (Sundaram *et al.,* 2011) to adjust the level of contrast enhancement, which in turn gives the resultant image a strong contrast and brings the local details for micro-calcifications detection, the technique tested by MIAS database. Also, for mammogram image enhancement, (Mohan and Ravishankar, 2013) are used CLAHE based on local contrast modification (LCM), the LCM-CLAHE method used to highlight the finer hidden details in mammogram images as micro-calcifications and to adjust the level of contrast enhancement. The proposed method tested on MIAS database, the performance measured by peak signal to noise ratio (PSNR), the experimental results show the superiority of the method comparing with other enhancement techniques as HE and CLAHE.

1.1.1.2.Contrast stretching

Also called normalization, aims at rescaling intensity values through the analysis of the image histogram so that there is a greater separation between foreground and background gray level distribution. A simple example of contrast stretching is a linear rescaling of the gray-level distribution in the image (Morrow *et al.,* 1992), which denoted by following equation:

$$y = kx + m \tag{3-4}$$

where x is the input gray-scale image, y is the output image, k and m are non-zero transformation parameters. Relying on the values of k and m, contrast will be increased or decreased. Other typically transformation function is given by (Thangavel *et al.,* 2005)

$$y = \begin{cases} \alpha x & 0 \le x < a \\ \beta(x-a) & a \le x < b \\ \gamma(x-b) & b \le x < L \end{cases} \tag{3-5}$$

where, the parameters a and b can be obtained by examining the histogram of the original mammogram, the slope α, β, and γ are chosen greater than unity in the region of stretch, and L is the maximum gray value of the original image. According to (Morrow *et al.*, 1992), a non-linear transform function is represented in exponential form:

$$y = kx^p \tag{3-6}$$

where k is a factor to rescale the output image to the range of the input image. This technique can remove the uniform background, but it is difficult to remove the noise which gray-level are similar to objects such as micro-calcifications (Thangavel *et al.,* 2005). It was used for instance in (Langarizadeh *et al.*, 2011).

1.1.1.3.Gradient operators

This technique aims at reducing the low frequency information and amplifying the high frequency details. Some usual gradient operators are defined by convolution masks, such as Sobel masks. Some more elaborated methods are unsharp masking. In unsharp masking, a blurred mammographic negative is placed in register with a positive image to obtain a difference image containing the details in the image, and then the final image is obtained by amplifying the difference image and added to the blurred positive image (Chan *et al.*, 1998). Unsharp masking enhanced the sharpness of the borders of mass lesion. However, it is easy to create unwanted and conspicuous edge effect, so increase image noise. According to (Cernadas *et al.*, 1996), the unsharp masking can be represented by:

$$C = \gamma . A + (1 - \gamma) . B \tag{3-7}$$

where A represents the source image, B is the image obtained by applying a low pass filter to image A, and γ is an enhancement factor. If: $\gamma \epsilon (0,1)$ the effect is blurring, if: $\gamma > 1$ the effect is sharpening the details. In (Chan *et al.,* 1987), an investigation study conducted, to study the effect of unsharp-mask filtering on the detection of micro-calcifications in digital mammography. Twenty normal cases and 12 cases with micro-calcifications were included. The performance evaluated using ROC and LROC (ROC with localization) analysis. Results showed that the detectability of micro-calcifications is improved by unsharp-mask filtering. To reduce the noise effects of conventional linear unsharp-masking method, Siddharth *et al.,* (Sidharth *et al.,* 2012) introduced a new unsharp masking algorithm using a non-linear enhancement function. The input image is simultaneously processed using the improved high pass filter which is insensitive to noise and the non-linear enhanced function; both images are then combined to get the final enhanced image. Results showed that the proposed algorithm succeed to enhance the edge of masses and suppress the background noise.

1.1.1.4.Fixed neighborhood statistical enhancement (FNSE)

This technique uses statistical properties in a pixel neighborhood to estimate the background, suppress it and increase local contrast. This enhancement technique presented (Gordon and Rangayyan, 1984) to aid diagnosis of breast cancer without requiring additional X-ray dose. The mammographic image is first digitized using dynamic range expansion procedure; a pixel operator is then applied to the image, which performs contrast enhancement according to a specified function. The FNSE technique has better performance to remove non-homogeneous background from mammogram images (Thangavel *et al.,* 2005).

1.1.1.5.Adaptive neighborhood contrast enhancement (ANCE)

This technique is similar to (FNSE), however, there is some adaptation of the size of the neighborhood to the local properties. The details can thus be enhanced while introducing fewer artifacts. A preference study applied (Sivaramakrishna *et al.,* 2000) to compare the performance of four enhancement techniques (gradient operators, CLAHE, ANCE, and wavelet-based enhancement) applied on forty mammogram images containing masses and

micro-calcifications. For microcalcifications, results showed that, among the different algorithms, the ANCE algorithm is the most preferred in 49% of the interpretations.

1.1.1.6.Wavelet based enhancement

Wavelet transform is a powerful tool for filtering images, it decomposed an image into a set of frequency channels, which having a constant bandwidth in a logarithmic scale (Bai and Liu, 2011). The wavelet transform is suitable for enhancing mammogram images as it helps in visualization of features without amplifying noise. This technique was used in (Yousefi, 2015). A new enhancement method based on wavelet introduced in (Devi and Mini, 2015). The proposed method used stationary wavelets transform (SWT), the absolute maximum wavelet coefficient, and unsharp-masking. The method applied on MIAS database, the performance is evaluated using measures as contrast and (PSNR) and found to be used as a pre-processing step in CAD for all categories of mammographic images as masses, and calcifications.

1.1.2. Pectoral muscle identification

The pectoral muscle is a large fan shaped muscle that covers much of the front upper chest, and it's appearing in mammogram image in MLO view. The representation of the pectoral muscle on the MLO view is a key component in assessing the adequacy of patient positioning and therefore, the adequacy of the acquisition. The pectoral muscle represents a high density area, which can strongly affect the result of image processing. Hence, for better detection accuracy of abnormalities, the pectoral muscle should be first detected and removed.

Several methods have been proposed to identify pectoral muscles in mammograms. Some researchers (Suckling *et al.,* 1995) used a multiple linked self-organizing neural network to segment the pectoral muscle. However, to generate satisfactory results, this method requires a good training dataset. Masek (Masek *et al.*, 2001) employed both a threshold-based algorithm and a straight line fitting technique to segment the pectoral muscle (Masek *et al.,* 2001). Ferrari (Ferrari *et al.*, 2004) highlighted the pectoral muscle using a set of Gabor wavelet filters. Ma (Ma *et al.*, 2007) describes two segmentation methods based on adaptive pyramids

and minimum spanning trees. Camilus (Camilus *et al.,* 2010) used a graph-cut based image segmentation method and Bezier curve smoothing. For muscle detection, a discrete time Markov chain and active contour model proposed in (Wang *et al.,* 2010). In (Santle *et al.,* 2011) the pectoral muscle was identified based on watershed transformation.

Although, there are many different approaches for pectoral muscle identification, till now the most used segmentation algorithm is based on Hough transform (Ferrari *et al.,* 2000), (Weidong and Shunren., 2003) to find a straight line in the mammogram.

Hough Transform

The Hough transform (Hough, 1962) is a technique, which can be used to isolate objects having a particular shape within an image. For pectoral muscle segmentation, the Hough transform was first exploited by (Karssemeijer, 1998). There are two types of Hough transform, the classical one and the geometrical one. For pectoral muscle identification, we restrict the main focus to the classical transform. The classical Hough transform is most common used for detection of regular curves such as lines and circles.

For line detection in a plane (Duda, and Hart, 1972) as it is the case for identifying the pectoral muscle, each input measurement (coordinate point) indicates its contribution to a globally consistent solution: the physical line, which gave rise to that image point. The conventional equation for describing a given line uses parametric definition based on the normal line:

$$x\cos\theta + y\sin\theta = r \tag{3-8}$$

where r is the length of the normal from the origin to this line, and θ is the angular orientation of the normal axis with respect to the X-axis. It occurs that for each point (x, y) on the line, r and θ constants, see [Figure (3-2)].

In an image analysis context, a binary edge detector can estimate the coordinates of the points of the edge segments (x_i, y_i) in the image. Therefore these points serve as constants in the parametric line equation, while r and θ are the unknown variables, which have to be estimated are seek. All possible (r, θ)values defines by each(x_i, y_i), in Cartesian image space map to curves in the polar Hough parameter space were plotted. This point-to-curve

transformation is the Hough transformation for straight lines. The transform is implemented by quantizing the Hough parameter space into finite intervals, which are called accumulator cells. As the algorithm runs, each (x_i, y_i) is transformed into a discretized (r, θ)curve and the accumulator cells which lie along this curve are incremented.

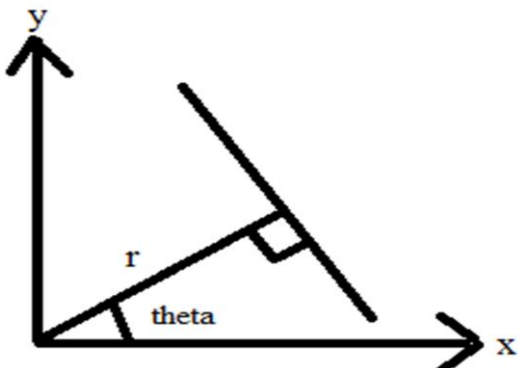


Figure (3-2): Parametric descriptions of a straight-line for Hough transform

1.2. Segmentation techniques used in mammogram

Segmentation is an important issue in image analysis. In a segmentation process the regions of interest corresponding to a specific object are isolated. There are two kinds of segmentation in mammogram images: breast region isolation, and suspicious region isolation. Segmentation algorithms for gray images are generally based on two basic properties of image intensity values: discontinuity and similarity (Maitra *et al.*, 2012)

1.2.1. Segmentation using one single view

Segmentation using a single view of a mammographic image relies on the principle that pixels inside a mass have different characteristics from the other pixels within the breast area. The characteristics used can be simply related to intensity values, texture measures or morphological features. Techniques used for the detection and segmentation of masses can be divided into supervised and unsupervised approach.

Supervised segmentation: also known as model-based segmentation relies on prior knowledge of objects and background regions that are being segmented. Supervised segmentation methods also include template matching approaches, in which the training set contains templates or patterns of objects that can be detected (Tourassi *et al.*, 2003). Numerous researchers (Constantinidis *et al.*, 1999) (Constantinidis *et al.*, 2000) used matching approach in segmentation of mammogram images, as in (Lai *et al.*, 1989), for template matching step: suspicious areas are identified by thresholding the cross-correlation values, and a percentile method is used to determine a threshold for each film, then two tests are used to remove false alarms from the resulting candidates. The main limitation of model based segmentation methods is their reduced effectiveness in case of irregular masses with spiculated margins that are difficult to isolate (Marcin, 2017).

Unsupervised segmentation: this method relies on dividing the image into areas that are different or uniform with regard to some *a priori* defined features. There are three main groups of unsupervised segmentation approach: region-based methods, contour-based methods, and thresholding-based methods.

1.2.2. Region-based segmentation methods

Region-based segmentation relies on the principle of homogeneity; the criteria of homogeneity such as gray level, texture, shape, and model. The basic purpose of region based segmentation method is to segment an entire image R into smaller sub-images (R_i), i=1, 2, 3,…, N., which satisfy the following conditions:

$$R = \bigcup_{i=1}^{N} R_i \; ; \; R_i \cap R_j = \emptyset \tag{3-9}$$

$$P(R_i) = True, i = 1,2, \dots, N \tag{3-10}$$

when, R_i and R_j are adjacent:

$$P\left(R_i \cup R_j\right) = False, \; when \; i \neq j \tag{3-11}$$

where: P(R_i) is logical predicate defined over the points in set R_i

Two basic strategies of region-based methods are region growing, and split and merge approaches.

1.2.2.1.Region growing method

In 1976, Zucker reviewed region growing algorithms(Zucker, 1976). These methods are based on finding a set of seed pixels in the image, and then grow iteratively and aggregate with the pixels that have similar property (homogeneity criterion), [Figure (3-3)]. The key issue of region growing is to find a criterion that decides whether the gray level values of neighboring pixels are similar, within a specific deviation from the seed or current region, and this criterion depends on the applied enhancement method (Rangayyan *et al.,* 1997). The other key issue is to find the suitable seeds; there are numerous automatic seed selection methods (Lee *et al.,* 2000).

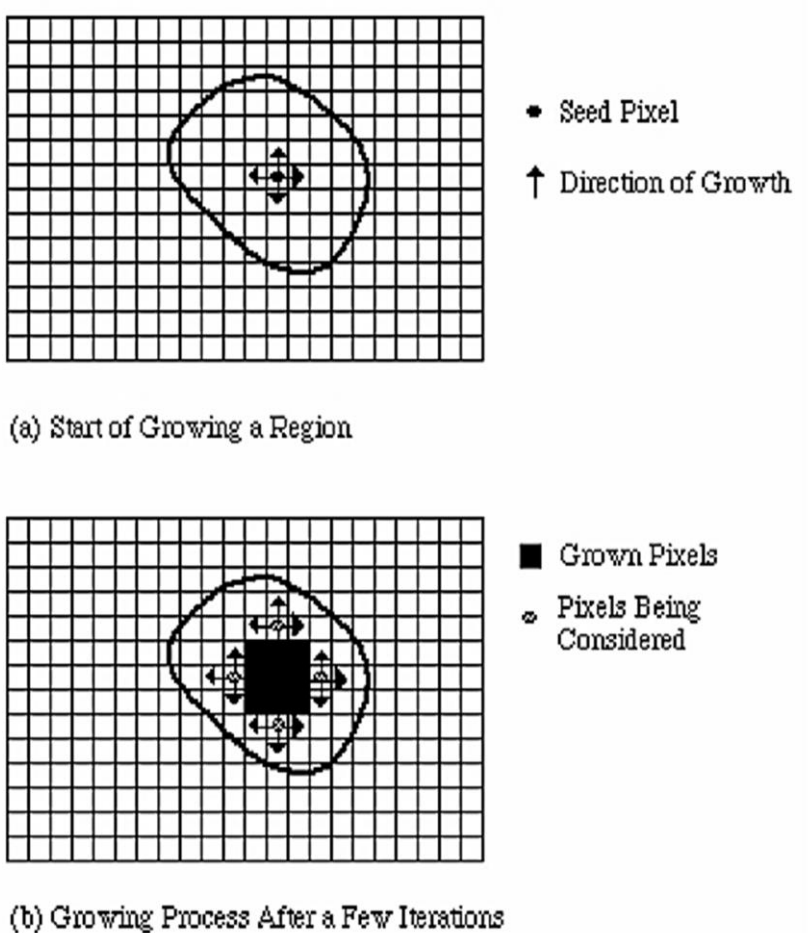


Figure (3-3): Region growing criteria

Region growing algorithms have been widely used for the segmentation of mammographic mass. Huo and Giger (Huo *et al.,* 1995) developed a semi-automatic region growing algorithm, by manually placing the seed points then automatically computing the growing region. Kupinski and Giger (Kupinski and Giger, 1998) introduced two approaches of region growing, the first one incorporated the Radial Gradient Index, which is a measure of the average proportion of gradient, which are radially directed outwards. The second approach is based on a probabilistic method in which the probability of belonging to one region is modeled by a non-Gaussian distribution; the background being modeled using a uniform probability. Guliato (Guliato *et al.,* 1998) proposed a pixel based algorithm, the proposed algorithm preserved the transition between masses and normal tissue to segment the mass boundary. Petrick (Petrick *et al.,* 1999) introduced gradient information with the objective to reduce merging of adjacent and overlapping structures, and select seed point by defining local maxima in the original image. Zheng (Zheng *et al.,* 2003) used an edge image as the starting point; the image was obtained by subtracting two filtered images using different Gaussian filters. Mudigonda (Mudigonda *et al.,* 2001) used multilevel threshold to detect closed edges for mass segmentation. Robottino (Robottino *et al.,* 2008) proposed a region growing segmentation algorithm for mass contour extraction, the algorithm starts from one pixel (seed) and then, expands the area around the seed to include nearby pixels falling within a threshold range.

1.2.2.2.Split and merge method

The split and merge technique is another classical region-based segmentation method. The process consists of recursively splitting the image until all the regions meet a homogeneity criterion. In an accompanying step, all adjacent regions satisfying a second homogeneity criterion are merged, [Figure (3-4)]. For mammographic mass segmentation, this approach has been used by (Rangayyan *et al.,* 1997). They begin with a hand-selected region of interest containing a single mass.

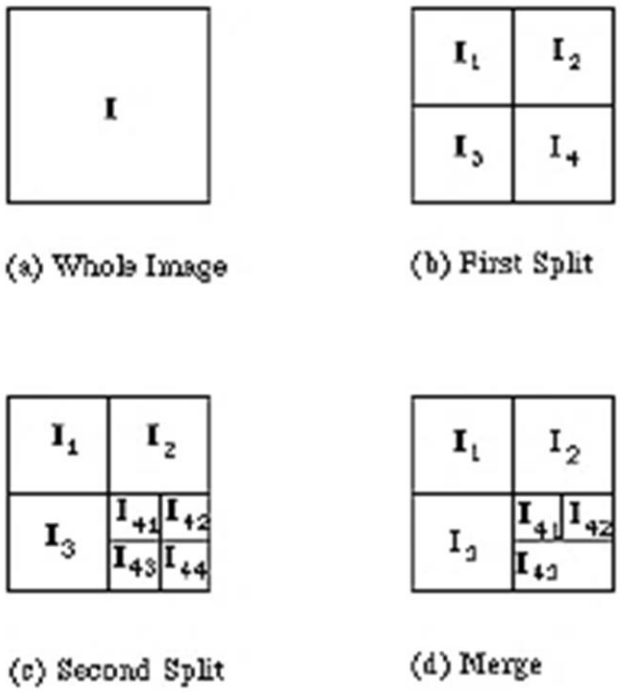


Figure (3-4): Region splitting and merging criteria

1.2.3. Contour –based methods

Image segmentation techniques based on edge detection have been introduced since 1965 by Roberts (Roberts, 1965). Contour based methods work on finding the boundaries of the regions of interest. Algorithms are based on filtering the image in order to enhance edges prior to the detection stage (Hmida *et al.*, 2017). Many operators: Robert gradient, Sobel gradient (Viton *et al.*, 1996), Prewitt gradient, Laplacian operator, etc. can be used to enhance edges. A good edge detector should satisfy three conditions (Canny, 1986). First condition is to minimize errors: probability of marking non-edge and losing edge pixels should be low. The second condition is that detected edge pixels should be as close as possible to real edges. The third condition is that the boundary width should be of one pixel.

Most of the research conducted using the contour-based methods segment masses rather than micro-calcifications. Contour methods are able to find the boundaries, but can be approximate. However, active contour methods are able to integrate image pixels into smooth connected borders that better delineate the shape being first approximated (Marcin, 2017). Active contour-based methods can be classified into snake-based methods and level sets,

which differ in their mathematical implementation. The boundary in snake evolves explicitly, while it evolves implicitly in level set.

For segmentation of suspicious mass regions in mammograms, Petrick (Petrick *et al.*, 1996) introduced a dedicated algorithm for edge detection, first enhancing the image using an adaptive density-weighted contrast enhancement filter in combining with Laplacian-Gaussian edge detector. Kobatake and Yoshinaga (Kobatake and Yoshinaga, 1996) presented a contour based algorithm for spicules detection using first the gradient information, then the morphological line-skeletons are extracted in order to detect spicules, a modified Hough transform is finally applied to extract lines passing near the center of mass. The ROI corresponding to the mass is delineated from these lines.

Yuan (Yuan *et al.,* 2007) developed a method for mass lesion segmentation on mammograms using a geometric active contour model. In their method, a radial gradient index (RGI)-based segmentation was applied to obtain an initial contour, and then automatic background estimation was applied to identify the effective circumstance of the lesion, and a dynamic stopping criterion was implemented to terminate the contour when it reached the lesion boundary. Song (Song *et al.*, 2009) developed an automatic mass segmentation method; the method used plane fitting and dynamic programming. First the plane fitting applied to a background to obtain the edge candidate points. Second, the dynamic programming applied to find the *best* contour of the mass from the edge points. Muralidhar (Muralidhar *et al.,* 2001) introduced a mass classification method based on the "snakules" segmentation method, which is an evidence-based active contour algorithm.

1.2.4. Clustering methods

Clustering methods are among the most commonly used techniques for the delineation of both breast masses and calcifications. Region clustering and region growing are very similar, but clustering does not need a starting point or some prior information, so clustering searches the regions directly (Cheng *et al.,* 2006). A variety of clustering algorithms have been introduced to make the segmentation more effective. Most common clustering algorithms applied to mammograms are the following:

3.2.4.1.K-means clustering

A traditional partition k-Means algorithms are under the group of squared error based clustering (MacQueen, 1967), which is characterized by easy implementation and low complexity. In statistics and machine learning, K-means clustering is a method which aims to partition image into (k) clusters in which each partition belongs to the cluster with the nearest mean. Let $X_n = (x_1, x_2, \dots, x_n)$, where each component is a d-dimensional real vector , the algorithm attempts to partition the (n) observations into k- sets, $S = \{S_1, S_2, \dots, S_k\}$ with (k<n). The basic algorithm as follows:

1. Pick (k) cluster centers, either randomly or according to some methods as (Hamerly and Elkan, 2002), so as to minimize the within-cluster sum of squares, as in following:

$$\underset{S}{argmin} \sum_{i=1}^{K} \sum_{x_j \in S_j} \left\| X_j - \mu_i \right\|^2 \tag{3-12}$$

where μ_i is the mean points in S_i. Given an initial set of k means $m_1^{(1)}, m_2^{(1)}, \dots, m_k^{(1)}$

2. Assign each observations in the image to the cluster, which minimizes the distance between the pixel and the cluster center, as:

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\| \leq \left\| x_p - m_j^{(t)} \right\| for\ all\ j = 1, \dots, k \right\} \tag{3-13}$$

where each x_p is assigned to exactly one $S^{(t)}$

3. Re-compute the cluster center's by averaging all of the pixels in the cluster, the new mean is calculated as follows:

$$m_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_j \in S_i^{(t)}} x_j \tag{3-14}$$

Repeat step 2 and step 3, until the algorithm has converged, as no pixels change clusters.

For mass segmentation, this algorithm has been used for instance by (Sahiner *et al.*, 1998) (Li *et al.*, 2002) (Matrins *et al.*, 2009). An adaptive k-means clustering algorithm implemented for breast image segmentation for micro-calcifications detection by (Patel and Sinha, 2010).

3.2.4.2.Fuzzy C-means clustering (FCM)

Is a method of clustering which allows one piece of data to belong to two or more clusters (Bezdek *et al.*, 1984). FCM is popularly used for automated segmentations; it can provide better results than K-means and modified K-means clustering algorithms (Christ and Parvathi, 2011). FCM algorithm is used to find cluster center that minimize a dissimilarity function. K-means algorithm takes the mean of the weighted cluster, while FCM algorithm considers that each point has weighted value associated with cluster. However, the performance of k-means and FCM algorithms are based on initial choice of weights (Ramani *et al.*, 2013).

The FCM algorithm (Yang *et al.*, 2005) attempts to: i) partition a finite collection of n elements $X = \{x_1, x_2, \dots, x_n\}$ into a collection of c fuzzy clusters, ii) given a finite set of data, the algorithm returns a list of c cluster centres $C = \{c_1, c_2, \dots, c_n\}$ and a partition matrix $W = w_{ij} \in |0,1|, i = 1,2, \dots, n,\ j = 1,2, \dots, c,$ where each element w_{ij} tells the degree to which element x_i,belong to cluster c_j. iii) The FCM aims to minimize an objective function:

$$\underset{c}{argmin} \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} \left\| x_i - c_j \right\|^2 \qquad (3\text{-}15)$$

where

$$w_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \qquad (3\text{-}16)$$

and

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}, \qquad (3\text{-}17)$$

where m is the hyper-parameter that controls how fuzzy the cluster will be. The higher it is, the fuzzier the cluster will be in the end.

For mass segmentation in mammogram, FCM was used by (Velthuizen, 2000) and recently by (Vijayalakshmi *et al.*, 2016). In (Pavan *et al.*, 2016), FCM clustering used to automatically segment fibroglandular and adipose tissues from breast mammography, also, an adaptive k-class fuzzy c-means clustering presented (Keller *et al.*, 2011) for fully-automated identification and quantification of breast density.

3.2.4.3.Expectation maximization clustering (EM)

The EM algorithm (Dempster *et al.*, 1977) provides an iterative and computationally simple algorithm based on the incomplete data (Chen and Lee, 1997). In EM clustering, alternating steps of expectation (E) and maximization (M) are performed iteratively till the results converge. The E step computes an expectation of the likelihood by including the latent variables as if they were observed, and maximization (M) step, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the last E step. The parameters on the M step are used to generate the next E step. Mathematically, E and M steps express as following (Gupta and Chen, 2011):

E-step: gives the estimate from the previous iteration$\theta^{(m)}$, compute the conditional expectation function as:

$$Q\left(\theta\middle|\theta^{(m)}\right) = \int_{X(y)} \log p\,(x|\theta) p\left(x\middle|y,\theta^{(m)}\right) dx \qquad (3\text{-}18)$$

where $\theta \epsilon \Omega$ for some set Ω space$p(x|\theta)$ is parametric density, $p\left(x\middle|y,\theta^{(m)}\right)$is the conditional probability distribution for the complete data x

M-step: the $(m+1)$ guess of θ is:

$$\theta^{m+1} = \underset{\theta \epsilon \Omega}{argmax}\, Q\left(\theta\middle|\theta^{(m)}\right) \qquad (3\text{-}19)$$

The EM algorithm is used for mammogram segmentation, in (Chen and Lee, 1997). A modified EM algorithm developed (Vedanarayanan and Nandhitha, 2017) to enhance cancer detection.

3.2.5. Thresholding methods

Threshold methods have been widely used for image segmentation, and they can be considered as a special case of partition clustering methods. Global thresholding has been widely used for mammogram mass segmentation (Dominguez and Nandi, 2007). It is based on the global information that can be summarized in a histogram. On the histogram, the regions with abnormality impose extra peaks while a healthy region has only a single peak (Cheng *et al.,* 2006). Local thresholding is better for mass detection than global thresholding, because a local threshold value is determined locally for each pixel based on the intensity values of the surrounding pixels. At least two parameters are considered in local thresholding, the window size and the threshold value (Kallergi *et al.,* 1992).

Kom (Kom *et al.,* 2007) developed a local adaptive thresholding technique for mass segmentation, the proposed algorithm was tested on 61 mammograms which containing masses, results obtained a sensitivity of 95.9%, and the area under ROC was 0.946 for enhanced image and 0.938 for non-enhanced image. An algorithm proposed (Makandar and Halalli, 2016) to detect region of mass using morphological threshold based segmentation technique. The algorithm applied on 55 images of MIAS database and the achieved accuracy is 94.5% in segmentation of mass.

In (Dominguez and Nandi, 2007), a method for detection of masses in mammogram is presented. As part of this method, a segmentation of regions based on multiple level thresholding is proposed; a set of features is computed from each of the segmented regions. The method was tested on 57 images of masses from MIAS database, including circumscribed, spiculated, and ill-defined masses. The proposed method achieved a sensitivity of 80% at 2.3 FPs/image.

3.2.6. Segmentation using multiple images

Mass segmentation using more than one image relies on the comparison of different mammographic images of the same woman (Kopans, 1998). These different mammographic images can be left and right mammograms, two mammographic views of the same breast, or same view mammograms taken at different times.

The comparison of left and right mammograms is often based on the similarity of the internal structure of both breasts. Mass segmentation method is then based on bi-lateral subtraction. Left and right images are first aligned and subsequently subtracted (Méndez *et al.,* 2003). The comparison of two views of the same breast is known as ipsilateral comparison. Generally, CC and MLO views are compared, abnormalities are segmented on both views independently and results are compared point-to point (Sun *et al.,* 2004), the evaluation using the fibroglandular tissue in the two views. The comparison of the same breast with different time intervals, are based on the radiologists' evaluation of how suspicious regions evolved (Marti *et al.,* 2001).

(Xu *et al.,* 2010) and (Marti *et al.,* 2006) presented algorithms for mass detection using left and right mammograms. (Pu *et al.,* 2008) and (Sun *et al.,* 2004) presented ipsilateral multi-view CAD schemes to detect masses in mammograms. (Timp *et al.,* 2005), (Wai and Brady, 2005), (Timp and Karssemeijer, 2006) are developed algorithms using the same view of mammograms taken at different time intervals.

3.3. Feature extraction categories

Feature extraction is used to denote a piece of information from raw data which is more suitable for classification purpose. Many features have been extracted to characterize the abnormalities of mammograms (Pradeep *et al.,* 2012). The main categories of features are beside gray level intensities, texture, geometric, and gradient features, the following are details of texture features which are used in this research, and brief descriptions of geometric and gradient categories.

3.3.1. Texture features

Texture features represent an attempt to characterize gray level variations between adjacent pixels in the region of interest (Li *et al.,* 1997). Generally, texture features could be studied on at least two levels, statistical and structural (Rosenfeld and Lipkin, 1970). The statistical features estimates statistics of the gray-level histogram in different orders (Bharati *et al.,* 2004), and the structural features describes the architectural composition of the tissue using well-defined elements (Caldwell *et al.,* 2000). Texture features have proven to be

useful to differentiate normal breast tissues from abnormal lesions with masses or micro-calcifications.

There are different statistical textural approaches: first-order, second-order and higher-order texture. In first-order, information on texture is extracted from the histogram of image intensity; this approach depends only on individual pixel values and not on the correlation or co-occurrence of neighboring pixel values. Mean value of gray levels, variance, kurtosis, and skewness are examples of first-order statistics features (Garra *et al.,* 1993). In second-order, information on texture is based on the probability of finding a pair of grey-levels at random distances and orientation over an entire image; it takes into account correlations, it based on gray-level co-occurrence matrix (GLCM). Finally, higher-order statistics include run-length measures, which describe texture by measuring "runs" of consecutive pixels with similar gray levels and calculate statistics based on the number and length of such runs (Galloway, 1975).

3.3.1.1.Grey level co-occurrence matrix (GLCM)

The Grey Level Co-occurrence Matrix GLCM is a well-established robust statistical tool for extracting second order texture information from image. A co-occurrence matrix offers greater information about the inter pixel relationship, periodicity and spatial grey level dependencies.

Mathematically, a GLCM is a matrix where the number of rows and columns is equal to the number of gray levels G in the image. The matrix element $P(i,j|\Delta x,\Delta y)$is the relative frequency with which two pixels separated by a pixel distance$(\Delta x,\Delta y)$, occur within a given neighborhood, one with intensity i and the other with intensity j at a particular distance d and a particular angle θ. Given an $M \times N$ neighborhood of an input image containing G grey levels from 0 to $G-1$, let $f(m,n)$ be the intensity at sample m, line n of the neighborhood.

Then

$$P(i,j|\Delta x,\Delta y) = WQ(i,j|\Delta x,\Delta y) \tag{3-20}$$

where

$$W = \frac{1}{(M-\Delta x)(N-\Delta y)} \tag{3-21}$$

$$Q\langle i,j|\Delta x,\Delta y\rangle = \sum_{n=1}^{N-\Delta y}\sum_{m=1}^{M-\Delta x} A \tag{3-22}$$

and

$$A = \begin{cases} 1,\ if\ f(m,n) = i\ and\ f(m+\Delta x, n+\Delta y) = j \\ 0,\ elsewhere \end{cases} \tag{3-23}$$

Co-occurrence matrices are calculated in the normalized database to ensure that their features are not influenced by the regions limitation. The elements of $P_d[i,j]$ can be normalized by dividing each entry by the total number of pixel pairs. Normalized GLCM $N[i,j]$ defined by following equation:

$$N[i,j] = \frac{P[i,j]}{\sum_i \sum_j P[i,j]} \tag{3-24}$$

3.3.1.2.Fractal based texture features

Structural approaches characterize texture as being composed of simple primitives called "texels" (texture elements). The structural features include fractal dimension as estimated by the box-counting approach (Peitgen *et al.*, 1992), the local binary pattern which capture intensity variations between central and neighboring pixels, and the edge enhancing index feature which describes the directionality of flow-like structures within the breast (Zheng *et al.*, 2015).

The fractal refers to the complex pattern that recurs at various scales. Fractals have proven to be a mathematically elegant of generating textured surfaces (Pentland, 1984). In fractal based texture analysis the fracture will measure the geometric complexity, which describes the spatial pattern of textures (Srinivasan and Shobha, 2008). Shapes of fractal objects keep invariant under successive magnifying or shrinking the object. Hence, fractal geometry can be applied to overcome the scale problem of texture. The fractal dimension and fractal signature are good parameters to measure descriptive value of a region (Don *et al.*, 2012). The fractal dimension (FD) is a statistical quantity which measure roughness of a geometric region; it provides the degree of linear independence and correlation between the

available features (Mandelbort, 1982). Since the cancer grows in an unexpected way, we can also expect malignant masses to have high FD, if we focus exclusively on morphology.

3.3.2. Geometric features

Geometric features describe the geometric properties of the region of interest (ROI), The basic geometric features are area, perimeter, and compactness (Matsubara *et al.*, 1997). Geometrical shape, margin and texture features, are used (Valarmathie *et al.*, 2016) to classify masses into benign or malignant. The shape features used are compactness, dispersion, eccentricity, elongatedness, and roundness. To classify micro-calcifications into benign or malignant, 17-shape features and 44 texture features are extracted in (Zadeh *et al.*, 2001).

3.3.3. Gradient features

The gradient feature vector is a quantity comprising of gradient magnitude as well as its directional component; these components are computed by applying derivatives in both horizontal and vertical directions. Gradient operator generates a 2D gradient vector at each image point such that the vector points towards the direction of the largest possible intensity increase or decrease and its magnitude corresponds to the ratio of changes in intensities in that direction (Aggarwal *et al.*, 2015). The Sobel operator is a usual gradient operator to compute the gradient images.

3.4. Image classification

Image classification is an important task after extracting the features because it aims at assigning the extracted objects into different categories. Generally, there are two main classification strategies:

1. **Supervised classification**: the labeling is a priori known and the algorithms such as minimum-distance classification, and support vector machine (SVM) can be used to optimize the classification.

2. **Unsupervised classification**: the labeling is unknown and the algorithms proposed some classes based for instance on clustering in the image, using K-means and hierarchical clustering approaches.

3.4.1. Classification of breast based on density

Breast density assessment is an important component of the screening mammography report and conveys information to referring clinicians about mammographic sensitivity and the relative risk of developing breast cancer (Winkler et al, 2015). The origins of breast density classification are due to Wolfe (Wolfe, 1976), who showed the correlation between breast density and the risk of developing breast cancer, classifying the parenchymal patterns into four categories. Subsequently (Boyd *et al.*, 1995), showed a similar correlation between the relative area of dense tissue and mammographic risk. They developed a method to measure a percentage of breast densities from mammography using a computer-aided technique and divided mammograms into six categories. Many factors influence tissue density, such as age, endogenous and exogenous hormones, lactation, and previous chemotherapy and radiation therapy.

3.4.1.1.Qualitative classification of breast density

There are several approaches for qualitative classification of breast density. According to (D'Orsi, 1998), the breast can be classified into three categories, depending on the relative amounts of glandular tissue versus adipose tissue:

1. Glandular breast: mainly in women under 30 years
2. Fatty and glandular breast: mainly for women between 30 and 50 years
3. Fatty breast: mainly in women over 50, since at the end of the reproductive life, the breast loses fibrous mass and turn into fat.

The Wolfe classification (Wolfe, 1976) is a visual classification method, which classified breast into four categories:

1. N1 corresponds to fatty normal breast
2. P1 corresponds to prominent ducts occupying less than 25% of the breast

3. P2 corresponds to prominent ducts occupying between 25% and 75% of the breast
4. Dy corresponds to breast dysplasia and is extremely dense.

Using mammographic parenchymal patterns, Tabár et al. have proposed a mammographic modeling scheme based on mixture of four building blocks composing the normal breast: nodular, linear, homogeneous, and radiolucent block. Nodular densities corresponds to Terminal Ductal Lobular Units (TDLU); linear densities correspond to either ducts, fibrous or blood vessels; homogeneous densities correspond to fibrous tissue which appears as bright areas in mammographic images; radiolucent densities are related to fibrous tissue which appears as dark areas in mammographic images (Tabar *et al.*, 2004).

Presently, the most used breast density classifications is the BIRADS classification (Muhimmah *et al.*, 2006), which was developed as a quality assurance tool, and covers the significant relationship between increased breast density and decreased mammographic sensitivity in detecting cancer (Sickles, 2007). The classification of BI-RADS is divided into four categories according to their density:

1. Category1: breast is mostly made up of fat: breast density $< 25\%$.
2. Category2: breast density is between 25% and50%.
3. Category3: breast density is between 51% and75%.
4. Category 4: breast is extremely dense with breast density$> 75\%$.

Due to the strong positive correlation between breast cancer and breast density, researchers have developed different methods, possibly integrated in CAD systems for semi-automated and automated tissue classification of mammogram images.

(Saha et al, 2001) used a scale-based fuzzy connectivity method to extract dense tissue regions from mammographic images. (Selvan *et al.*, 2006) used a heuristic optimization approach to estimate the breast density model parameter set. The approach was applied to different categories of mammogram from MIAS database, it yielded lower floor of estimation error in 109 out of 112 cases (97.3%), and 101 out of 102 cases (99.0%), for the number of regions being five and eight respectively.

(Petroudi *et al.*, 2003) proposed an algorithm, which defined texture classes as a statistical distribution over texton dictionaries developed from a training set. A classification was done

using an appropriate distance measure for the data, which was obtained from the training dataset. (Torrent *et al.*, 2008) presented a comparison of multiple thresholding algorithms, FCM clustering algorithm, and one region based algorithm for segmenting fatty and dense tissue in mammogram images. The performance of algorithms is evaluated using ROC analysis. Results demonstrated that the use of region information is useful to obtain homogeneous region segmentation.

3.4.1.2.Quantitative calculation of breast density

Currently, U.S. Food and Drug Administration (FDA) approved software for fully automated volume density percentage calculation. Then the calculated volume density percentage is converted into the appropriate BI-RADS density category. These software are:

Quantra software (Ciatto *et al.*, 2012): the fibroglandular density per pixel is estimated by using known image acquisition parameters, including breast thickness; and the values for each pixel are then added to determine the volume of the fibroglandular tissue.

Volpara software (Jeffreys *et al.*, 2010**)**: a pixel value representing fat is first identified automatically by the software to provide a reference value; each individual pixel in the breast are then compared with the reference value to estimate x-ray attenuation and generate a density map; the volumetric breast density determined using the calculation of fibroglandular tissue divided by the total volume of tissue within the breast.

3.4.2. Lesions detection and classification

The goal of the detection stage is to assist radiologists in locating abnormalities on the mammogram. The presence of masses or of clusters of micro-calcification in mammograms is an indicator of early stage breast cancer, so most of approaches focused on the detection and classification of masses and micro-calcifications. Various CAD systems have good performance in the detection of micro-calcifications and masses.

3.4.2.1.Detection and classification of micro-calcifications

Micro-calcifications usually appear brighter than their surroundings. In dense breasts as in some younger women, suspicious areas almost invisible and micro-calcifications easily misdiagnosed. Micro-calcifications have a variety of size, shapes, and distributions. Mathematical morphology has been suggested to enhance their detection, since it can be applied to extract small details without affecting other image details (Durate *et al.,* 2011), (Thangaraju *et al.,* 2012). Other strategies have been proposed for detecting micro-calcifications, including random field models (Karssemeijer, 1993), fuzzy logic (Cheng *et al.,* 1998), SVM and artificial neural networks (Ren, 2012).

In recent years, researches devised methods based on multi-scale representation for several applications related to mammographic image processing. Wavelet analysis is for instance well adapted to generate such a representation (Zhang *et al.,* 1998) (Juarez *et al.,* 2006). Wavelets have indeed finite square supports and are ideal for capturing point discontinuities but not edges, so, using wavelets is very effective in micro-calcifications detection (Sampat *et al.,* 2005). In (Mini *et al.,* 2004), micro-calcifications detection using wavelet based edge detectors was performed. Detectors used are Marr-Hildreth (M-H) detector, and Canny detector. Both detectors achieved 95% detection accuracy for the MIAS database. In (Bouyahia *et al.,* 2009), different wavelet techniques for automatically micro-calcifications detection are applied on MIAS database. Techniques are: un-decimated wavelet transform, wavelet packets transform, two-dimensional multiscale product, and one-dimensional wavelet transform modulus maxima. The obtained detection rates of micro-calcifications are: 85%, 87%, 91%, and 95% respectively

The parenchymal and ductal patterns in mammograms have high local self-similarities which are the basic property of fractal objects. These tissue patterns can be “reproduced” by fractal models, and extracted from the original image. A fractal modeling of background tissues for the enhancement of micro-calcifications was presented in (Li *et al.,* 1997) (Sankar and Thomas, 2010). When self-similar geometrical objects are defined, their irregularities can be tested by analyzing their fluctuations at a different resolution. This property was found useful in breast medical images of breasts, where the fractal dimension (FD), a value describing how the irregular structure of objects is replicated at different scales can be used to detect abnormalities (Shanmugavadivu and Sivakumar, 2013).

3.4.2.2.Detection and classifications of masses

Masses are characterized by their shape and margin properties. Numerous studies focused on the detection of spiculated masses which present higher risk of malignancy. As spiculated masses are characterized by spicules radiating in various directions, some approaches calculate the edge orientation at each pixel, as in (Kegelmeyer *et al.*, 1994), the algorithm for the detection of spiculated lesions was applied on 85 four-view clinical cases and achieved 100% sensitivity with a specificity of 82%. Automatic mass contour extraction scheme have been proposed (Nakagawa *et al.*, 2004) based on active contour model (Snake). In this technique, the central point of a mass is determined, then an initial contour is arranged, and a control point's direction of movement is limited to directions radiating from the central point. The method applied on 53 digitized mammograms having masses. The results compared with correct segmentation describe by doctor's sketches, the number of cases corresponded with the correct one at $81-100\%$ and $61-80\%$ were 25 and 12 respectively. So, results of this technique show that it may be effective in improving performance of CAD system.

Feature extraction based on single-pixel has poor performance for mass detection, because it does not take into account the spatial arrangement of the pixels which is a very important factor to discriminate mass lesions. In contrast, region based feature extraction has good performance in mass detection because it takes into account this spatial information (Berber *et al.*, 2013).

A multi-resolution and multi-orientation wavelet transform have been used for mass detection (Qian *et al.*, 1999) (ke *et al.*, 2009). An algorithm (Abdul-Jaleel *et al.*, 2014) proposed to classify mass lesion to benign or malignant based on discrete wavelet transform (DWT). The algorithm applied on 148 mammogram images taken from MIAS database. The classifiers used are K-nearest neighbor (K-NN), support vector machine (SVM), and radial basis function neural network (RBFNN). It is found that the best performance 94.6% obtained using the RBFNN with DWT features.

For mass lesion detection, a fractal based methods have been presented (Guo *et al.*, 2009) (Shanmugavadivu *et al.*, 2016). In (Beheshti *et al.*, 2014), a fractal method for detection and diagnosis of mass lesion in mammogram is presented. The method consisted of two steps, firstly, the lesions discriminated automatically using fractal dimension, and secondly, using fractal features to identify the roughness in mass contours and determine the extent of

speculation or smoothness of the masses. The proposed method applied on a set of images selected from local hospital and MIAS database. Results achieved high accuracy of 99% and 90% in the classification of spiculated and ill-defined malignant masses from benign tumors, respectively. A study of four methods to compute the fractal dimension of the contours of breast masses was presented (Rangayyan and Nguyen, 2007). For calculating FD, the ruler method and box counting method are computed from 1-dimensional (1D), and 2-dimensional (2D) signature derived from the contour. The methods were applied to different dataset included MIAS database, results achieved by MIAS database are: 80%, 81%, 80%, and 75% respectively for 1D ruler, 2D ruler, 1D box counting, and 2D box counting methods. In (Matsubara *et al.*, 1997), the masses are classified into benign and malignant by the change of fractal dimension, the performance of classification was 100% in thirteen mammograms. (Mudigonda *et al.*, 2001), proposed a method for the detection and classification of masses in mammographic images. The masses are segmented by establishing intensity links from the central portions of masses into the surrounding areas, and analyzing oriented flow-like textural information in mammograms. Features based on flow orientation in adaptive ribbons of pixels across the margins of masses are proposed to classify the region detected as true masses or (FPs). Moreover, the approach is used five texture features based on GLCM and the features in a logistic regression method to classify masses as benign or malignant. The method applied on 56 images including 30 benign, 13 malignant, and 13 normal cases, which selected from MIAS database. The mass detection algorithm achieved a sensitivity of 81% at2.2 $FPs/image$. The mass classification algorithm achieved success rate of 100% and 63% in detecting malignant and benign masses respectively.

3.5. Examples of CAD system for breast cancer detection

Researchers used numerous methods for mammograms classifications. Most of these techniques are supervised machine learning methods (Duda *et al.*, 2000) such as: decision tree (Vibha *et al.*, 2006), linear discriminative analysis LDA (Costa *et al.*, 2007), SVM (Singh *et al.*, 2006) (Liu *et al.*, 2012), ANN (Kallergi, 2004), and Bayesian networks (Burnside *et al.*, 2000). Table (3-1) shows summary of some CAD systems applied on MIAS database for breast cancer detection

Table (3-1): Summary of some CAD systems for breast cancer detection

Reference	MIAS-Dataset Number	Features	Classifier	Results (%)
(Vibha et al., 2006)	N/A	Wavelet	Random forest decision classifier (RFDC)	90 Acc.
(Alolfe et al., 2009)	188	Wavelet, first order statistics, GLCM, Shape, FD	Combined (SVM/LDA)	87.5Sensitivity 90 Specificity
(Jasmine et al., 2009)	N/A	Wavelet	ANN	87 Sensitivity
(Rejani and Selvi, 2009)	75	Morphological features	SVM	88.75 Sensitivity
(Dheeba and Selvi, 2010)	322	Gabor features	Radial basis function networks (RBFNN)	85.2 Sensitivity
(Tahmasbi et al., 2011)	N/A	Zernike moments	Multi-layer perceptron (MLP)	97.6 Sensitivity 97.5 Specificity
(Kabbadj et al., 2012)	16	Statistical and Geometric features	SVM	99.6 Sensitivity 99.11 Specificity
(Shanthi and Bhaskaran, 2014)	192	Texture features	Self-adaptive resource allocation network (SRAN)	98.44 Acc.
(Tariq, 2017)	322	GLCM	ANN	99.3 Sensitivity 100 Specificity 99.4 Acc.
(Kaur et al., 2019)	322	Speed-up robust features (SURF)	Multi-class SVM	98 Acc.

4. Materials and Methods

The process flow diagram for the proposed algorithm is shown in [Figure (4-1)].The mammogram images used in this research are taken from the MIAS database, and the program is written by the MATLAB software, version R2013b. Firstly, mammogram images are preprocessed, segmented, then features are extracted, tissue characterized, and finally the mammogram image classified to normal or abnormal. Moreover, the abnormal lesion classified to benign or malignant.

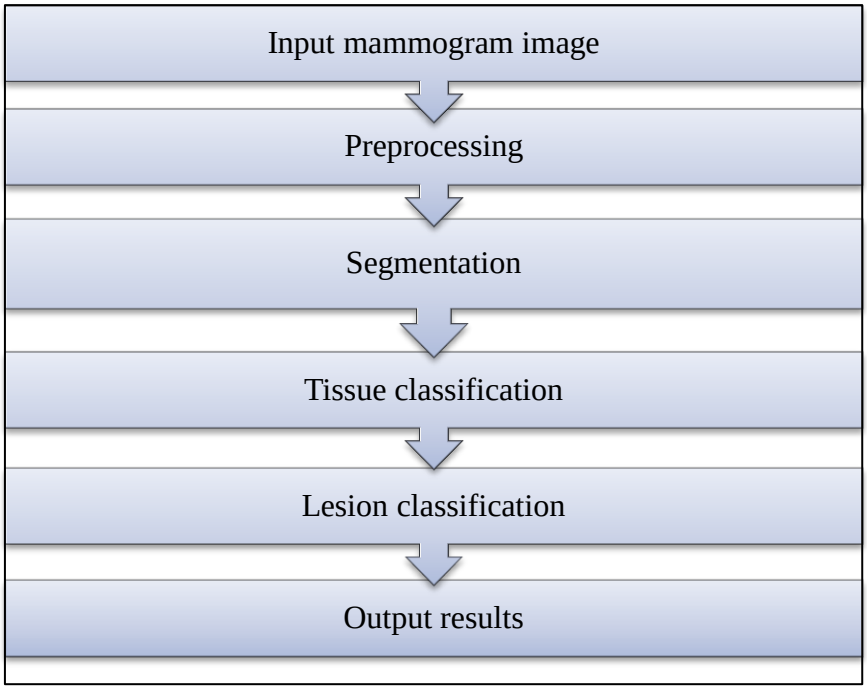


Figure (4-1): Research methodology block diagram

4.1. MIAS database

The MIAS database published in 1994 contains 161 studies, patients with ages ranging from 50 to 65 (Serhat *et al*., 2005), each study includes an image of each breast in (MLO) view, these results of 322 mammography images. Films taken from the United Kingdom National Breast Screening Program have been digitized to 50 micron pixel edge and a gray-scale resolution to 8 bits per pixel (bpp), with a Joyce-Loebl scanning microdensitometer SCANDIG-3. Resolution reduced to 200 microns (mini-MIAS) and fitted to 1024×1024

boundary box. The image files are available in portable network graphics (PNG) format, [Figure (4-2)]. The mammogram images are investigated and labeled by expert radiologists.

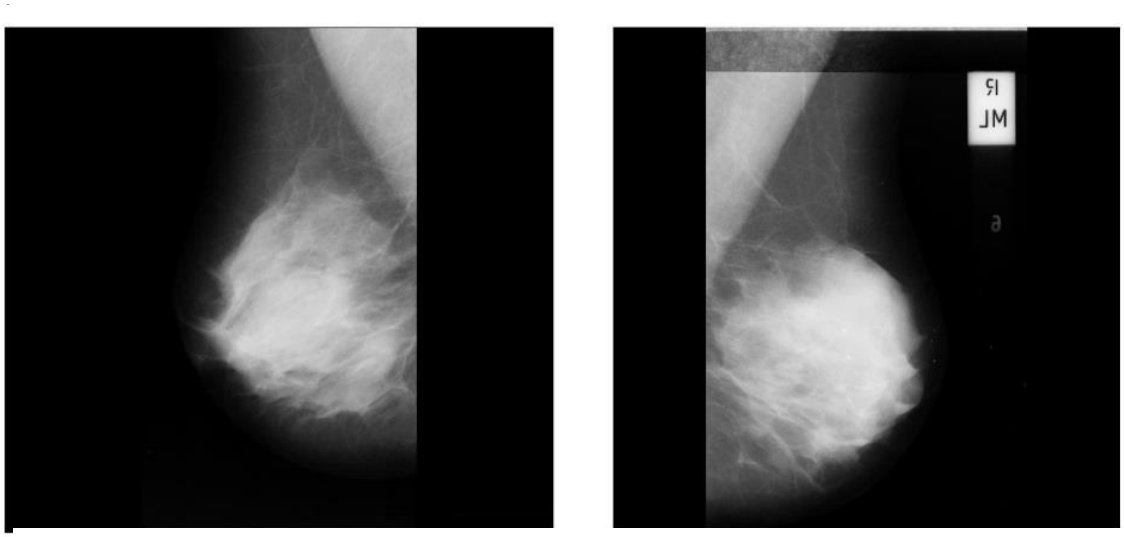


Figure (4-2): Example of MIAS database, left and right breast for one patient, (mdb001 and mdb002)

MIAS detailed information

There is text file contains information of all images [Table (4-1)], the list is arranged in pairs of films, and the information's are:

1st column contains MIAS database reference number, the list is arranged in pairs of films, the left (even reference numbers) and right mammograms (odd reference numbers) of a single patient.

2nd column contains character of background tissue F for Fatty, G for Glandular and D for Dense tissue type.

3rd column contains class of abnormality present, which are

CALC	Calcification
CIRC	Well-defined/circumscribed masses
SPIC	Spiculated masses
MISC	Other, ill-defined masses
ARCH	Architectural distortion
ASYM	Asymmetry
NORM	Normal

4th column contains severity of abnormality, which are (B) for benign, or (M) for malignant.

5th, 6th, and 7th columns contain the (x, y) coordinates of center of abnormality, and approximate radius (in pixels) of a circle enclosing the abnormality, respectively.

Table (4-1): Example of MIAS database details

Reference Number	Tissue Type	Abnormality Type	Abnormality Severity	x-Coordinates	y-Coordinates	Radius
mdb001	G	CIRC	B	535	425	197
mdb002	G	CIRC	B	522	280	69
mdb003	D	NORM				
mdb004	D	NORM				
mdb028	F	CIRC	M	338	314	56
mdb029	G	NORM				

4.2. Technique for preprocessing of mammogram

A mammography images involves excessive parts, [Figure (4-3)], these parts are not important in processing of mammogram image. Some parts of images have some notes and labels which consist of information like name, date, etc., these labels are bright as gray level of some breast tissues which will influence the classification and will not give the desired results (Ponraj *et al*., 2011). The presence of noises and artifacts can disturb the detection of breast cancer and reduce the rate of accuracy in analysis (Chaabani *et al*., 2010). Removing the excessive parts, all irrelevant information, noises and artifacts are the aim of preprocessing.

As the mammogram images are texture in nature, it is very important to get proper enhancement for these images. In low contrast images, the minor difference between normal and the malignant tissue is not discernable and make the interpretation very difficult (Sundaram *et al*., 2011). Hence, preprocessing enlarge the intensity difference between objects and background to produce reliable image for processing step.

In mammograms, the pectoral muscle has nearly homogeneous gray level values and exhibits as a high intensity region. Failure in segmenting the pectoral muscle may cause a higher number of false positives for breast cancer detection (Yanfeng *et al*., 2013). So a method represented to automatic pectoral muscle removal from mammogram images.

The preprocessing code is attached at (appendix A). Also, more detail is in preprocessing published paper attached at (Appendix B).

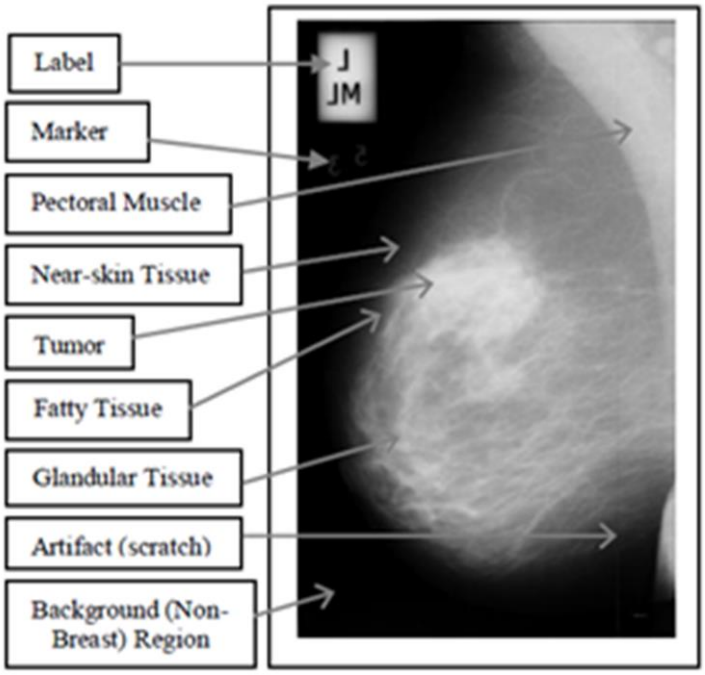


Figure (4-3): Components of mammogram image

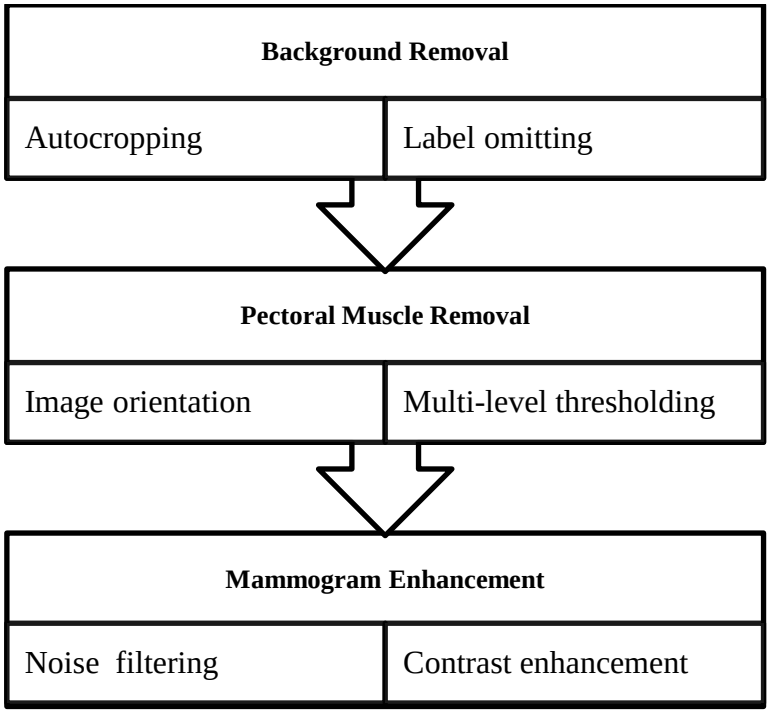


Figure (4-4): A flowchart represents preprocessing process

4.2.1. Auto-cropping & labels omitting

MIAS images have high density noise which characterized by high values of optical densities, such as labels or scanning artifacts. For isolated breast region from irrelative regions, the main common step is image binarization, by converting the mammogram image from grayscale form to binary form using multi-level thresholding (Chaki *et al*., 2014).

1. Auto-cropping

For omitting extra image parts which are on both sides of mammogram image, a double precision image is gotten from the original image and multiply them in (unit 16) format, then using Otsu's threshold (Otsu, 1979) to get binary image, finally sweeping the first row of the image from the right corner to left and left to right to reach first white points (non-zero value). Then the image is cropped from beginning the corner till the non-zero values for both sides. [Figure (4-5)] demonstrate the procedure block diagram of auto-cropping mammogram image.

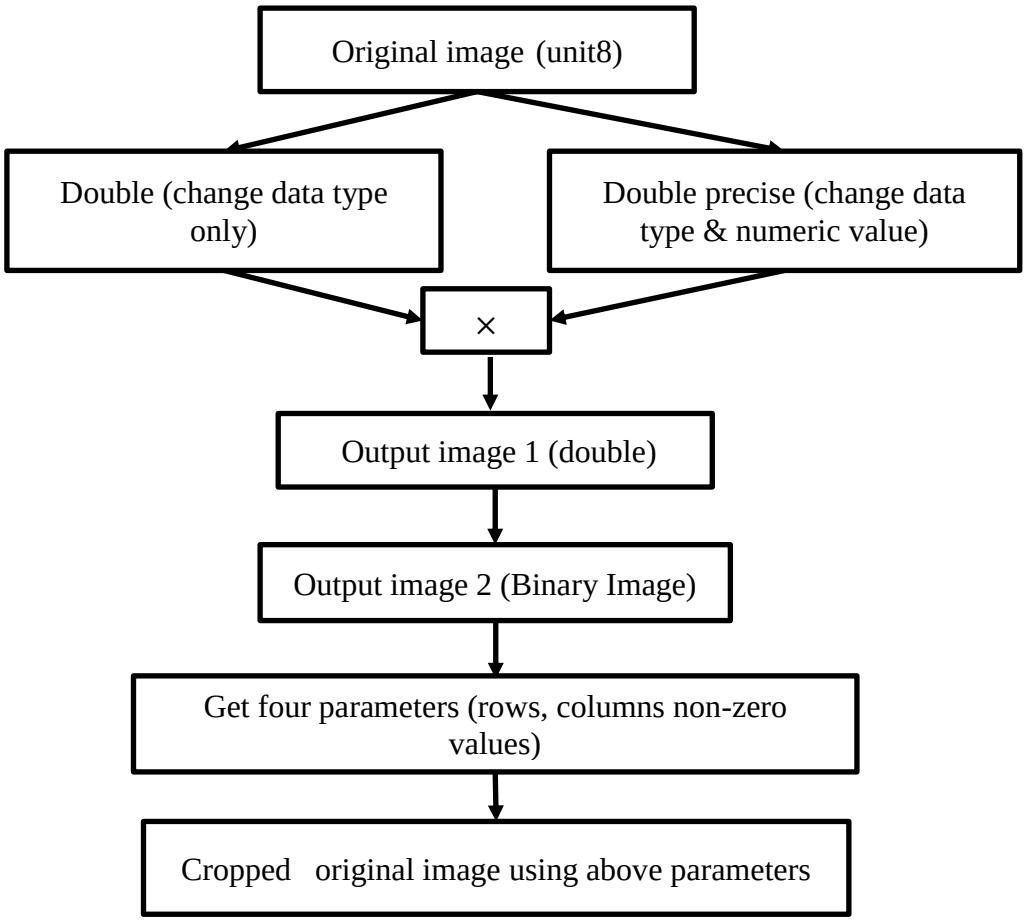


Figure (4-5): Block diagram of auto-cropping procedure

2. Label omitting

To omit the useless information as marker (small letters appear on the mammogram), a successive morphological operation are applied (Haralick and Shapiro, 1992) on the binary image; a disk type structuring element (Sreedhar and Panlal, 2012) is used here.

To omit labels, connected component technique (Yapa and Koichi, 2007) is used. For the binary image, the breast region identified as the largest connected component, and all other connected components which are small than breast region are discard. So, the final image is a breast region mask.

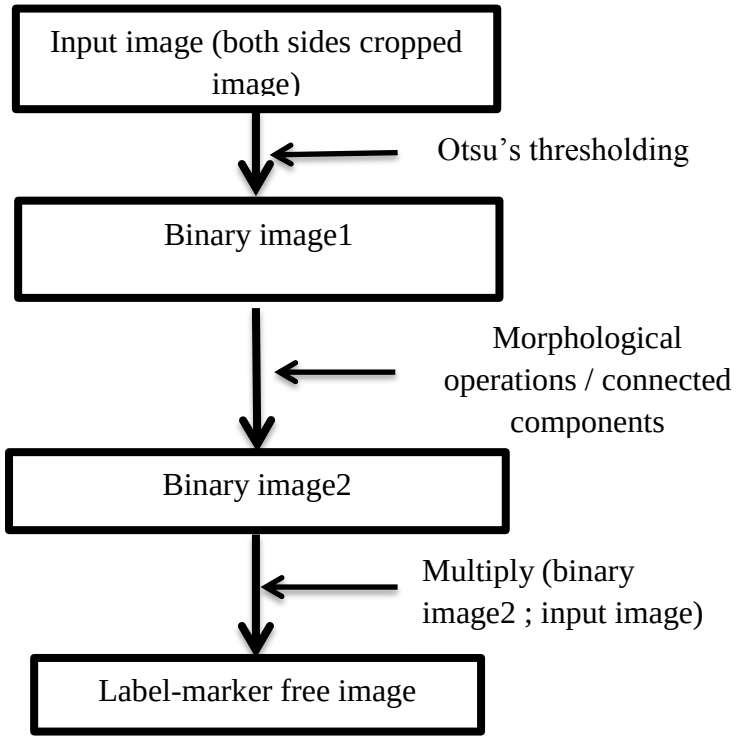


Figure (4-6): Block diagram of removing marker/labels procedure

4.2.2. Mammogram enhancement using median filter and CLAHE technique

There are different types of noises in MIAS database, the common noises present are salt and pepper noise and impulse noise. Contrast is the difference in visual properties that makes an object distinguishable from other objects and the background.

In this research, two different techniques have been used, to reduce the noise in the image and bring the enhanced and quality oriented data for good performance of CAD algorithm. The techniques are median filter (Huang *et al*., 1979) and CLAHE filter (Zuiderveld, 1994).

4.2.2.1. Median filter

Median filter run through the image pixel by pixel, replacing each value with the median value of neighboring pixels (mask or window). The pixels of the mask are ranked in the order of their gray levels, and the median value of the group is stored to replace the noisy value. The median filtering output is:

$$g(x,y) = med\{f(x-i,y-j) \quad ,i,j \in W \tag{4-1}$$

where : $f(x,y)$ is the original image, W is the two dimensional mask; the mask size is $(n \times n)$, n is commonly odd number

In this research median filter implemented assuming a mask size (3×3).

4.2.2.2. CLAHE technique

Contrast-limited adaptive histogram equalization CLAHE technique, is a special case of histogram equalization technique (HE) (Gonzalez, 1992), that functions adaptively on the image to be enhanced. The pixel intensity is thus transformed to a value within the display range proportional to the pixel intensity's rank in the local intensity histogram. The CLAHE operates in small regions in the image called tiles rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the distribution parameter. The CLAHE filter originally developed for medical imaging to reduce the noise and edge shadowing effect produced in homogeneous area. The CLAHE technique is described by following steps.

Step 1: Mammogram is divided into a number of non-overlapping contextual regions of equal sizes (tiles numbers). Experimentally set to 8×8

Step 2: The histogram of each contextual region is calculated.

Step 3: Setting clip limit$[0, 1]$, for clipping histograms. The clip limit is a threshold parameter, higher setting of clip limit result in more contrast. Experimentally is set$(t = 0.01)$

Step 4: Histogram redistribution, to not exceed the clip limit. In our experiment, uniform histogram distribution is selected.

Step 5: Histogram modification, according to transformation function:

$$t(r_k) = \sum_{j=0}^{k} p_r(r_j) \tag{4-2}$$

where

$$p_r(r_j) = \frac{n_j}{n} \tag{4-3}$$

is the probability density function of the input mammogram image grayscale value j, n is the total number of pixels in the input mammogram image and (nj) is the input number of grayscale value j

Step 6: The neighboring tiles are combined using bilinear interpolation and the mammogram image grayscale values are altered according to the modified histogram.

4.2.3. Pectoral muscle removal based on Otsu's threshold

It is important to determine the orientation of the mammogram. The mammogram image is transformed so get unidirectional images and the pectoral muscles in upper left corner of the image. In this work, images contain right breast are simply flipped to the left. The flipping criterion is based on MIAS database images arrangement and text file.

The pectoral muscles have a slightly higher intensity compared to rest of the breast tissue and appear in upper left corner of mammogram. So, it is important to detect the pectoral muscle and isolate it from the region of interest. In this research, segmentation based on thresholding approach is used. Mammogram image segmented using multi-level threshold technique which based on Otsu method.

4.2.3.1. Multiple thresholding

The multilevel threshold segments the pixels into several distinct groups in which the pixels of the same group have gray levels within a specific range (4-4). The Otsu's method can be applied for multiple thresholding segmentation (Bindu and Prasad, 2012), (Arora *et al*., 2008), the optimal thresholds t_1^* and t_2^* can be computed as equation (4-5).

$$g(x,y) = \begin{cases} a, & if\ f(x,y) > T_2 \\ b, & if\ T_1 < f(x,y) \leq\ T_2 \\ c, & if\ f(x,y) \leq T_1 \end{cases} \quad (4\text{-}4)$$

$$\sigma_B^2(t_1^*, t_2^*) = \max_{0<t_1<t_2<\mathrm{L}-1} \sigma_B^2\,(t_1, t_2) \quad (4\text{-}5)$$

Experimentally,$level$ (3)selected for segmentation. Using connected component labeling; the pectoral muscle identifies then removed according to its edge shape.

For non-straight line edge shape, pectoral muscle directly cropped using connected labeling. For straight line edge shape, a straight line refinement is applied by starting from upper left margin point and search horizontally and vertically to points where the pixel brightness change, then a triangle shape connected. Pectoral muscle lies within this triangle. Finally the demarked rectangle is cropped out from the original mammogram.

4.3. Segmentation of breast regions in mammogram

In CAD system, mass segmentation is an important step, which separate a mass from its background and captures the shape and boundary of the mass. However, it is more difficult to detect masses than micro-calcifications because their features can be obscured by normal breast parenchyma. Moreover, masses have large variability in morphology (shape and size), also, there is large number of features that have been used to detect and classify masses (Sickles, 2007).

Breast masses develop from the epithelial and connective tissues of breast, mass is a localized swelling, protuberance, or lump in the breast, and often occurred in the dense area of the breast tissue, have smoother boundaries than micro-calcifications and have many shapes. The shape and margin of mass are two most important criteria in distinguishing benign from malignant mass. Radiologist, found that a mass with poorly defined shape is

more likely to be malignant than a well-circumscribed mass; and a mass with ill-defined margins or spiculated lesion is much more likely to be malignant than a mass with smoothed margins (Cheng *et al*., 2006).

4.3.1. Mass--segmentation dataset

Masses are include circumscribed and spiculated masses. According to the MIAS database, 42 of the 322 mammogram images contain masses, as [Table (4-2)]. In this work, 41 mammograms are analyzed and one mammogram image (mdb059) is ignored because it is not including ROI location identification in database annotation.

Table (4-2): Mass distribution in MIAS

		Benign			Malignant			Total
		Fatty	Gland.	Dense	Fatty	Gland.	Dense	
Masses	Circumscribed	10	6	3	2	2	N/A	23
	Spiculated	2	4	5	3	3	2	19
	Total	12	10	8	5	5	2	42

In this work, two techniques of mass segmentation are proposed, as in flowing flowchart, Figure [(4-7)]. The first method based on threshold, while the second based on pseudo-color.

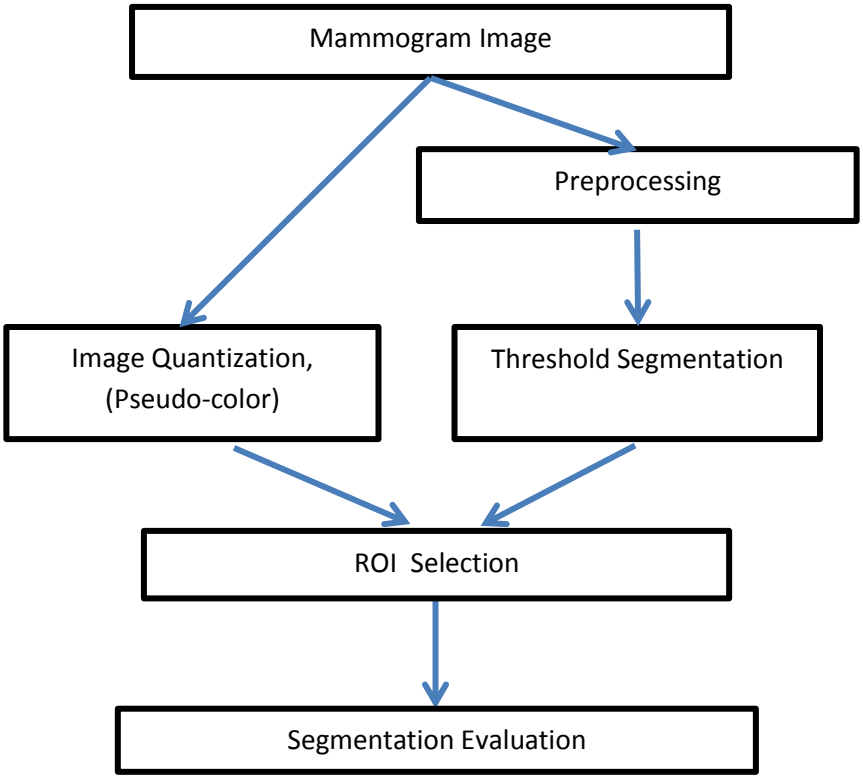


Figure (4-7): Mass segmentation strategy flowchart

4.3.2. Threshold-based segmentation technique

For image segmentation, a threshold based segmentation method is used, it considered as a special case of partition clustering methods. The main objective of this method is to classify the pixels of the image into two classes: those pertaining to an object and other pertaining to the background. For mammogram mass segmentation, the main drawback of this approach is the assumption that masses have more or less uniform density compared to the local background.

Here, intensity histogram based segmentation approach (Otsu) is used; Otsu segmentation algorithm aims to find split points on intensity histogram which separates whole intensity histogram into groups whose intra-class variances minimum. Thresholding, partition the mammogram into several regions, the suspicious area is an area that is brighter than its surrounding, which has almost uniform density with varying size.

In this experiment, multi-level threshold technique applied; the threshold is set to $level\ 4$. The process separates objects within ROI, which eroded and dilated by morphological operators. According to level threshold, mammogram may include unique ROI or more than

one of ROI. In case of unique threshold, the method considers as full-automated segmentation, while in case of multi ROI segmentation the method considers as semi-automated technique. For semi-automated, a set of regions of the mammograms are segmented as suspicious ROI, radiologist could select the mass according to morphological base of circumscribed or spiculated lesion.

4.3.3. Image quantization based segmentation technique

Quantization, in image processing is a lossy compression technique achieved by compressing a range of values to single quantum value. Quantization in terms of color histograms refers to the process of dropping the number of bins by taking colors that are very similar to each other and putting them in the same bin.

Pseudo color (false color) image processing consists of assigning colors to gray value based on a specified criterion. The principal cause of using color, that human can discern thousands of color shades and intensities compared to shades of gray (Zahedi *et al*., 2011). Preliminary study (Xu *et al*., 2010) indicates that lesion areas on the processed pseudo-color images could be more easily distinguished than on the original gray images. There many methods to realize pseudo-color of gray images such as the filtering in frequency domain, the equal density pseudo-color coding methods which include density segmentation coding, function transformation and complementary pseudo-color coding (Hu *et al*., 2012).

Function transformation method is shown in Figure [(4-8)]. The idea of this approach is to perform three independent transformations on the gray level of any input pixel. The three results are then fed separately into the red, green, and blue channels of a color.

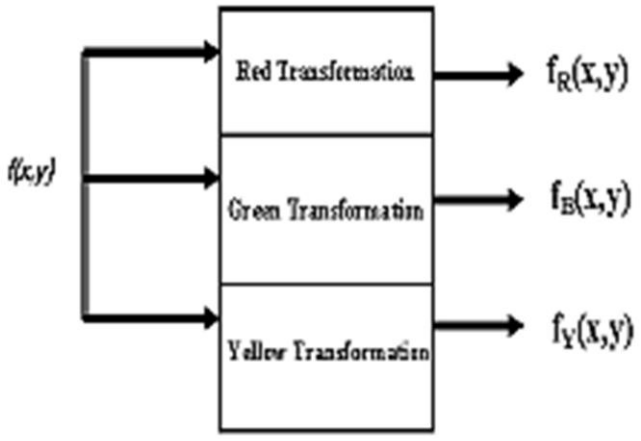


Figure (4-8): Gray-level to color transformation

In this experiment, the mammogram image (grayscale image) is quantized using$level7$; the output image then contained $(7+1)$ discrete integer pseudo-color. For mammogram images, highlighting a specific range of colors in a mammogram is useful for separating objects from their surroundings; however masses intensity is higher, so the color will differ. In pseudo-color segmentation, the operator could select rectangle window for the suspicious region (ROI) as a mask for further processing.

4.4. Texture analysis

Texture is a surface property; texture analysis refers to the branch of imaging science that is concerned with the description of characteristic image properties by textural features.

The main image processing disciplines in which texture analysis techniques are used are classification, segmentation, and synthesis. In image classification the goal is to classify different image region into distinct groups (Pietikainen, 2000). Texture analysis is a useful computational method for discriminating between pathologically different regions on mammogram.

The feature is defined as a function of one or more measurement which specifies some characteristic of an object. Feature extraction, is the transforming the input data into a set of features, it is a valuable step for mammogram classification, this step is responsible for extracting all possible features that are expected to be effective in diagnosing an ROI in the mammogram, without concerning the disadvantages of excessive dimensionality.

Experimentally, two types of textural features are extracted; the first one is based on statistical approach while the second is based on structural approach. Features are: Haralick features which are old, well known features and on SFTA features which are new established features. Both of features are introduced in following paragraphs.

4.4.1. Haralick Features

Haralick (Haralick *et al*., 1973) developed a set of fourteen scalar textural features driven from (GLCM), Haralick features are widely employed texture features which describes the correlation in intensity of pixels that are next to each other in space.

For Haralick features, Assume G is the number of gray level used, μ is the mean value of P. μ_x ,μ_y, $\sigma_x$and $\sigma_y$are the means and standard deviations of P_x and P_y. $P_x(i)$ is the ith entry in the marginal-probability matrix obtained by summing the rows of $P(i,j)$:

$$P_x(i) = \sum_{j=0}^{G-1} P(i,j)\,,\ P_y(j) = \sum_{i=0}^{G-1} P(i,j) \tag{4-6}$$

$$\mu_x = \sum_{i=0}^{G-1} iP_x(i)\,,\ \mu_y = \sum_{j=0}^{G-1} jP_y(j) \tag{4-7}$$

$$\sigma_x^2 = \sum_{i=0}^{G-1}(P_x(i) - \mu_x(i))^2\,,\ \sigma_y^2 = \sum_{j=0}^{G-1}(P_y(j) - \mu_y(j))^2 \tag{4-8}$$

Then

$$P_{x+y}(k) = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P\,(i,j),\ i+j=k,\ for\ k = 0,1,\cdots,2(G-1) \tag{4-9}$$

$$P_{x-y}(k) = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P\,(i,j),\ |i-j|=k,\ for\ k = 0,1,\cdots,G-1 \tag{4-10}$$

1. **Homogeneity (Angular second moment)**: measure the local uniformity of the gray levels. When pixels are very similar, the ASM value will be large.

$$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\{P(i,j)\}^2 \tag{4-11}$$

2. **Contrast:** is a measure of gray level variations between the reference pixel and its neighbor.

$$\sum_{n=0}^{G-1} n^2\{\sum_{i=1}^{G}\sum_{j=1}^{G} P(i,j)\} \tag{4-12}$$

3. **Correlation**: is the linear dependency of gray values in the co-occurrence matrix

$$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{(ij)P(i,j)-\{\mu_x\mu_y\}}{\sigma_x\sigma_y} \tag{4-13}$$

4. **Sum average:**

$$\sum_{i=0}^{2G-2} iP_{x+y}(i) \tag{4-14}$$

5. **Variance:**

$$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-\mu)^2P(i,j) \tag{4-15}$$

6. **Sum variance**:

$$\sum_{i=0}^{2G-2} i^2\,P_{x+y}(i) \tag{4-16}$$

7. **Difference variance:**

$$\sum_{i=0}^{G-1} i^2\,P_{x-y}(i) \tag{4-17}$$

8. **Entropy:** is the randomness or the degree of disorder present in the image. The value of entropy is the largest when all elements of the co-occurrence matrix are the same and small when element are unequal:

$$-\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}P(i,j)\log\big(P(i,j)\big) \tag{4-18}$$

9. **Sum entropy**:

$$-\sum_{i=0}^{2G-2}P_{x+y}(i)\log\Big(P_{x+y}(i)\Big) \tag{4-19}$$

10. Difference entropy:

$$-\sum_{i=0}^{G-1} P_{x-y}(i)\log\left(P_{x-y}(i)\right) \tag{4-20}$$

11. Inertia (Inverse difference moment) :

$$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{1}{1+(i-j)^2}P(i,j) \tag{4-21}$$

12. Info. measure of correlation 1:

$$\frac{HXY-HXY1}{max\{HX,HY\}} \tag{4-22}$$

where: $HXY=-\sum_i\sum_j p(i,j)log(p(i,j))$, HX, HY are the entropies of $p_x$and p_y, $HXY1=-\sum_i\sum_j p(i,j)\log\{p_x(i)p_y(j)\}$

13. Info. measure of correlation2 :

$$(1-\exp[-2(HXY2-HXY)])^{\frac{1}{2}} \tag{4-23}$$

Where: $HXY2=\sum_i\sum_j p_x(i)p_y(j)\log\{p_x(i)p_y(j)\}$

14. Max. correlation coefficient:

Is square root of the second largest eigenvalue of Q, where:

$$Q(i,j)=\sum_k\frac{p(i,k)p(j,k)}{p_x(i)p_y(k)} \tag{4-24}$$

4.4.2. Segmentation-based fractal texture analysis (SFTA)

A group of researchers (Costa *et al*., 2012) proposed a new feature extraction algorithm, the Segmentation-based Fractal Texture Analysis (SFTA). SFTA introduced for both medical imaging and general domain texture feature extraction. They applied SFTA for three different dataset, ROIs of lung CT scan dataset which used in (Costa *et al*., 2011), and two publicly available dataset: KTH-TIPS dataset (Caputo *et al*., 2005) and Textured Surfaces Dataset (Lazebnik *et al*., 2005). In Costa's research, SFTA showed superior performance when

compared with widely employed texture extraction methods such as Haralick features and Gabor filter banks.

SFTA extraction algorithm consists of two parts: firstly, decomposing the input grayscale image into a set of binary images, secondly, for each resulting binary image, compute three parameters (region's boundaries fractal dimension, region's mean gray level, and region's size).

1. Decomposition grayscale image

Decomposing original grayscale image into a set of binary images based on Two-Threshold Binary Decomposition (TTBD) technique (Costa *et al*., 2012). The TTBD compute a set T of threshold values, which obtained by selecting equally spaced gray level values. This achieved by using multi-level Otsu threshold. Then, TTBD algorithm decomposing the input grayscale image $I(x, y)$selecting pairs of threshold from T and applying two-threshold segmentation according to following equation:

$$I_b(x,y) = \begin{cases} 1 \; if \; t_e < I(x,y) \leq t_u \\ 0, \quad elsewhere \end{cases} \tag{4-25}$$

where I_b denotes the binary image, $t_e$and t_u denote lower and upper threshold values respectively.

According to applying the two-threshold segmentation to the input image using all pairs of contiguous threshold from $T \cup \{n_l\}$ and all pairs of thresholds$\{t, n_l\}, t \in T$, where n_l corresponds to the maximum possible gray level in $I(x, y)$. The number of resulting binary image is $2n_t$,where n_t is a user defined parameter that corresponds to the number of threshold values. The set of binary images is obtained by applying the two threshold segmentation

2. SFTA features extraction

After applying TTBD, for the resultant binary image the SFTA algorithm computes feature vectors which are region area (size), mean gray level, and boundaries' fractal dimension, The mean gray level and size (pixel count) compute directly from binary image.

The region boundaries of a binary image $I_b(x,y)$ are represented as a border image denoted by $\Delta(x,y)$and computed as follows:

$$\Delta(x,y) = \begin{cases} 1 \; if \; \exists (x',y') \in N_8[(x,y)]: \\ \quad I_b(x',y') = 0 \;\; \wedge \\ \quad I_b(x,y) = 1, \\ 0, \quad otherwise \end{cases} \tag{4-26}$$

where $N_8[(x,y)]$ is the set of pixels that are 8-connected to (x,y).

The fractal dimension D computed from each border using the box counting algorithm (Schroeder, 1992). The algorithm, divide the image into a grid composed of squares of size $\epsilon \times \epsilon$, then, counting the number $\overline{N}(\epsilon)$ of squares of size $\epsilon \times \epsilon$ that contains at least one pixel of the object. Fractal dimension D corresponds to the slope of line, which is fitted on $\log \overline{N}(\epsilon)$ vs $\log \epsilon^{-1}$ curve.

Algorithm 1 shows SFTA extraction algorithm. The SFTA feature vector V_{SFTA} dimensionally corresponds to the number of binary images obtained by TTBD multiplied by three.

Algorithm 1: SFTA Extraction algorithm [Costa *et al*., 2012]

Require: Grayscale image I and number of thresholds n_t.
Ensure: Feature vector V_{SFTA}.
1: $T \leftarrow \textbf{MultiLevelOtsu}(I, n_t)$
2: $T_A \leftarrow \{\{t_i, t_{i+1}\} : t_i, t_{i+1} \in T, i \in [1..|T|-1]\}$
3: $T_B \leftarrow \{\{t_i, n_t\} : t_i \in T, i \in [1..|T|]\}$
4: $i \leftarrow 0$
5: **for** $\{\{t_\ell, t_u\} : \{t_\ell, t_u\} \in T_A \cup T_B\}$ **do**
6: $I_b \leftarrow \textbf{TwoThresholdSegmentation}(I, t_\ell, t_u)$
7: $\Delta(x,y) \leftarrow \textbf{FindBorders}(I_b)$
8: $V_{\text{SFTA}}[i] \leftarrow \textbf{BoxCounting}(\Delta)$
9: $V_{\text{SFTA}}[i+1] \leftarrow \textbf{MeanGrayLevel}(I, I_b)$
10: $V_{\text{SFTA}}[i+2] \leftarrow \textbf{PixelCount}(I_b)$
11: $i \leftarrow i+3$
12: **end for**
13: **return** V_{SFTA}

4.5. Feature selection

It is very difficult to predict which feature or feature combination will achieve better in classification rate. Generally, when the number of features is large but the number of training sample is small, this yield to a situation called the curse of dimensionality (Aoki and Kudo, 2008), this situation will weak the classifier. Moreover, few features used in classifier can keep the classification performance robust (Giger *et al*., 2000). So, an optimized subset of features must selected to gain the highest performance.

Removing irrelevant features may lead to improved accuracy and increased interpretability of the classification model. Features can be selected in many different ways. One scheme is to select features that correlate strongest to the classification variable; this has been called maximum-relevance selection. On the other scheme, features can be selected to be mutually far away from each other while still having high correlation with the classification variable. This scheme termed as Minimum Redundancy Maximum Relevance (mRMR). mRMR selection method has been found more powerful than the maximum relevance selection method (Peng *et al*., 2005)

For feature selection (mRMR) selection method is used, which is an approximation to maximizing the dependency between the joint distribution of the selected features and classification variable.

Haralick features and SFTA features are extracted, and then feature selection method applied for individual features and for both Haralick and SFTA features. Selecting seven optimize subset of features from Haralick features, SFTA features, and Haralick-SFTA combination features. Difference performances are obtained as a result of different features.

mRMR algorithm

The mRMR is a feature selection approach that tends to select features with a high correlation with the class and a low correlation between each other's. In this algorithm, the features are ranked according to the minimal-redundancy-maximal-relevance criteria (Peng *et al*., 2005). In case of continuous features, relevance can be calculated by using the F-statistic, while redundancy calculated by using Pearson correlation coefficient. For discrete features, relevance and redundancy calculated by using mutual information [Figure (4-9)].

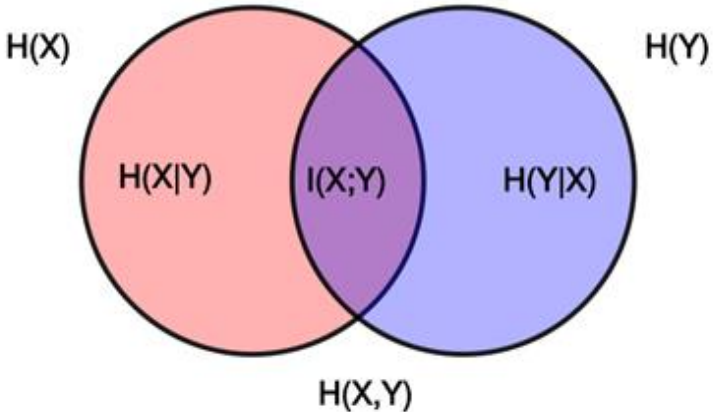


Figure (4-9): Relevance between features, mutual information

The mutual information defines dependency of variables. Assume individual are (H(X), H(Y)), joint (H(X,Y)), and conditional entropy for a pair of correlated subsystems X, Y with mutual information I(X;Y):

$$I(x,y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)}\, dxdy \tag{4-27}$$

where: p(x, y): joint distribution function of X and Y

p(x), p(y): marginal probability distribution functions.

Maximize Relevance:

$$\max V_I = \frac{1}{|S|} \sum_{i \in S} I(h,i) \tag{4-28}$$

where: S is the set of features; $I(i,j)$ is mutual information between feature i and j

Minimal Redundancy:

$$\min W_I = \frac{1}{|S|^2} \sum_{i,j \in S} I(i,j) \tag{4-29}$$

4.6. Classifications

Classification is the process of taking decision that best matches the membership of the object. The task is composed of many stages; the goal is to associate the appropriate class labels with the test image.

Support Vector Machine

The support vector machine (SVM) is a classification method introduced in 1992 (Boser *et al*, 1992). It is based on the principle of structural risk minimization, which targets at minimizing the error made by the learning machine during training rather than minimizing the mean square error (Vapnik, 1998). In the field of medical imaging, SVM has proven to be a good classifier in the diagnosis of mammograms (Balakumaran and Vennila, 2011). SVM is a supervised learning algorithm based on the concept of hyperplane to separate a set of objects with maximum margin (Cortes and Vapnik, 1995). SVM works to give optimal hyperplane equation, as following

$$f(x) = \sum_{i=1}^{N} \alpha_i \, w_i (x_i \cdot x) + b \qquad (4\text{-}30)$$

where $X\{x_1, x_2, \dots, x_N\}$ is training set containing Nfeature vectors, w is known as the weight vector, α_i is the Lagrange multiplier of a dual optimization problem that describes the separating hyperplane $w_i(x_i . x)$, and b is the threshold parameter of the hyperplane.

4.7. Tissue characterization

The sensitivity of CAD in the detection of breast cancer is impacted by breast density. In mammography, breast tissue presents differently: darker regions indicate fat and clearer regions indicate glandular, thus characterizing the breast tissue type in mammograms can act as a primary step to detect cancer and reduce false positive. Breast density is a way to describe the types of tissue that make up the breast. The amount of each of these tissues varies in women. The dense breast more likely to develop a cancer (Ursin *et al*., 2005) so breast tissue density type consider as indicator for cancer risk, moreover the detection of breast

cancer in dense breast are harder than detection cancer that surrounded by fatty tissue (Diamant *et al*., 2012). According to BI-RADS category, there is only a minimal and insignificant difference in the sensitivity of mammography between the densest breast in a lower density category and the least dense breast in the next higher density category [ACR, 2013].

Breast density is a radiological concept based on the proportion of radiopaque glandular tissue relative to radiolucent fatty tissue. On other words, breast density is the ratio of the area of dense (white) tissue on a mammogram divided by the total area of the imaged breast. There is no criterion standard for determining breast density. However, this classification in most cases depends on the skills of the technician or doctor who performs the exam. Commonly used methods of breast density assessment range from subjective visual estimation to quantitative calculations of area and volume density made with complex computer algorithm (Winkler *et al*., 2015).

4.7.1. Tissue types dataset

MIAS database classifies breast density into three classes, that is, fatty (F), fatty glandular (G) and dense glandular (D), [Figure (4-10)]. By using Random Sample without Replacement (RSWR) method (Teuhola and Nevalainen, 1982), the dataset is divided according to tissue types into a learning dataset of 160 images and a testing dataset of 162. This experiment mainly restricted to cases, one case being the association of the left and of the right breast images of one patient. So there are 80 cases in the learning dataset and 81 cases in the test database. [Table (4-3)] presents the distribution learning and testing datasets according to their tissue type.

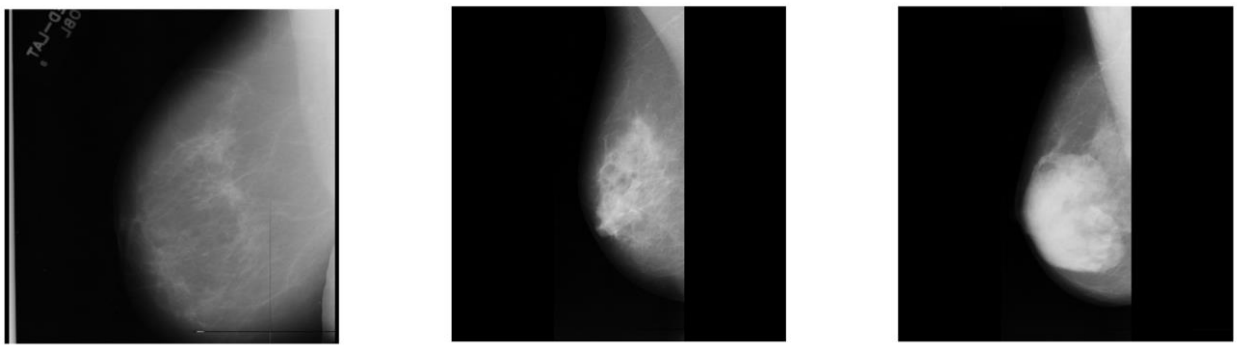


Figure (4-10): Three tissue types examples of MIAS database, fatty, glandular, and dense (from *left to right*)

Table (4-3): Tissue types distribution of selected dataset

Class	***Learning (cases)***	***Testing (cases)***	***Total (cases)***	***Total (images)***
Fatty	27	26	53	106
Glandular	25	27	52	104
Dense	28	28	56	112
Total	80	81	161	322

4.7.2. Feature extraction

Two types of texture features are computed, Haralick texture features and Segmentation-based Fractal Texture Analysis (SFTA) features.

For the thirteen-Haralick texture features, angle is set to 45 , displacement to 1, and 8 gray levels. For SFTA method, features are extracted using SFTA (I, nt) function, where I corresponds to input image and (nt) corresponds to the number of selected thresholds, features are returned as $(6 * nt - 3)$ vector. In this experiment, the number of thresholds is set to 4, the resulting set of binary images is $(2n_t - 1)$ as in [Figure (4-11)], and providing 21 SFTA features for each image denoted F1 till F21.

In this stage, for extracting features, three concurrent studies are investigated. The key of the three approaches is a window size. In order to obtain a good texture characterization, it is desirable to work with large windows, since they obviously contain more information.

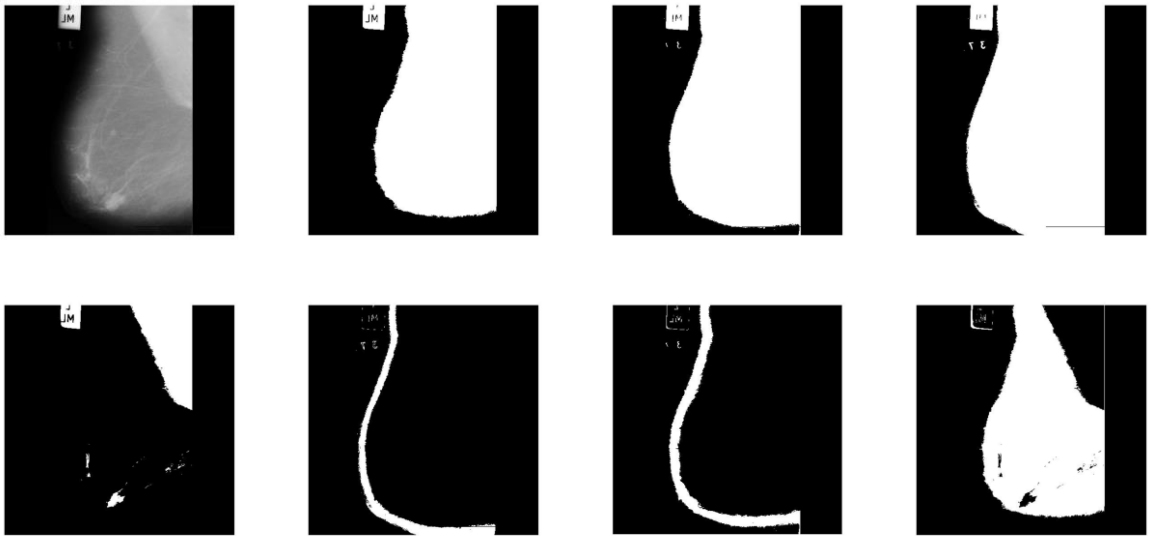


Figure (**4-11):** Example of original image and 7th resultant binary image from SFTA extraction algorithm, threshold set to 4, (from *left to right* and *top to bottom*)

4.7.2.1. The 1st study

1. Apply preprocessing stage to all mammogram images [Figure(4-12)]
2. Compute texture features for each image, and then compute average values of right and left breast (one value per case).

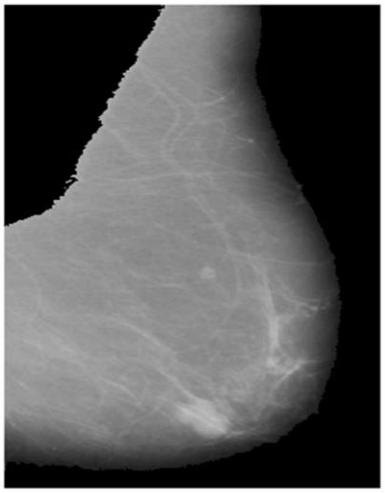


Figure (4-12): Breast region selection of the 1st study

4.7.2.2. The 2nd study

1. Apply CLAHE technique to mammogram image for contrast enhancement.
2. Select a center point (x, y) manually (in the center of the breast region), then a window of (50×50) pixels centered on the previously selected point is generated automatically.
3. Generate four windows of (50×50) pixels, at a distance of $150\ pixels$ from the center point, as shown in [Figure (4-13)].
4. Compute texture features for each window, compute average of any texture feature of the five windows per image, then compute the average of any texture feature of right and left breast (case).

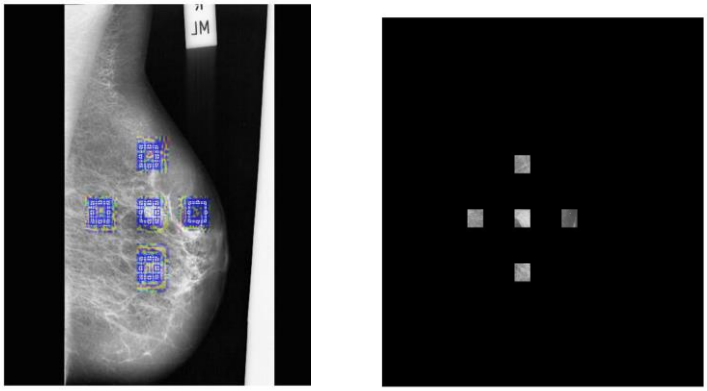


Figure (4-13): The 2nd study, five windows selection (blue windows)

4.7.2.3. The 3rd study

1. Apply CLAHE technique.
2. Select one window of (128 × 128) pixels manually, located approximately behind the nipple.
3. Compute texture features inside the window, then average values of right and left breast (one value per case).

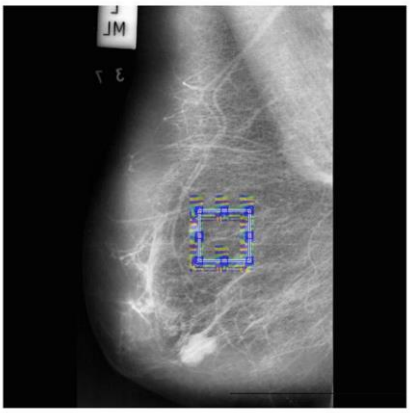


Figure (4-14): The 3rd study, one window selection (blue window)

4.7.3. Classification

For classification, Support Vector Machine (SVM) classifier is built on a linear kernel function. Different kernel functions are used and then obtained different classification

hyperplane, the best classification accuracy obtained by using polynomial function. Three classifications are conducted for each features extraction study.

1. **First classification study (C_1)**: The first study applied binary classification to differentiate between fatty tissue and non-fatty tissue, the non-fatty tissue consists of glandular and dense tissue types.
2. **Second classification study (C_2):** The second study used binary classification to classify glandular tissue and dense tissue
3. **Third classification study (C_3)**: The third study used multi classification process to distinguish between the three tissue types (Fatty, Gland and Dense)

For each study, mRMR selection method is applied to select best features from:

1) Haralick features only.
2) SFTA features only
3) Combined SFTA and Haralick features.

Features descriptors as shown in [Table (4-4)]

Table (4-4): Extracted features for tissue classification experiments

Method Name	Setting	Number of Features	Description
Haralick	angle: 45°, displacement: 1 gray levels: 8	14	H1,H2,…,H13, H14 (Homogeneity, Contrast, Correlation, Variance, Inertia, Sum-average, Sum-variance Entropy, Sum-entropy, Diff-variance, Diff-entropy, Info-correlation1, Info-correlation2, Max.corr Coeff) respectively
SFTA	Number of threshold: 4	21	F1, F2, ….F21
Combined Haralick & SFTA	angle: 45°, displacement: 1 gray levels: 8 ; Number of threshold: 4	35	Combination of Haralick and SFTA (H1,H2,…H14 and F1, F2,…F21)

4.8. Normal-abnormal classification

MIAS dataset includes normal and abnormal images as distributed in [Table (4-5)].

Table (4-5): The distribution of MIAS database classes

Class	Benign	Malignant	Total
Microcalcification	12	13	25
Circumscribed	19	4	23
Spiculated mass	11	8	19
Ill-defined mass	7	7	14
Architectural distortion	9	10	19
Asymmetry	6	9	15
Normal	0	0	207
Total	64	51	322

4.8.1. Dataset

MIAS database contains 207 normal mammogram images, and 115 abnormal images. For this experiment, the learning and testing data sets are randomly selected. The learning contains 108 images; the testing dataset contains 110 images as [Table (4-6)], datasets are selected regardless tissue types. CLAHE technique applied to all dataset.

Table (4-6): Normal-abnormal dataset distribution

	Learning	***Testing***	***Total***
Normal	56	59	115
Abnormal	52	51	103
Total	108	110	218

4.8.2. Feature extraction

Haralick and SFTA features are extracted from region of interest (ROI). For normal mammogram, a ROI is a rectangle box selected on the center of image, for abnormal mammogram, a ROI selected according to abnormality coordinates (x, y) given by MIAS information details,

Two experiments are applied:

1. Fixed window size (64×64): extracted Haralick features and extracted SFTA features at different threshold values $nt = \{2,3,4,5,6,7,8\}$, and test the classification
2. Different window sizes: extracted Haralick features and SFTA in different size of ROIs, set SFTA threshold at $(nt = 3)$, the different ROIs size are: $\{(8 \times 8), (16 \times 16), (32 \times 32), (40 \times 40), (50 \times 50), (80 \times 80), (100 \times 100), (128 \times 128)\}$

4.8.3. Classification

To classify normal and abnormal mammograms, mRMR selection method is used to select the reliable features from:

1) Haralick features only.
2) SFTA features only
3) Combined SFTA and Haralick features.

In this research, SVM classifier is built on a linear kernel function. Different kernel functions are tested, and then obtained different classification hyperplane, but the best classification accuracy obtained by using linear function

4.9. Benign-malignant classification

All the abnormal learning and testing datasets are selected according to the previous study (4.8.1.**)**. [Table (4-7)], shows the abnormal distribution.

Haralick and SFTA texture features are extracted from ROI, the selected window size is (32×32), and the threshold for SFTA is set to three. SVM classifier built using individual texture types and combined Haralick with SFTA features.

Table (4-7): Benign-Malignant dataset distribution

	Learning	***Testing***	***Total***
Benign	29	29	58
Malignant	23	22	45
Total	52	51	103

5. Results

5.1. Output of preprocessing stage

To limit the processing to breast region, we applied the preprocessing stage using three steps: the first step is omitting the excessive parts which are in the both sides of image, as in [Figure (5-1)]. [Figure (5-2)] shows morphological operations applied on mammogram image. So, the white region at binary image could shrink by erosion or expand by dilation operation. Example of omitting unwanted marker and label is in [Figure (5-3)] and [Figure (5-4)]. The second step is mammogram enhancement, by noise filtering as shown in [Figure (5-5)], [Figure (5-6)], and contrast enhancement [Figure (5-7)], and the last one is distinct the breast region and put all images in one direction for pectoral muscle removal.

In general, muscles have high intensity than tissue, so pectoral muscle have maximum threshold, and we obtained the perfect result by using level-three in multi-thresholding for pectoral muscle segmentation. Pectoral muscle edges have different curvatures (convex, straight line, and concave). So the straight line technique is efficient on straight line edges while directly cropped technique efficient on convex and concave edges [Figure (5-8)] [Figure (5-9)].

The algorithm showed the ability to eliminate background, remove pectoral muscle and enhance the image contrast without losing any information from the image. The results show reducing of image size and consequently, minimizing the computational time of processing stage. A set of samples of input and output image are shown in [Figure (5-10)]

A proposed method applied on 160 images and the obtained accuracy is (96%) for background and pectoral muscle elimination. Out of 160 images, 6 images was success in background elimination but failed to identify the pectoral muscle, because images are extremely dense, the density of pectoral muscle same as whole breast, so it’s difficult to segment the pectoral muscle as in [Figure (5-11)].

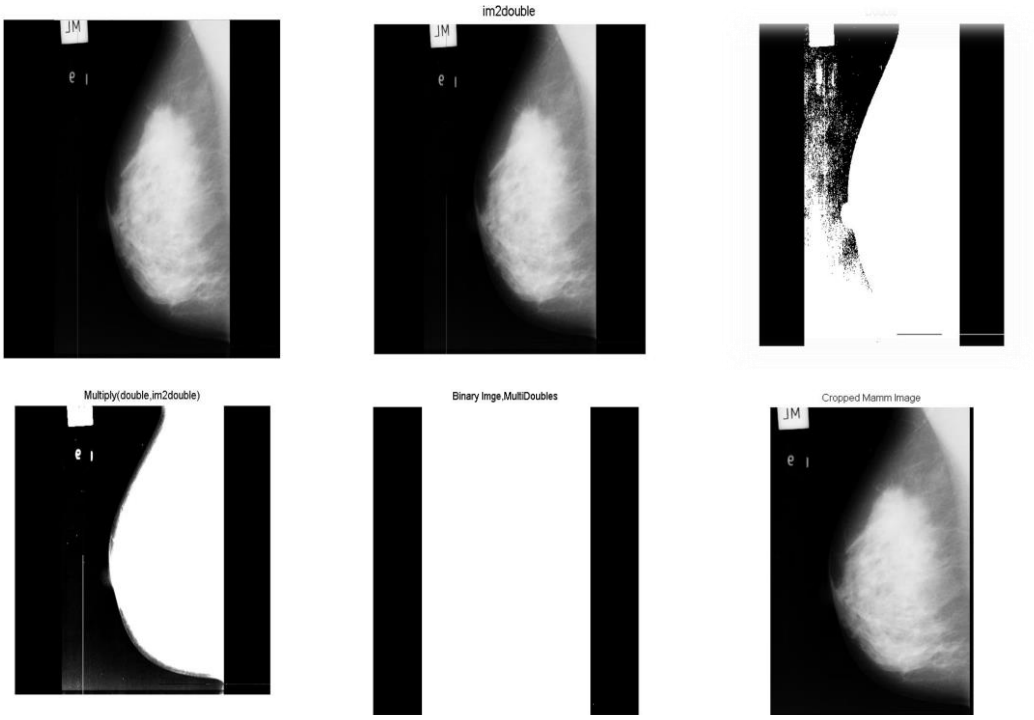


Figure (5-1): Auto-cropping result: original image (mdb003), double image, double precise image, double image result of double and double precise images multiplication, binary image, output cropped image

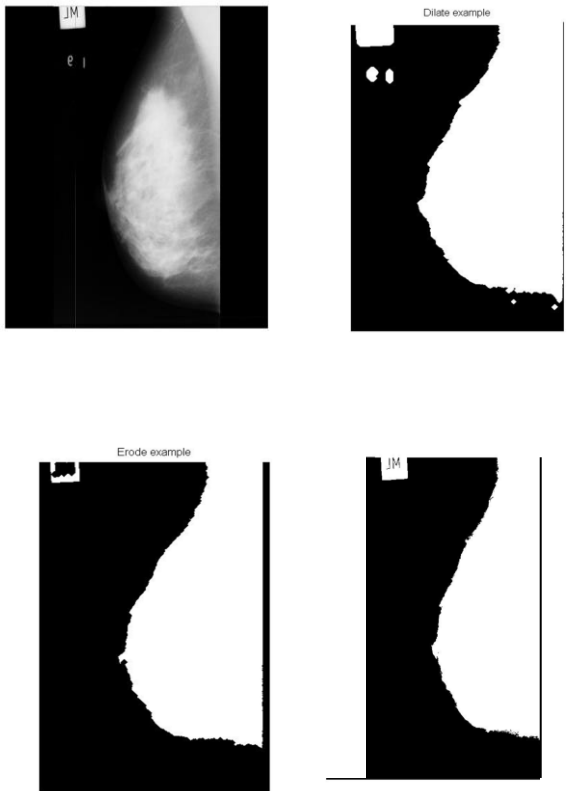


Figure (5-2): Morphological operations output: Original image, binary images after (erosion, dilation, opening) respectively

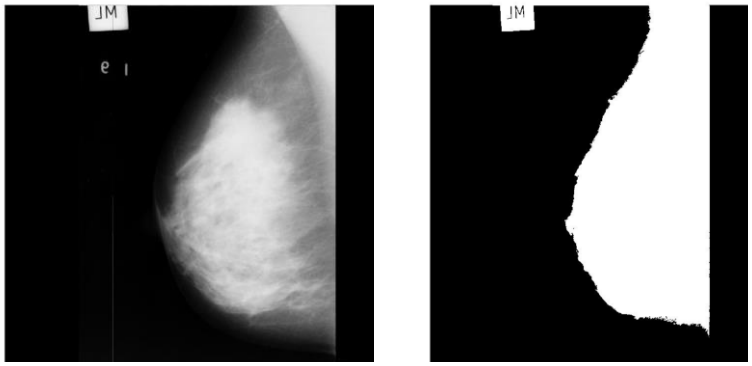


Figure (5-3): Marker removal: original image, binary image after apply morphological operation,

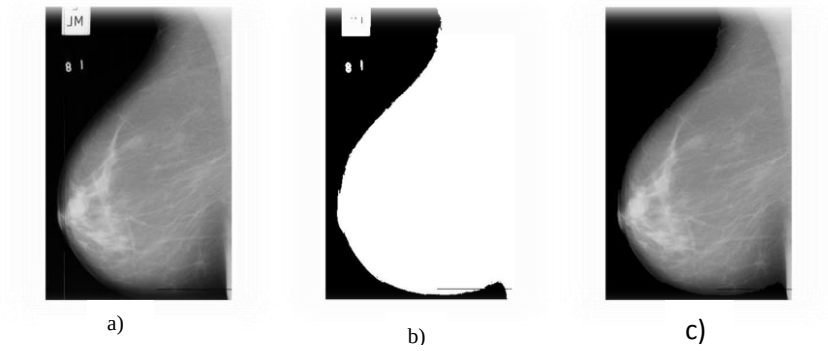


Figure (5-4): Label Omitting: (a) Cropped image, mdb195 (b) Binary image. (c) Label free image

Figure (5-5): Median filtering used for preprocessing: (from *left* to *right*) original image, binary image with noises, and binary image after filtering

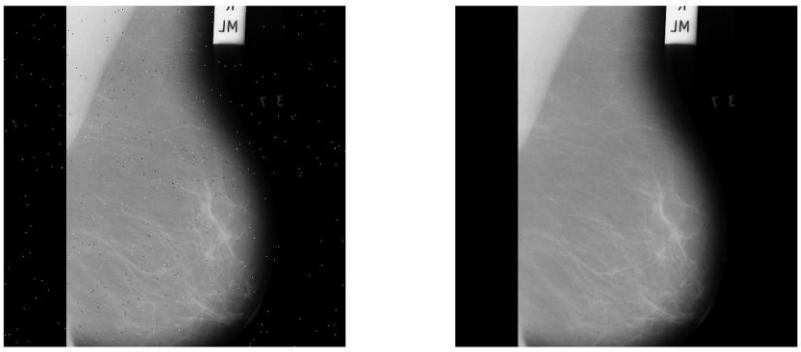


Figure (5-6): Noise filtering: noisy, and de-noised image
(from *left to right)*)

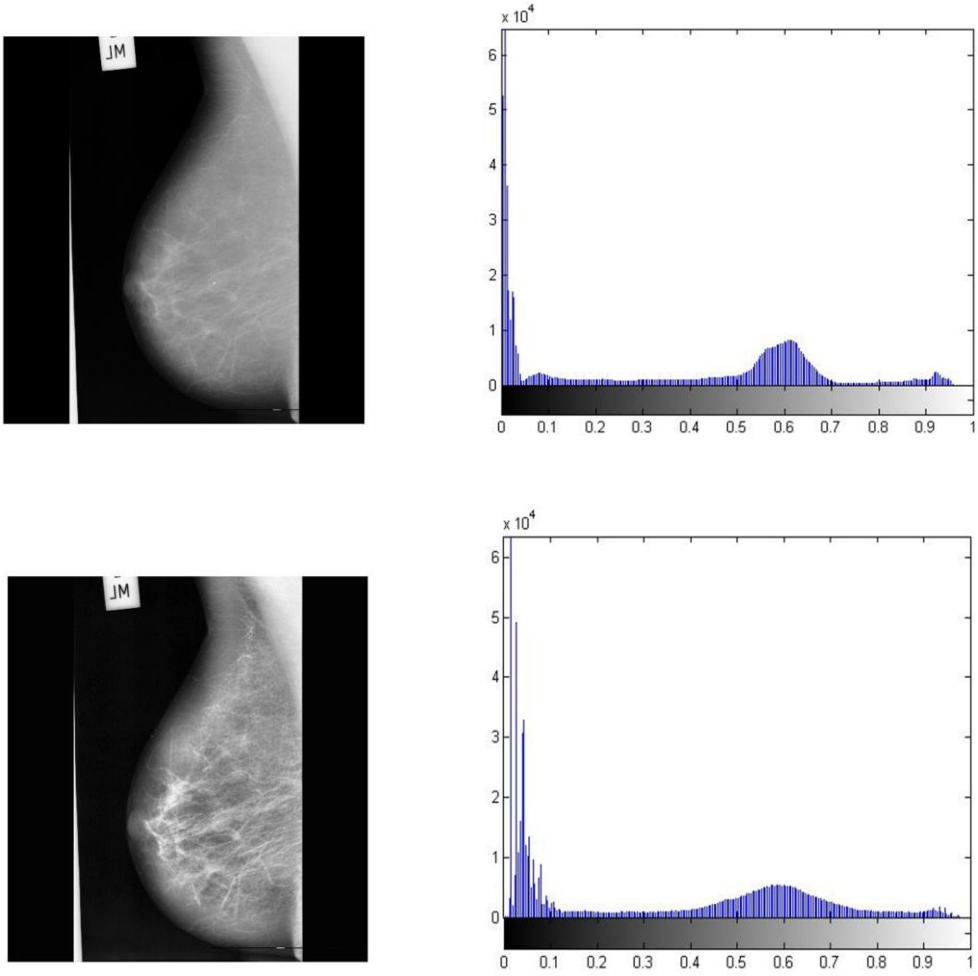


Figure (5-7): CLAHE technique: original image and histogram, enhance image and histogram (from *left to right* and *top* to *bottom*)

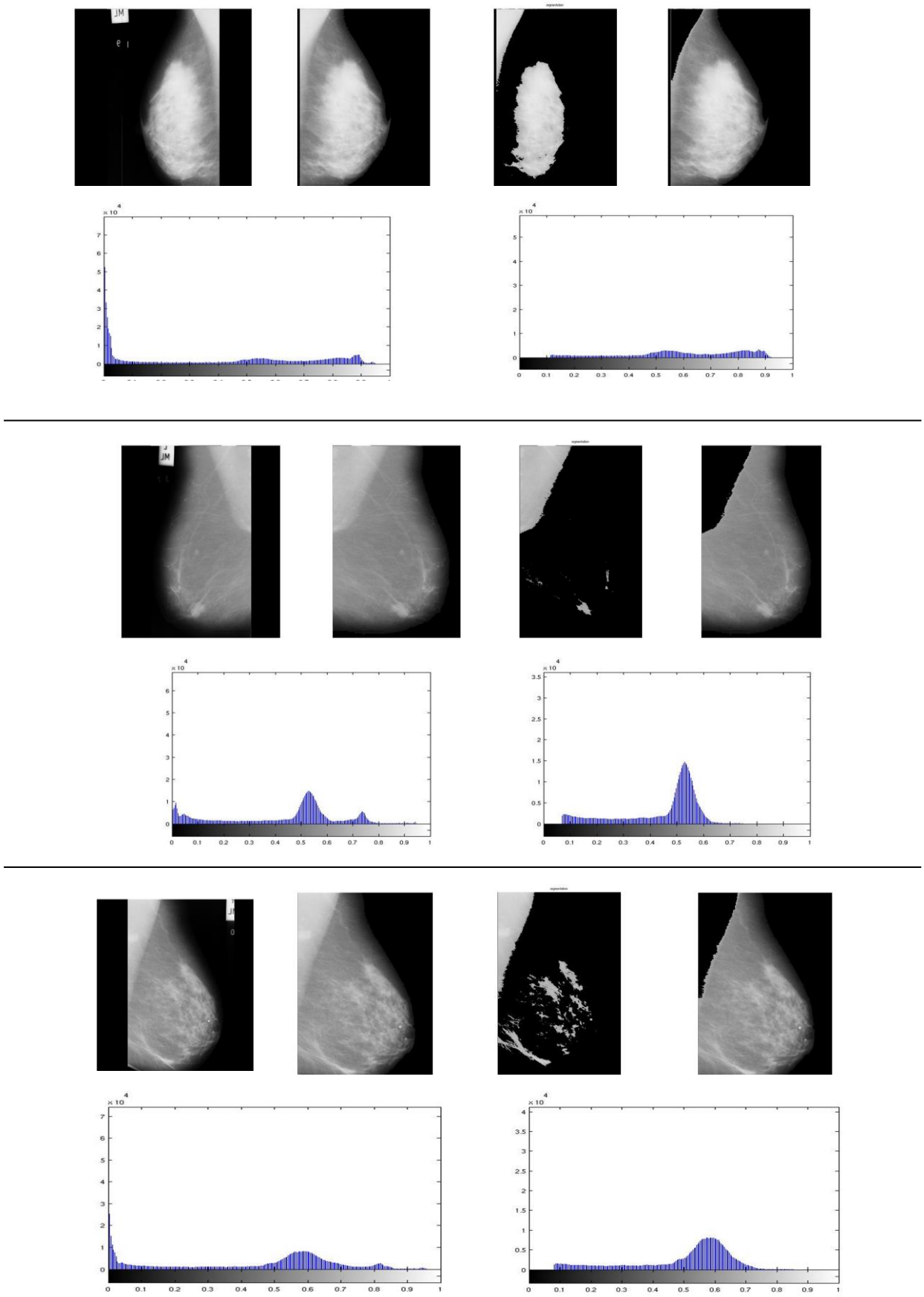


Figure (5-8): Pectoral muscle removal based on segmentation and connected components: original image and input-output histograms, images are 003, 005, 008

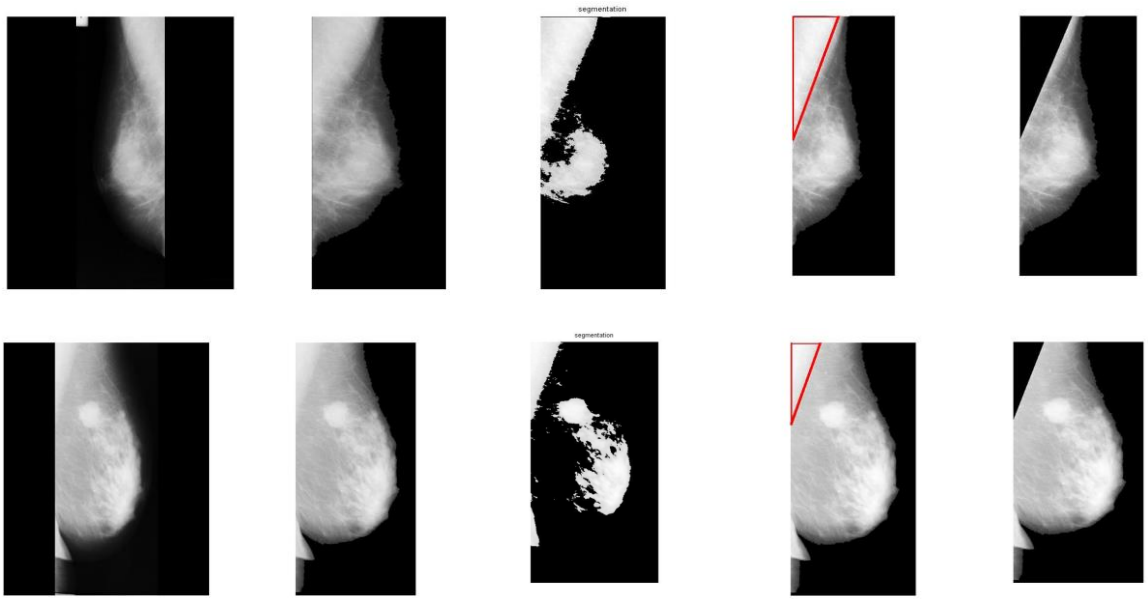


Figure (5-9): Pectoral muscle removal based on straight line refinement

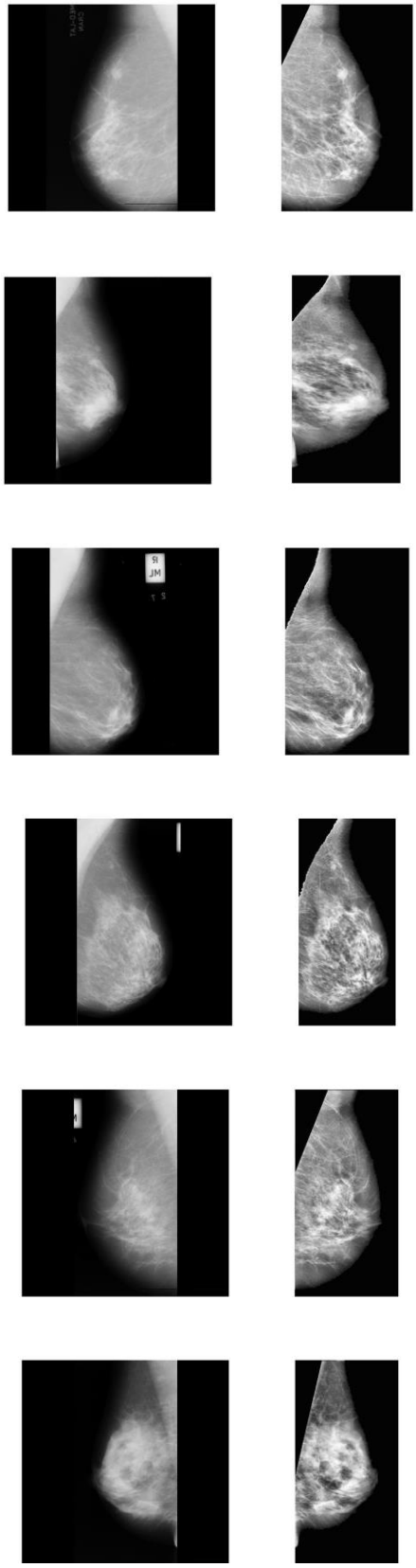


Figure (5-10): Examples of success preprocessing stage: input and output mammogram (from *right* to *left*), images are: mdb023, mdb072, mdb096, mdb100, mdb103, mdb225 (from *top* to *bottom*)

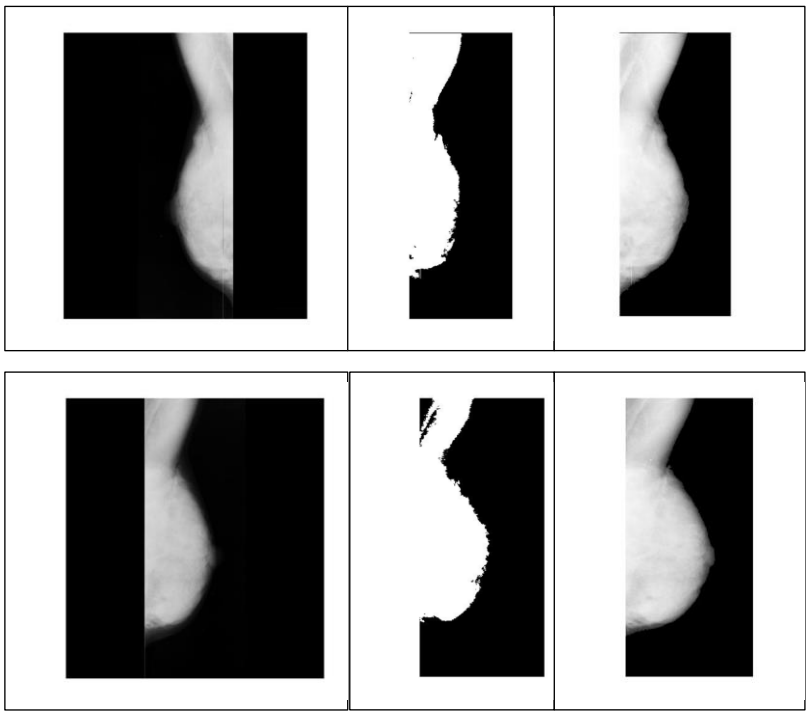


Figure (5-11): Example of images failed in preprocessing stage: image 053, image 054 (*top* to *bottom)* Input image, binary, and output image (from *left* to *right*)

5.2. Segmentation outcomes

Two types of segmentation are conducted, the breast region isolation and suspicious region isolation

5.2.1. Breast region segmentation

This process is applied for segment the breast region from the background. Results in previous stage (preprocessing) showed the breast region segmentation by background elimination and pectoral muscle removal using multilevel thresholding, $level - 3$ for grayscale image.

Other segmentation results obtained by convert the grayscale image to RGB image, and then applied quantization. Results showed the ability of quantization to segment the breast region and ROI as in [Figure (5-12)].

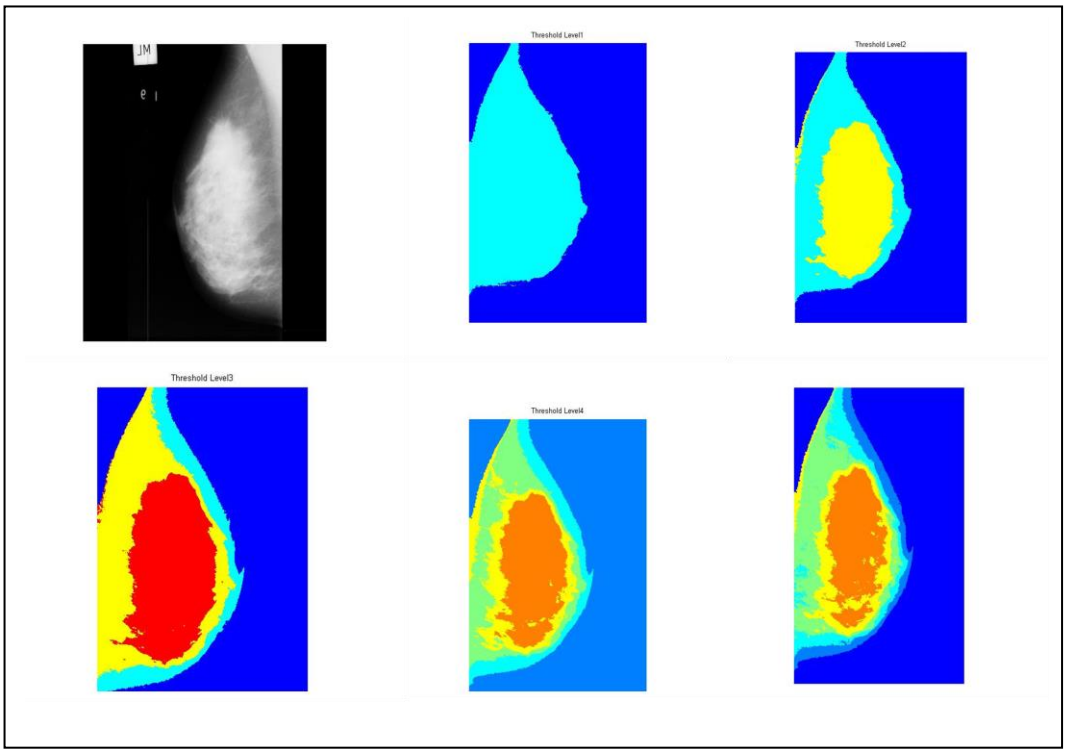


Figure (5-12): Breast region segmentation based RGB quantization

5.2.2. Mass lesion segmentation

In this research, the suspicious regions are masses which are circumscribed or spiculated. The proposed methods (Thresholding, Pseudo-color), tested on 41 abnormal images of MIAS database. The efficiency of the algorithm is measured by the following equation

$$Efficiency = \frac{Number\ of\ True\ identification\ of\ mass}{Total\ abnormal\ images} \quad (5\text{-}1)$$

$$Error = \frac{Number\ of\ False\ identification\ of\ mass}{Total\ abnormal\ images} \quad (5\text{-}2)$$

[Figure (5-13)] and [Figure (5-14)] are examples of mass segmentation based on thresholding. The method successfully identified 36 masses among 41 abnormal images. However, five lesions are not identified properly on the fourth level of Otsu's threshold because of their poor intensity. The efficiency of algorithm is reported as 87.8% and error rate is 12%.

Pseudo-color segmentation method [Figure (5-15)], [Figure (5-16)] obtained less accuracy than thresholding segmentation method, the method identified 32 masses. The efficiency is78%, and error rate is22%.

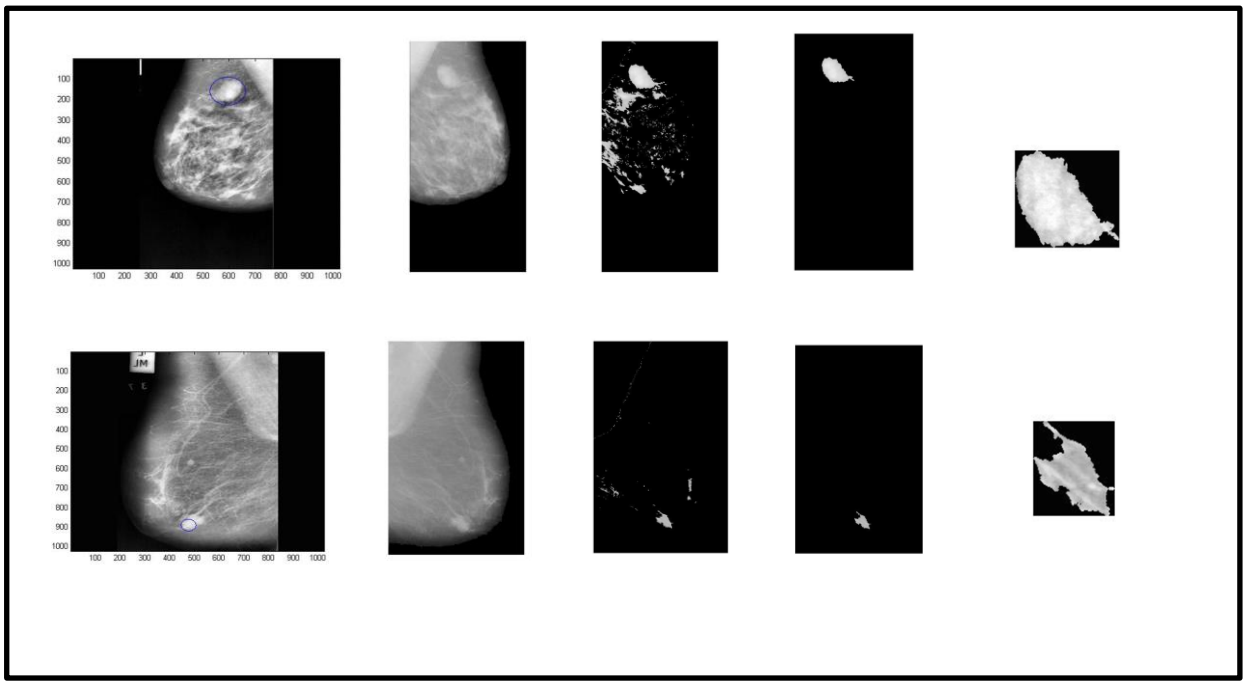


Figure (5-13): Examples of mass segmentation-based thresholding method

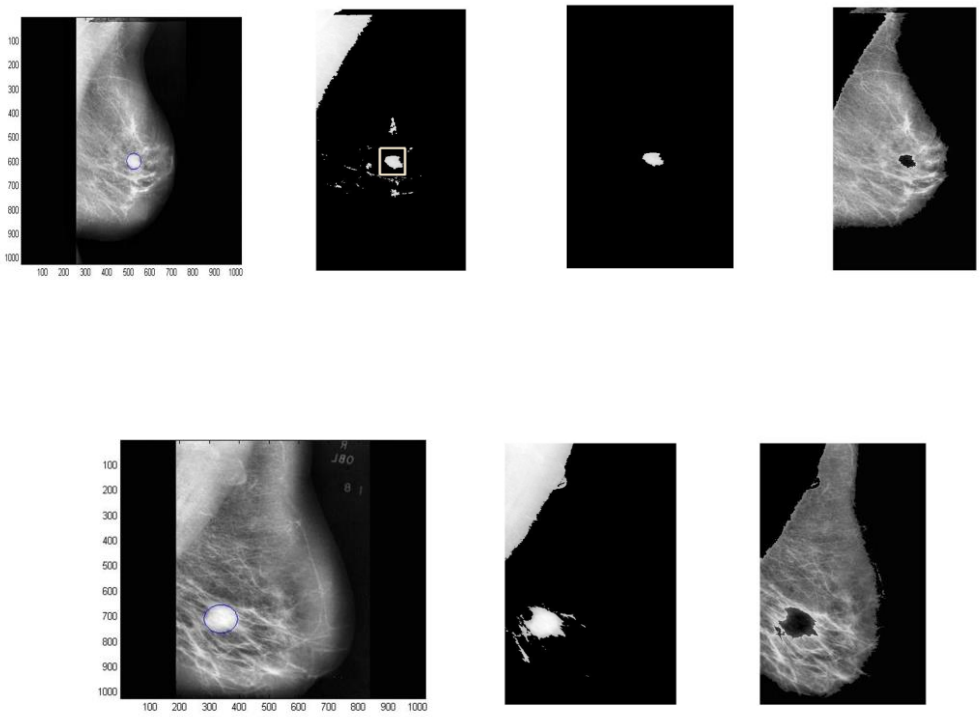


Figure (5-14): Automatic mass segmentation-based thresholding , several suspicious region, uni-suspicious region (from *right to left* and *top to bottom*)

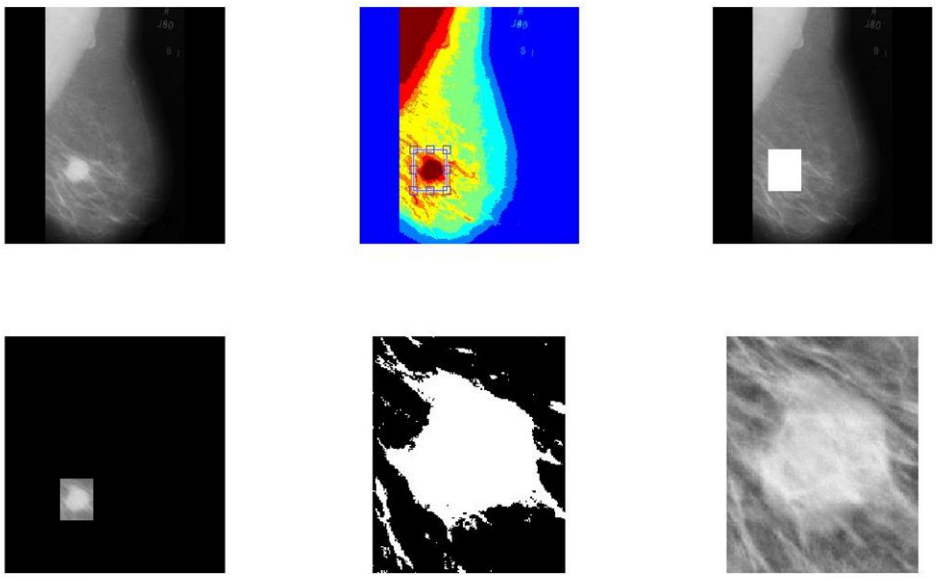


Figure (5-15): Mass segmentation steps

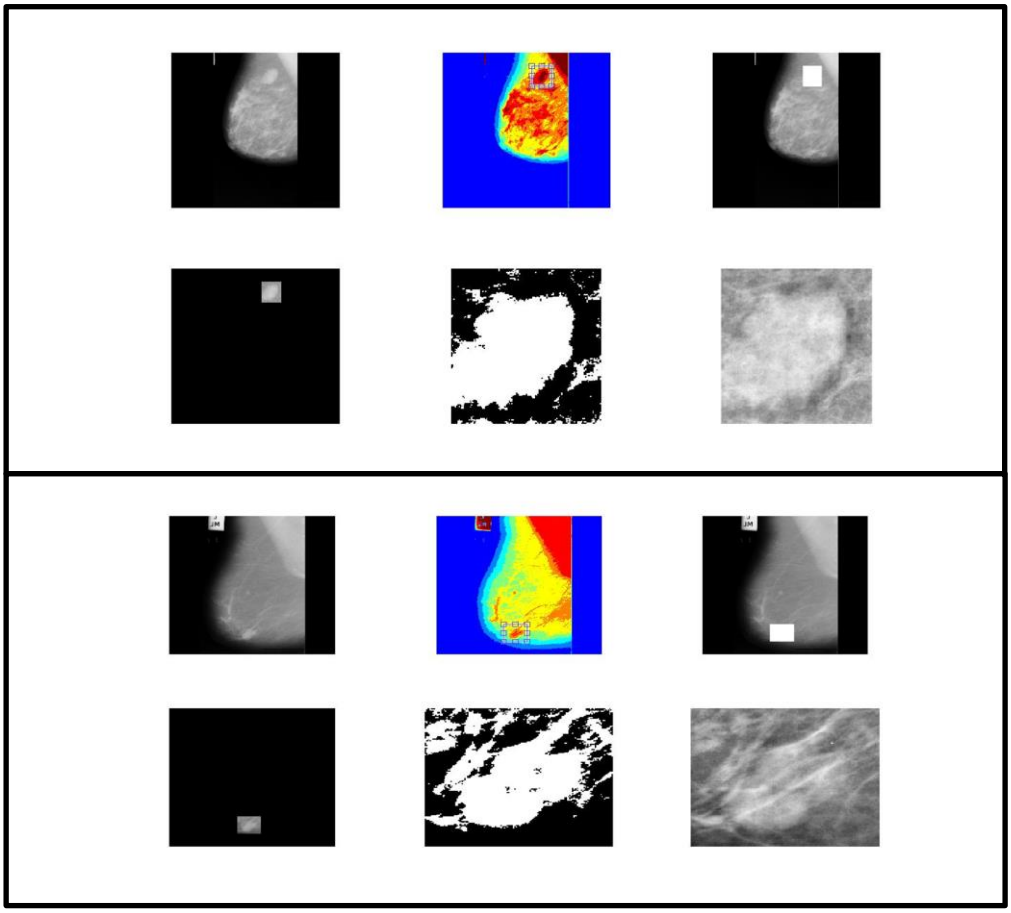


Figure (5-16): Pseudo-color based- segmentation examples

5.3. Tissue Classification performance

According to mRMR selection method, the selected features are as in [Table (5-1)].

In this work, three studies are investigated for extracting features, and the best accuracy gained is 88% for distinguishing fatty tissue from non-fatty in the 1st study (preprocessed image selection) using SFTA features. Moreover, the best accuracy 78 % gained in differentiated between dense tissues and glandular in the 3rd study (window (128×128)) using combined Haralick and SFTA features.

For tissue classification we gained more better accuracy when calculate the average of feature for many points for one breast compared with one point, and the best accuracy gained by calculating the average feature for both breast (right and left) for a case study [Figure (5-17)] . [Figure (5-18)], illustrated the increment of quadratic classifier performance when an average feature of two breast sides computed.

Table (5-1): Features selection in tissue classification task

Feature extraction	**Study1: Preprocessed image**			**Study2: Five windows (50×50pixel) per image**			**Study3: Window 128×128pixel**		
	S_1C_1	S_1C_2	S_1C_3	S_2C_1	S_2C_2	S_2C_3	S_3C_1	S_3C_2	S_3C_3
Haralick	H1, H4, H6, H7, H8, H12, H13	H1, H2, H3, H9, H10, H12, H13	H1, H3, H4, H6, H7, H10, H13	H2, H3, H6, H7, H9, H10, H13	H3, H5, H6, H7, H10, H11, H12	H2, H3, H6, H7, H9, H10, H13	H3, H5, H6, H7, H10, H12, H13	H3, H5, H7, H9, H10, H11, H12	H3, H5, H6, H7, H10, H11, H12
SFTA	F3, F10, F11, F12, F15, F19, F20	F2, F5, F10, F11, F12, F19, F21	F2, F5, F10, F11, F15, F19, F20	F2, F5, F8, F10, F11, F19, F21	F2, F5, F8, F11, F12, F19, F21	F2, F5, F8, F10, F11, F19, F21	F2, F5, F10, F11, F12, F19, F21	F2, F5, F10, F11, F12, F19, F21	F2, F5, F10, F11, F12, F19, F21
Combined	H1, F10, F11, F12, F15, F19, F20	H2, H3,H10 , F2, F5,F11, F12	H1, F2, F5, F10, F11, F19, F20	H3, H6, H7, H10, F11, F19, F21	H3, H5, H6, H7, H11, H12, F21	H3, H7, H9, H10, F11, F19, F21	H3, H7, H10, F2, F11, F19, F21	H5, H9, H11, H12, F2, F19, F21	H3, H7, H10, H12, F11, F19, F21

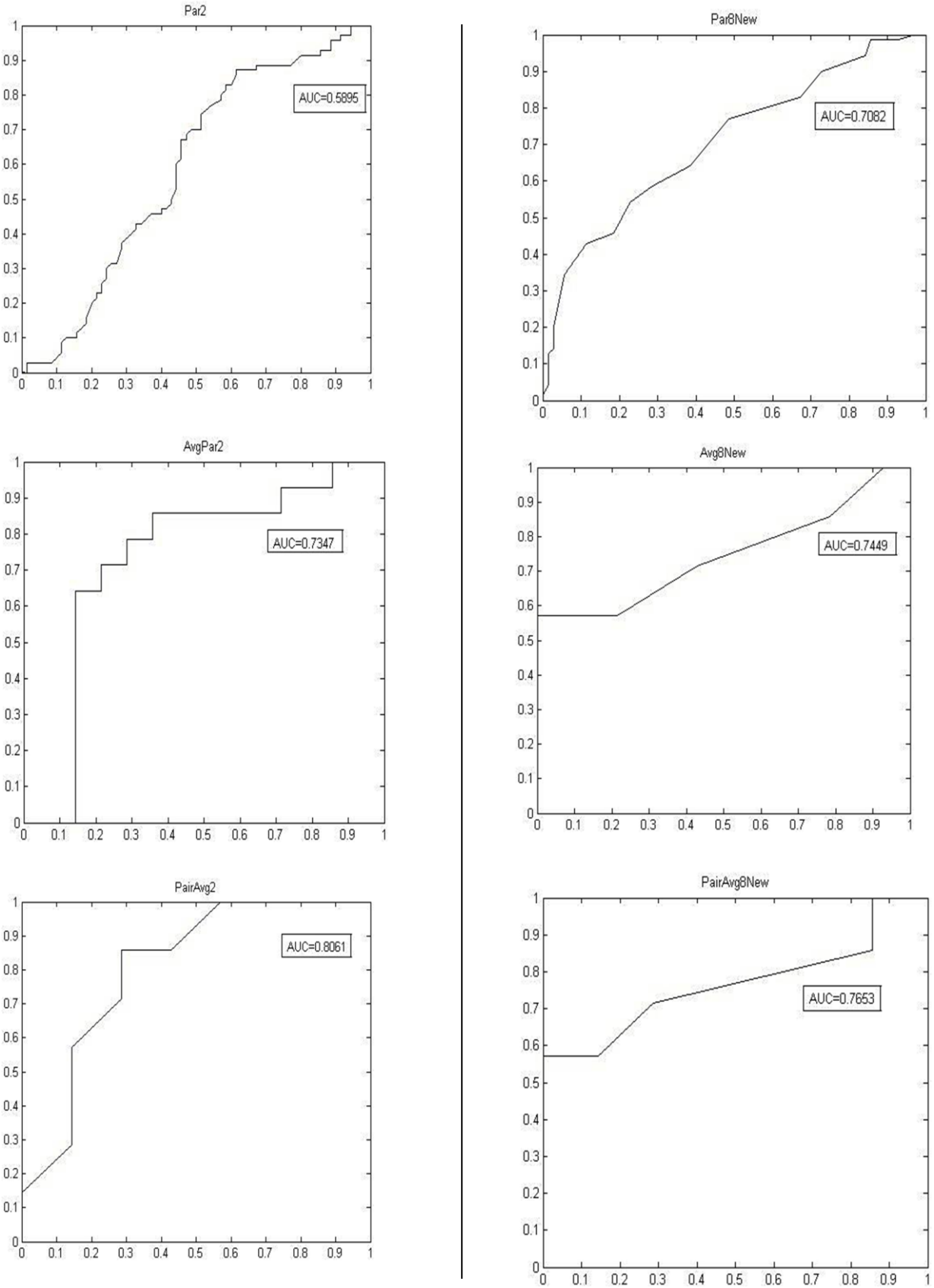
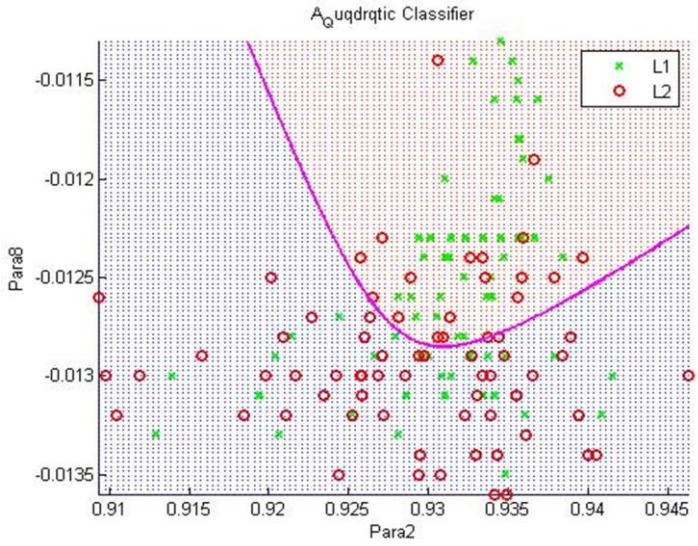
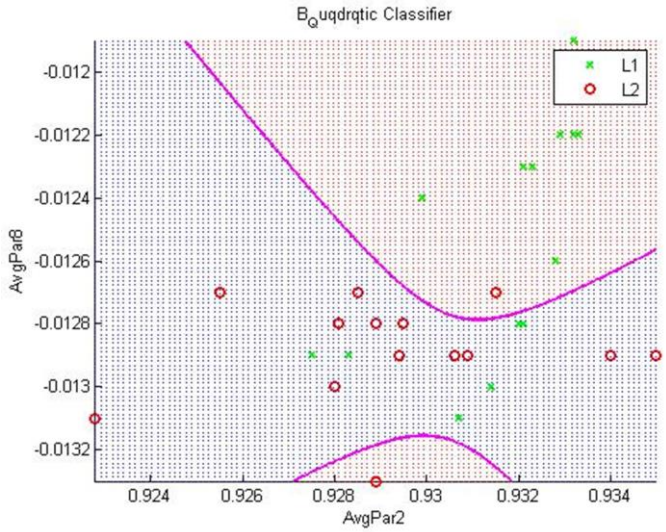


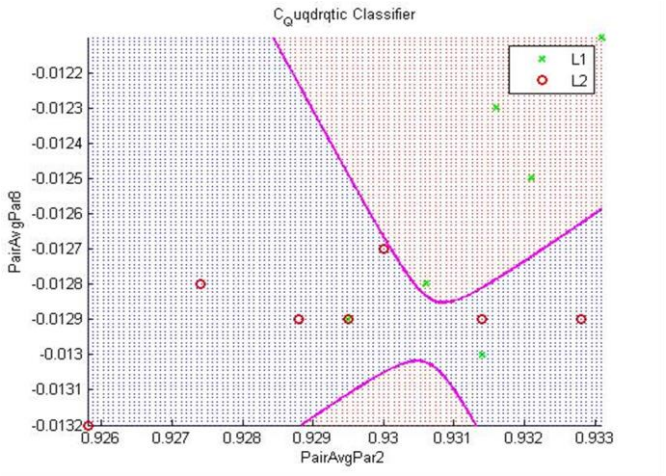
Figure (5-17): ROC curves (*x-axis*: sensitivity; *y-axis*: specificity), for individual feature, average feature, and pair of breast average feature, Feature1 (*left* side: *upper* to *lower*), Feature2 (*right* side from *upper* to *lower*)



AUC = 0.6643
Sensitivity = 0.7286
Specificity = 0.6000



AUC = 0.7143
Sensitivity = 0.8571
Specificity = 0.5714



AUC = 0.8571
Sensitivity = 1
Specificity = 0.7143

Figure (5-18): Quadratic classifier for individual feature parameter, average, and average of pair

5.3.1. Preprocessed image (S_1)

1. [Tables (5-2)] Fatty/ non-fatty classification (S_1C_1)

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	22	4
	D/G	10	45

a) Haralick features (H1, H6, H13, H12, H7, H8, H4), Acc. = 82.716%

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	23	3
	D/G	7	48

b) SFTA features (F10, F3, F20, F19, F15, F11, F12), Acc. = 87.654%

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	26	0
	D/G	14	41

c) Combined features (F10, H1, F20, F19, F15, F11, F12), Acc. = 82.716%

2. [Tables (5-3)] Glandular/ Dense classification (S_1C_2)

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	0	27
	D	0	28

a) Haralick features (H10, H12, H2, H13, H9, H1, H3), Acc. = 50.91%

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	21	6
	D	12	16

b) SFTA features (F5, F20, F19, F11, F2, F10, F12), Acc. = 67.27%

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	14	13
	D	15	13

c) Combined features (H10, H2, F5, F11, F2, F12, H3), Acc. = 49.09%

3. [Tables (5-4)] Fatty/ Glandular/ Dense classification (S_1C_3)

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	22	0	4
	G	8	4	15
	D	3	12	13

a) Haralick features (H1, H6, H3, H13, H10, H7, H4), Acc. = 48.15%

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	25	1	0
	G	6	10	11
	D	4	13	11

b) SFTA features ((F10, F2, F19, F15, F11, F5, F20), Acc. = 56.79%

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	25	0	1
	G	10	4	13
	D	4	17	7

c) Combined features (F10, H1, F2, F19, F20, F11, F5), Acc. = 44.44%

5.3.2. Five windows per image (S_2)

1. [Tables (5-5)] Fatty/ non-fatty classification (S_2C_1)

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	22	4
	D/G	11	44

a) Haralick features (H10, H13, H3, H7, H9, H6, H2), Acc. = 81.481%

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	20	6
	D/G	5	50

b) SFTA features (F19, F5, F10, F2, F21, F11, F8), Acc. = 86.419%

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	22	4
	D/G	13	42

c) Combined features (F19, H6, F21, H3, H10, H7, F11), Acc. = 79.012%

2. [Tables (5-6)] Glandular/ Dense classification (S_2C_2)

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	20	7
	D	14	14

a) Haralick features (H11, H3, H5, H12, H10, H7, H6), Acc. = 61.8182%

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	21	6
	D	12	16

b) SFTA features (F21, F8, F19, F5, F12, F11, F2), Acc. = 67. 2727%

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	17	10
	D	9	19

c) Combined features (H11, F21, H3, d) H5, H12, H6, H7), Acc. = 65.4545%

3. [Tables (5-7)] Fatty/ Glandular/ Dense (S_2C_3)

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	22	0	4
	G	13	3	1
	D	3	13	12

a) Haralick features (H3, H13, H9, H7, H10, H2, H6), Acc. = 45.679%

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	21	0	5
	G	10	6	11
	D	0	17	11

b) SFTA features (F19, F8, F10, F5, F21, F2, F11), Acc. = 46.9135%

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	22	0	4
	G	11	4	12
	D	2	14	12

c) Combined features (F19, H9, F21, F11, H3, H10, H7), Acc. = 46.9135%

5.3.3. Window size (128×128) (S_3)

1. [Tables (5-8)]Fatty/ non-fatty classification (S_3C_1)

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	25	1
	D/G	12	43

a) Haralick features (H3, H13, H5, H10, H12, H7, H6), Acc. = 83.951%

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	20	6
	D/G	10	45

b) SFTA features (F21, , F10, F2, F5, F12, F11, F19), Acc. = 80.247%

Breast Tissue Types		Predicted Class	
		F	***D/G***
True Class	***F***	21	5
	D/G	11	44

c) Combined features (F21, H3, F19, H10, H7, F11, F2), Acc. = 80.247%

2. [Tables (5-9)] Glandular/ Dense classification (S_3C_2)

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	27	0
	D	28	0

a) Haralick features (H12, H3, H9, H10, H5, H11, H7), Acc. = 49.09%

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	19	8
	D	8	20

b) SFTA features (F21, F2, F19, F5, F10, F11, F12), Acc. = 70.91%

Breast Tissue Types		Predicted Class	
		G	***D***
True Class	***G***	23	4
	D	8	20

c) Combined features (H12, F19, F2, H9, F21, H11, H5), Acc. = 78.182%

3. [Tables (5-10)] Fatty/ Glandular/ Dense (S_3C_3)

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	23	2	1
	G	12	12	3
	D	2	12	14

a) Haralick features (H12, H3, H5, H10, H11, H7, H6), Acc. = 60.494

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	20	6	0
	G	9	10	8
	D	1	7	20

b) SFTA features (F21, F2, F10, F5, F12, F19, F11), Acc. = 61.728%

Breast Tissue Types		Predicted Class		
		F	***G***	***D***
True Class	***F***	23	3	0
	G	11	12	4
	D	2	9	17

c) Combined features (F21, H3, H12, F19, H10, , F11, H7), Acc. = 64.198%

5.4. Normal-abnormal classification

As SFTA features, are new established feature extraction method, the reliability of SFTA is tested using the learning dataset. For normal mammogram, a ROI in the center of image ($radius = 50$), for abnormal mammogram, a ROI selected according to abnormality coordinates (x, y, r) given by MIAS information details, as following samples in [Figure (5-19)]

MIAS dataset samples:

```
mdb007 G NORM
mdb013 G MISC B 667 365 31
mdb014 G NORM
mdb015 G CIRC B 595 864 68
```

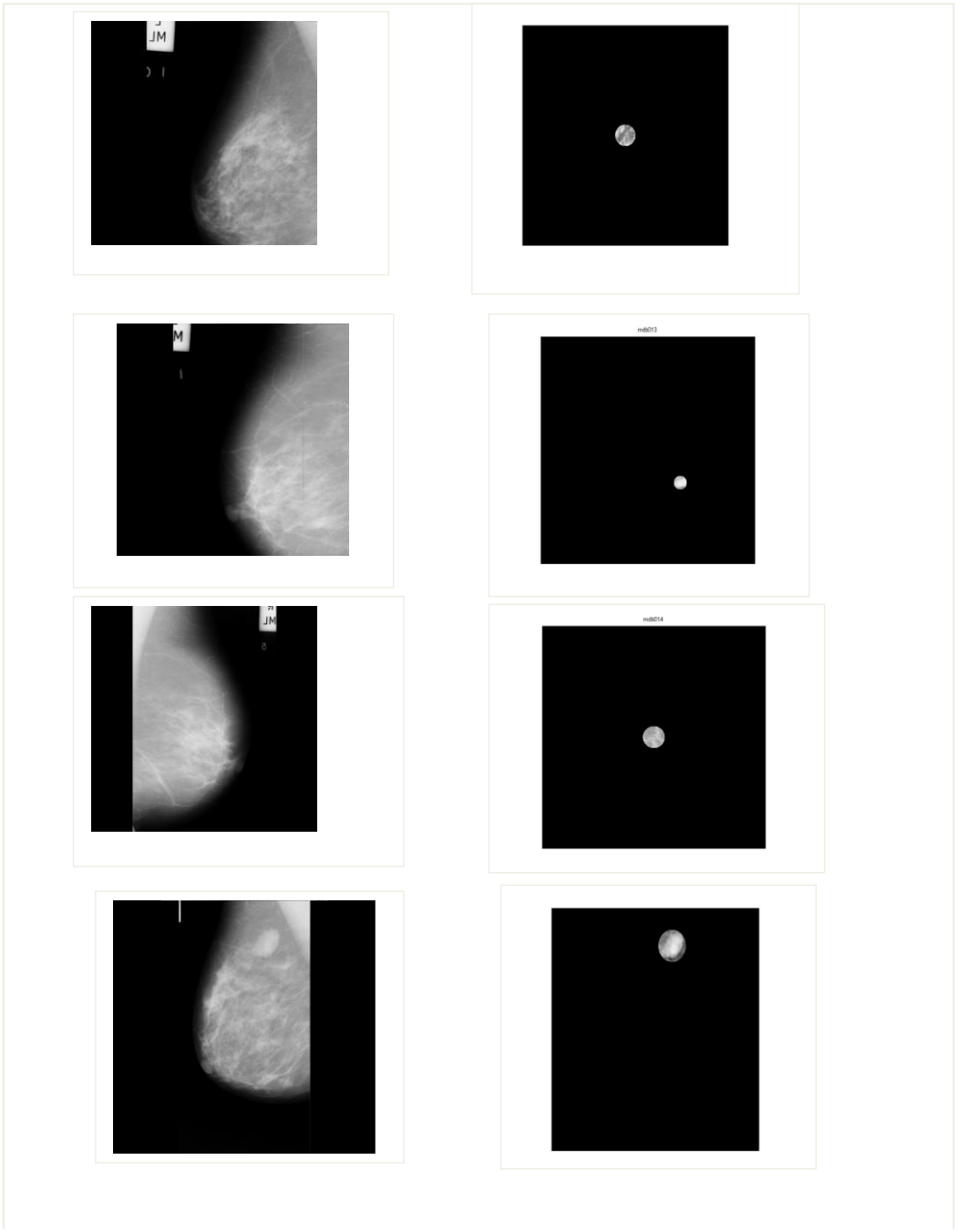


Figure (5-19): Various circular masks sizes, for abnormal images, and fixed for normal images; *(left* to *right*; *upper* to *lower*)

Figure (5-20), shows an automatic abnormality mask when user mark a point on suspicious region in RGB image.

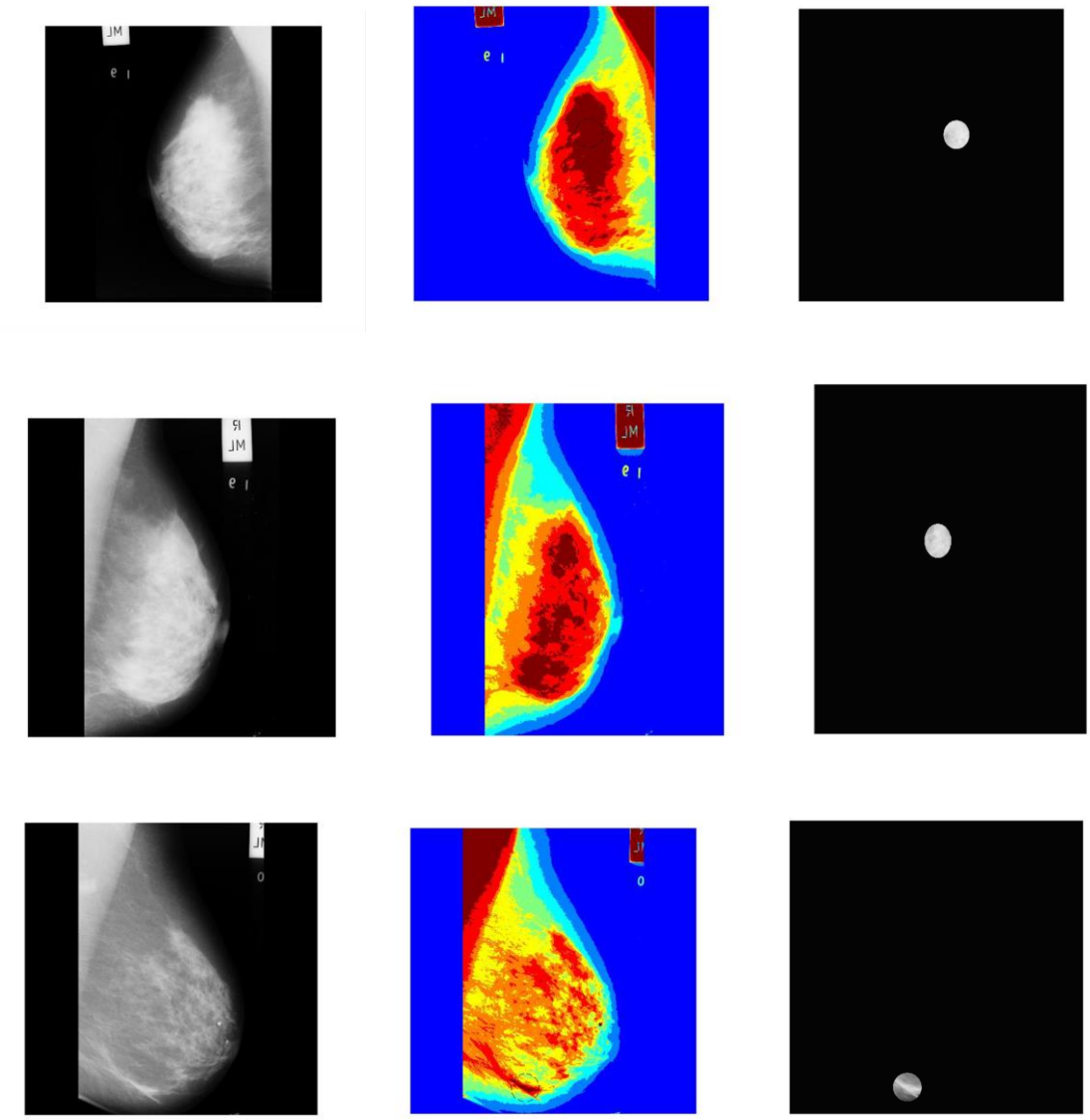


Figure (5-20): Automatic ROI selection with fixed size ($r = 50$)

The '*best*' SFTA features [Table (5-11)] computed from the areas of abnormalities on abnormal images, where the area (A) is given by: $A = \pi r^2$, and (r) is given on MIAS details. Results [Figure (5-21)], showed high correlation between the area of the abnormality circular mask and selected features.

[Table (5-12)] show the high accuracy of both Naïve and SVM classifiers for normal and abnormal classifications based on SFTA features.

Table (5-11): Performances of optimum SFTA features computed from different circular mask sizes'

Features	**F1**	**F3**	**F4**	**F6**	**F7**	**F12**	**F21**
AUC×100	0.538385	0.942503	0.555642	0.953302	0.673775	0.909505	0.92159

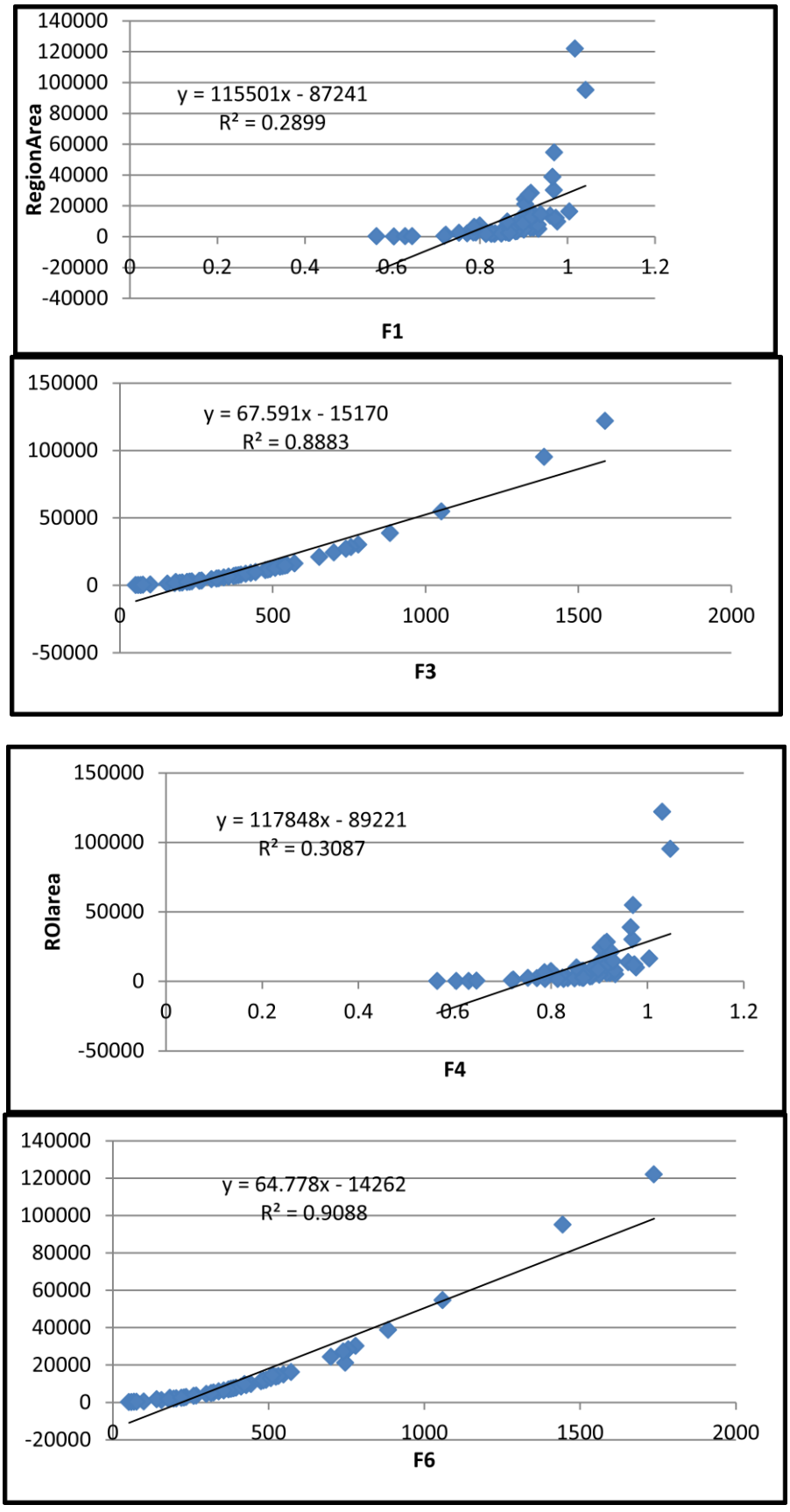


Figure (5-21): Correlation between features (F1, F3,F4, and F6) and area of abnormality, (*horizontal* axis: features; *vertical* axis: area ROI)

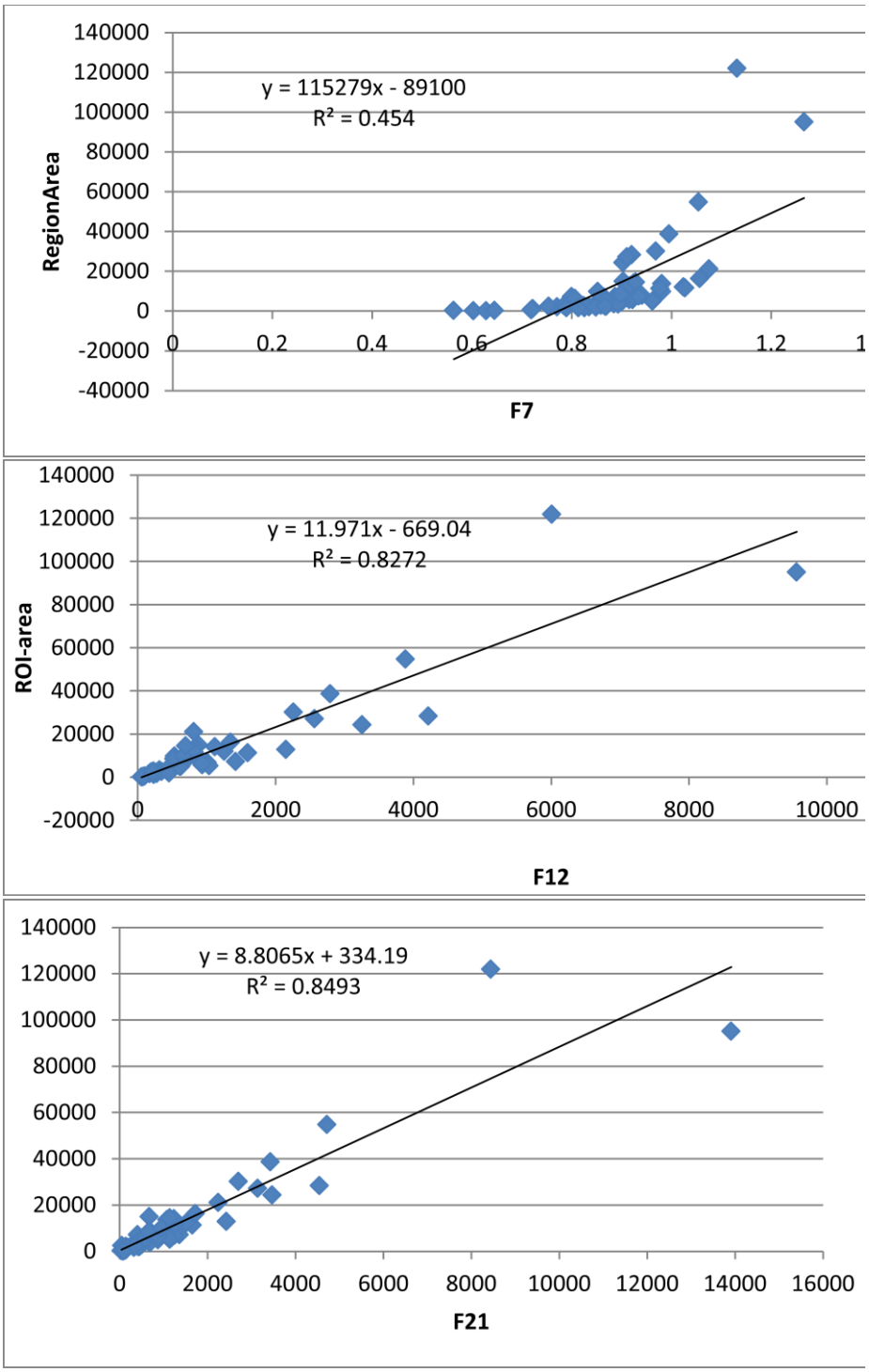


Figure (5-21): Correlation between features (F7, F12, and F21) and area of abnormality, (*horizontal* axis: features; *vertical* axis: area

Table (5-12): Confusion matrix for normal-abnormal classification based on SFTA features

Truth MIAS	SVM classification	
	Normal	Abnormal
Normal	109	0
Abnormal	4	52

Accuracy= 97.57%

Truth MIAS	Naive classification	
	Normal	Abnormal
Normal	108	1
Abnormal	0	56

Accuracy= 99.39%

As the setting-threshold for SFTA method and the window size play important factors in feature extraction and classification, we applied following tests:

1. Compute the classification accuracy, by test multi threshold values at fixed window size [Table (5-13)]. Experimentally approved the adequate threshold value is (3).
2. Fix the threshold to($level$ 3) and tested the classification accuracy for SFTA and Haralick features using different window sizes, [Figure (5-22)].
3. Select the window size, which gained better accuracy for the two types of features (Haralick, SFTA), [Table (5-14)].
4. Fix the threshold to ($level$ 3), and window size to (32×32), then select the features

- **Haralick feature** : H1, H3, H5-H6, H9-H11
- **SFTA features**: F2, F5, F7 - F9, F11, F14

Table (5-13): Classification accuracies based on SFTA features, using different threshold values and fixed window size (64×64)

Threshold Setting	Accuracy (%)
2	69.1
3	71.8
4	69.1
5	71.0
6	67.3
7	71.8
8	69.1

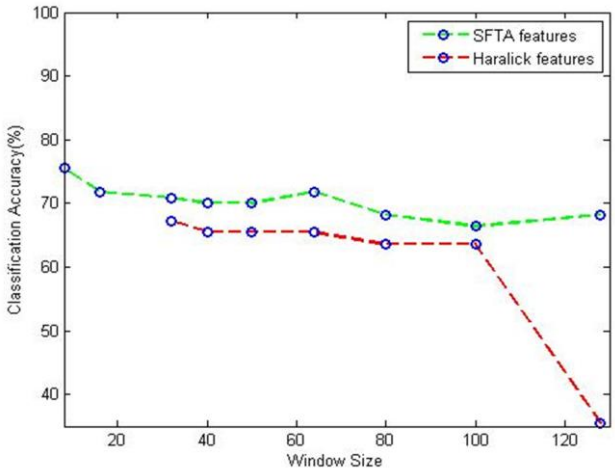


Figure (5-22): Classification accuracies of SFTA and Haralick features at different window sizes

Table (5-14): Classification accuracies based on SFTA and Haralick features, using different window sizes

	Accuracy (%)	
Window Size	**SFTA**	**Haralick**
8 × 8	75.5	NaN
16 × 16	71.8	NaN
32 × 32	70.9	67.3
40 × 40	70.0	65.5
50 × 50	70.0	65.5
64×64	71.8	65.5
80 × 80	68.2	63.6
100 × 100	66.4	63.6
128 × 128	68.2	35.5

5.5. Benign/ Malignant classification

For classification task, optimum features were selected:

1. Haralick features: H3, H4, H5, H9, H10, H11 and H12
2. SFTA features: F1, F2, F3, F6, F10, F12 and F14
3. Combined features: F1-F3, F6, F10, F14, H3-H5, H9-H12

Haralick and SFTA features performances on benign-malignant classification are tested using ROC curves as in [Figure (5-23)], [Figure (5-24)], performance results showed in [Table (5-15)], and the gained accuracy in [Table (5-16)]

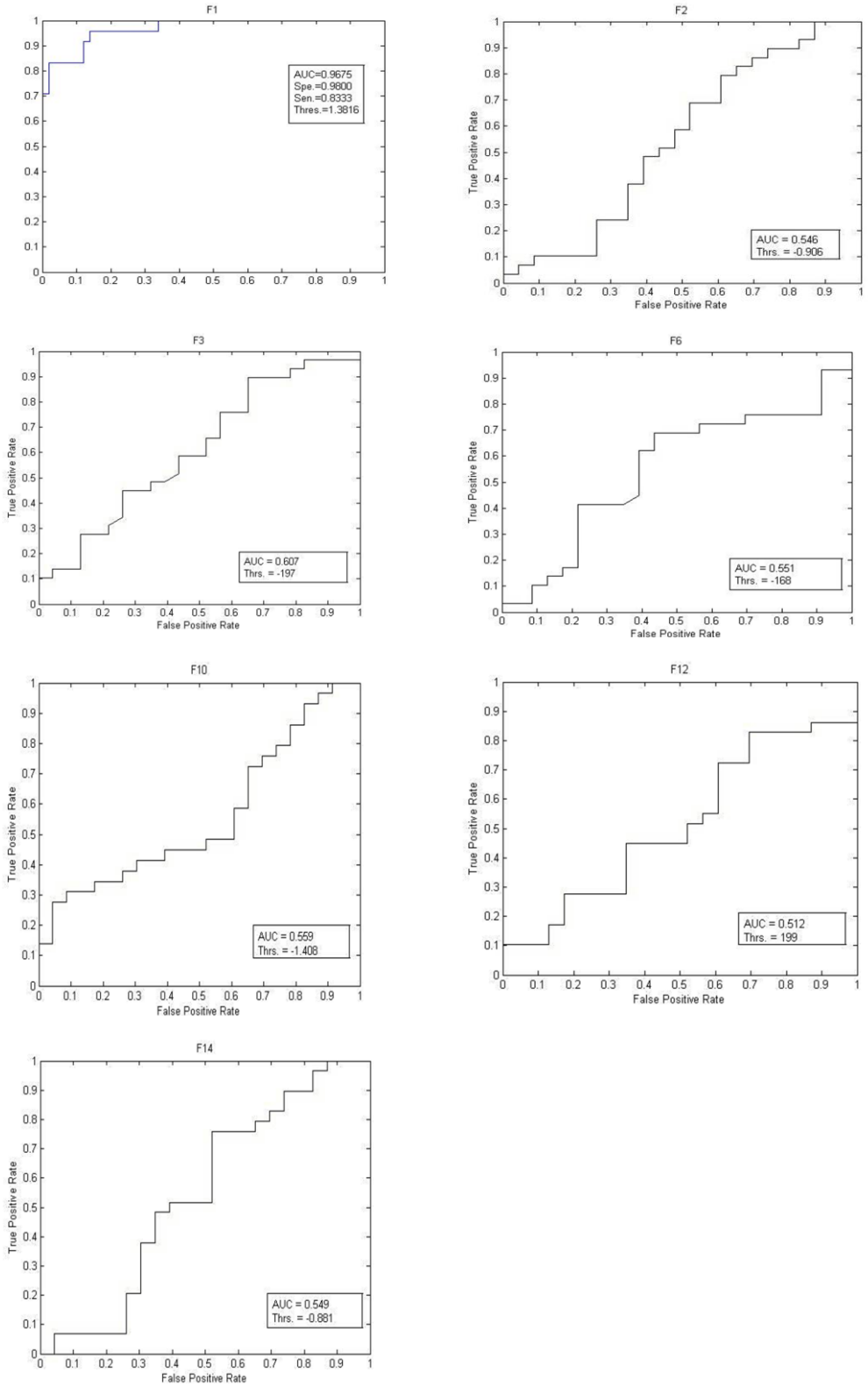


Figure (5-23): ROC curves for SFTA features (F1, F2, F3, F6, F10, F12 and F14) (*left* to *right*; *upper* to *lower*); (*x-axis*: false positive rate, *y-axis*: true positive rate)

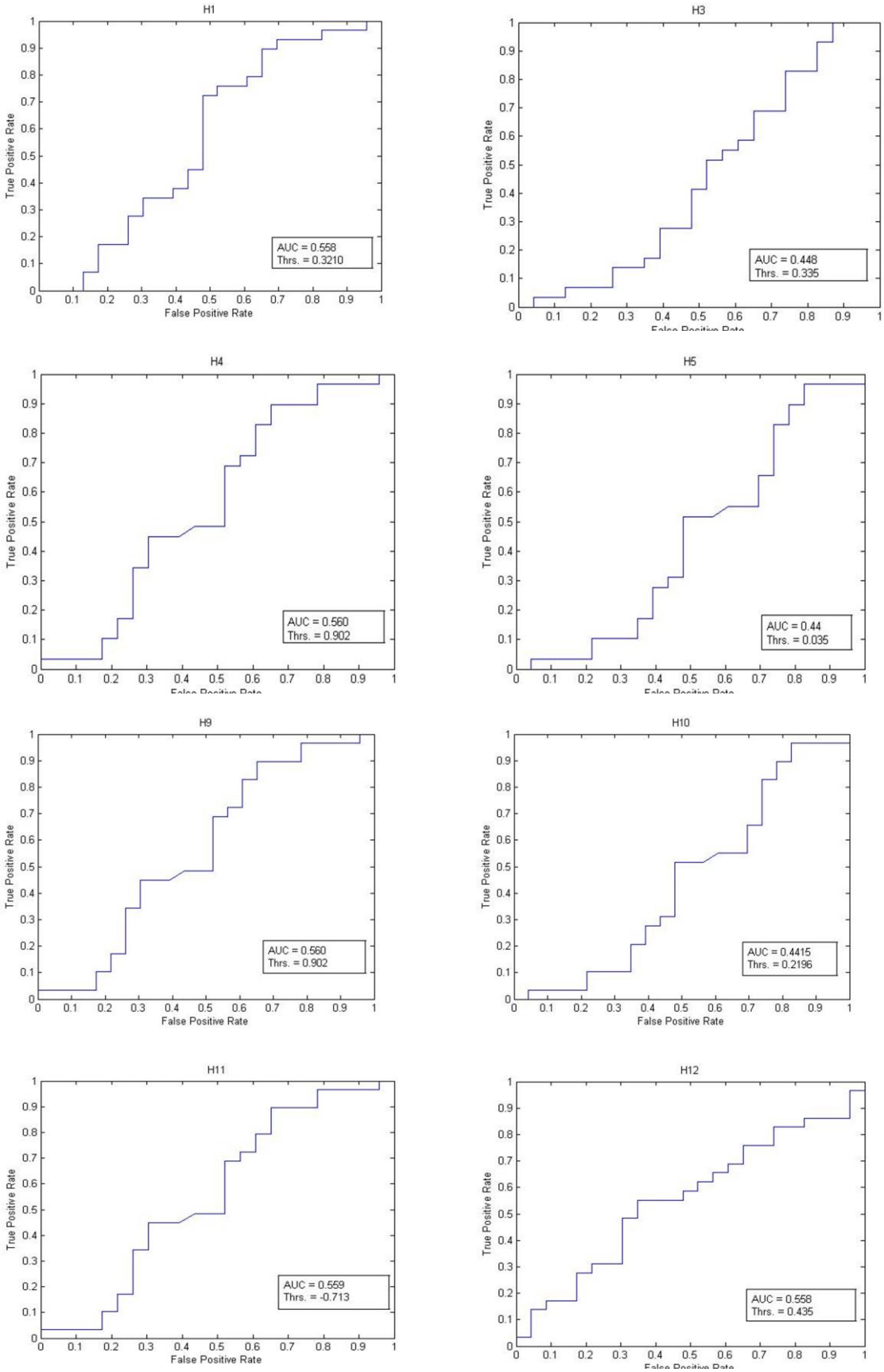


Figure (5-24): ROC curves for selected Haralick features, (H1, H3, H4, H5, H9, H10, H11 and H12) (*left* to *right*; *upper* to *lower*); (*x-axis*: false positive rate, *y-axis*: true positive rate)

Table (5-15): Performances of different features for benign/malignant classification

	SFTA							**Haralick**						
	F1	F2	F3	F6	F10	F12	F14	H3	H4	H5	H9	H10	H11	H12
AUC (× 100)	74	77	72	76	70	69	75	76	69	69	78	69	69	68

Table (5-16): Benign/ Malignant classification accuracies

	SFTA features	**Haralick features**	**Combined features**
Acc. (%)	74.5	68.2	71.8

6. Discussion

6.1. Using of MIAS database

MIAS is well-known, broadly, free and easy access database, the database has ground truth from the radiologists about characteristic of background tissue, type of abnormality, severity of abnormality, the coordinates of center and approximate radius (in pixels) of a circle enclosing the abnormality.

MIAS in MLO view, MLO is the most important projection, as it allows depicting most breast tissue, while, the CC view is taken from above, resulting in an image that sometimes does not show the area close to the chest wall (Maitra et al., 2012).

As we work on tissue calcifications, we extracted features from two breasts (right and left) for one patient. Our main work concentrates on case (patient) studies rather than individual image.

Mammographic databases play an important role in the development of algorithms for CAD system. MIAS databases allow comparison of results from different studies (Antoniou et al., 2009).

For all above advantages, MIAS database is used.

Also, The (MIAS) abbreviation also refers to Medical Image and Signals, the MIAS-GRID started mid-2002, a project by University of Oxford [http://www.robots.ox.ac.uk/Irc/grid_mias-grid.html].

6.2. Preprocessing

6.2.1. Unwanted regions removal and mammogram enhancement

Converting the mammogram image to binary image is the main role in preprocessing stage, the advantages of using binary image, that it needs smaller memory and faster execution time. Moreover, morphological operations and connected components techniques which used for unwanted regions removal were developed for binary image. Binarization process applied by using Otsu's thresholding, which is the most successful global thresholding method, it automatically performs histogram shape, and choose optimal threshold by maximizing the between class variance for choosing optimal threshold.

For enhancement, median filter and CLAHE technique were used, which widely used in medical image processing.

Median filtering is a nonlinear process, often used to remove impulsive or salt-and-pepper noise. Edges are of critical importance to the visual appearance of image, median filtering is used because of its useful in preventing edges in an image (Guohong and Wenming, 2010) and reducing impulse and random noises.

The advantageous of CLAHE technique, it is not discard the part of the histogram that exceeds the clip limit but redistribute the histogram equally among all histogram bins.

6.2.2. Pectoral muscle orientation and identification

For orientation, our method rely on the MIAS database arrangement of images, odd file number addressed the right breast image, while even file number addressed left breast image. So, we simply flip the images which contained odd file number.

Surveying other approaches, the simple method for pectoral muscle reorientation based on thresholding (Dehghani and Dezfooli, 2011), (Chaabani et al., 2010). The image divided into two halves in comparison with vertical axis as in [Figure (6-1)], then the grey level of threshold limit of each part was calculated. Surly, the threshold limit under the image which contains the breast region is more than the under of image which contains the background, after finding the breast direction, it putted to left direction.

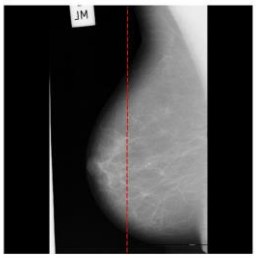


Figure (6-1): Pectoral muscle orientation, by image halves

For pectoral muscle segmentation, we set multi-level threshold to level-3, we based simply on fact that, background lies on the beginning of histogram, tissue lies in middle, and muscle lies in the third zone of histogram. Moreover, extensive experiments are conducted on MIAS mammographic images and threshold $level - 3$ showed best identification.

6.2.3. Comparative analysis

Numerous papers introduced different approaches for preprocessing mammogram images, here is comparative analyses introduce methods are used and results, comparing with our approach. See [Table (6-1)]

Sara and Mashalla in (Dehghani & Dezfooli, 2011) are introduced a preprocessing technique based on threshold to omitting excessive sides and put all images in one side, then eliminated labels and background based on region growing method. The result obtained 99% by applied on 60 images of MIAS database.

In (Camilus et al., 2011), pectoral muscle removed by using watershed transformation and merging algorithm. The method applied on 84 mammograms from MIAS database, the results obtained are 0.85% and 4.88% for mean false positive and mean false negative rates respectively comparing with manually identified pectoral muscle.

Researchers in (Maitra, 2012) applied three steps on MIAS database. First step is contrast enhancement then pectoral muscle detection and suppression, they used CLAHE technique and seeds region growing. Out of 322 mammogram images, fourteen images failed in output; the accuracy obtained was95%.

In (Yoon et al., 2016) a study applied on all MIAS database for pectoral muscle segmentation. Pectoral muscle detected by using the morphological method and the random sample consensus (RANSAC) algorithm. The results showed 92.2% accuracy.

Table (6-1): Summary of results and comparison with existing work for preprocessing

Authors	**Year**	**MIAS Database No.**	**Function**	**Method**	**Acc. (%)**
Raba et al.	2005	300	Pectoral muscle suppression	Region growing	98
Dehghani & Dezfooli	2011	60	Background elimination	Thresholding , and region growing	99
Camilus et al.	2011	84	Pectoral muscle identification	Watershed transformation, and merging algorithm	85
Maitra et al.	2012	322	Contrast enhancement, pectoral muscle suppression	CLAHE technique, and seeds region growing	95
Yoon et al.	2016	322	Pectoral muscle segmentation	Morphological operation, and random sample consensus algorithm	92
Our approach	2019	160	Background removal, and pectoral muscle suppression	Otsu’s thresholding	96

6.3. Segmentation

In this research, the mass segmentation technique applied is using one single view, segmentation methods are applied for both circular and spiculated masses, however, most of the algorithms are only able to detect a specific kind of mass, usually circular or spiculated masses. There are very few works using pseudo color in mammogram image processing (enhancement).

In some mammogram images, the result for segmentation method include the regions containing all mases even with some false positives (FP), however (FPs) will be removed at later step.

In this research simple algorithms based on thresholding and pseudo color are used, the segmentation method was determined if the proposed CAD is fully-automated or semi-automated.

6.4. Texture features

For all classification tasks, we based on Haralick and SFTA texture features

- **Haralick features:** it remains popular today, by virtue of good performance (Nixon and Aguado, 2008).
- **SFTA:** the rationale for using pairs of thresholds to compute the set of binary images is to segment objects that otherwise would not be segmented by regular threshold segmentation; this is especially TRUE for objects and structures whose gray level lies in the middle range of the input histogram (Costa et al, 2012). [Figure (6-2)] illustrates that there are three different zones, and breast region (tissue) lies in the middle range of histogram.

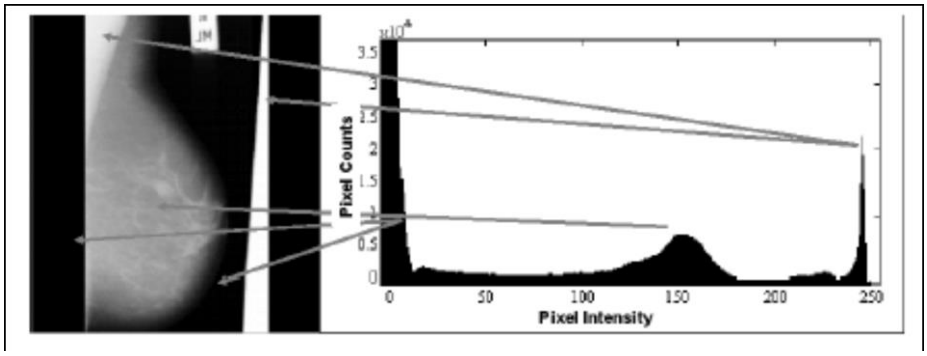


Figure (6-2): Histogram zones for mammogram images

Some of the factors that drastically influence the feature extraction results are:

1. The variability of the anatomy of the breast, every mammogram has different properties related to different tissue types and correspondingly variable brightness in the mammographic appearance.
2. The imaging conditions- shot noise, quantum mottle, patient movement, low contrast in mammograms due to low X-ray dosage and glare.
3. For micro-calcifications, faint micro-calcifications are lost in dense background
4. The superposition of certain breast structure. (Linguraru, 2002)
5. Window size for the computational methods utilized to extract features from textured images (Puig and Garcia, 2001). In this research the determination of tissue density is a qualitative measurement, considering the estimation of overall density. So, for tissue classification we used large windows [all breast region in preprocessed image, 5 windows per image, and (128×128) window] to obtain meaningful description of tissue type.

Table (6-2): Comparison between Haralick and SFTA texture features

Haralick Features	**SFTA Features**
Widely employed, statistical approach	New employed, structural approach
Orientations and scales dependent	Threshold dependent
Contains fixed number of features (14-features)	Number of features is dependent to user identifier of thresholds.

6.5. Tissue characterization

Breast density is not fixed but change over time. Some breasts may appear more or less dense when imaged using FFDM compared to SFM. Superior depiction of the skin line by digital mammography provides the observer with a more accurate (usually larger) estimate of the extent of the subcutaneous fat. However, no change in the distribution across density categories has been observed (D'Orsi et al., 2013).

[Table (6-3)] shows the common features which repeatedly in the three types of studies (preprocessed image, five windows per image, and window size (128×128).

Table (6-3): Effective features for tissue classification

	All Studies (S_1, S_2, and S_3)		
Features	Fatty-Non fatty (C_1)	Glandular-Dense (C_2)	Fatty-Glandular-Dense (C_3)
Haralick	H6, H13	H3, H10, H12	H3, H6, H7, H10
SFTA	F10, F11, F19	F2, F5, F11, F12, F19, F21	F2,F5, F10, F11, F19
Combined	F10, F11, F19	-	F11, F19

Obviously, the effective Haralick features are: H3, H6, and H10, which represent inertia, sum-average, and diff-variance. The effective SFTA features are: F11, F19, and F10. Moreover, in case of combined Haralick and SFTA features, all studies are SFTA features dependency.

The best accuracy 87% obtained for distinguishing fatty tissue from non-fatty, so as other researchers conclude (Chérel et al., 2008), the characterization of fatty breast tissue is easier than glandular or dense breast tissue.

6.5.1. MIAS database and BI-RADS assessment

Although MIAS database has three categories of tissue types, fat, dense, glandular, but there is no criterion standard exists for breast density assessment. Radiologists' visual assessment is known to be influenced by many factors. Even BIRADS breast density assessment method is subjective estimation method.

According to BI-RADS category, there is only a minimal and insignificant difference in the sensitivity of mammography between the densest breast in a lower density category and the least dense breast in the next higher density category (D'Orsi et al., 2013).

According to the ACR BI-RADS density classification standard, the MIAS database is divided into four density categories (Oliver et al., 2010) as in [Table (6-4)]

Table (6-4): Confusion matrix between the classification of MIAS according to its annotations (F, G, and D) and the consensus of three radiologists in BIRADS term, [Oliver et al., 2010]

	B-I	B-II	B-III	B-IV	Total
Fatty	83	23	0	0	106
Glandular	4	60	38	2	104
Dense	0	20	57	35	112
Total	87	103	95	37	322

6.5.2. Comparative analysis

In this section, researches related to breast density classification highlight, which using MIAS database.

Oliver et al. proposed several approaches focused on breast density. Examples of his researches: In (Oliver et al., 2005) proposed a new approach to classification of mammography according to the breast parenchymal density, the classification based on gross segmentation and the underlying texture contained within breast tissue the method applied on a set of 270 mammogram images of MIAS database, the obtained results for the leave-one-out classification method and k-NN classifier are 73% and 67%. In (Oliver et al., 2008) they used the whole set of MIAS database and 831 images from DDSM database, segmented the

breast into fatty and dense regions based on a two class fuzzy C-means clustering approach, then extracted 10 morphological features and 216 texture features finally used number of distinct classifiers (decision tree, Bayesian, and K-nearest neighbor), obtained accuracies are: 72%, 77%, and 82%, for four classes categories. The evaluation shows strong correlation between automatic and expert based Breast Reporting and Data System (BIRADS) mammographic density assessment.

(Sharma and Singh, 2014), presents a hybrid scheme for two class problem (fatty and dense) mammograms using correlation based feature selection (CFS), the classification performed using sequential minimal optimization (SMO). Texture analysis done on ROI of 322 images of MIAS database. The accuracy obtained is 96.46%.

(Mustra et al., 2012), based on histogram and GLCMs, 419 features were extracted from ROI of MIAS database, and then used different feature selection algorithm. Different selection methods tested with different classifiers. Using forward - backward feature selection method and K-NN classifier, they obtained different accuracies for different categories. Classification accuracies rate are: 91.6%, 82.5%, and 79.3% respectively for two-classes, three-class's, and four-class's categories.

(Silva and Menotti, 2012), used individual and combining various sets of statistical features for tissue classification, the classification is performed using SVM with RBF kernel. They used 320 MIAS mammograms images; obtained 77.18% accuracy rate for fully mammography by combining texture of image histogram intensity and co-occurrence matrix.

(Subashini et al., 2010) uses nine statistics features extracted from the image histogram and Support vector machine SVM classifier obtained accuracy 95% on the 43 MIAS database mammogram.

6.6. Normal / abnormal classification

In this research, different thresholds are tested: $nt = \{2,3,4,5,6,7,8\}$, with fixed window size(64×64). The best classification accuracy obtained at: $nt = \{3, 7\}$, as in [Figure(6-3)]. To have a reasonable number of features, we selected threshold $level3$, and ignored threshold $level7$

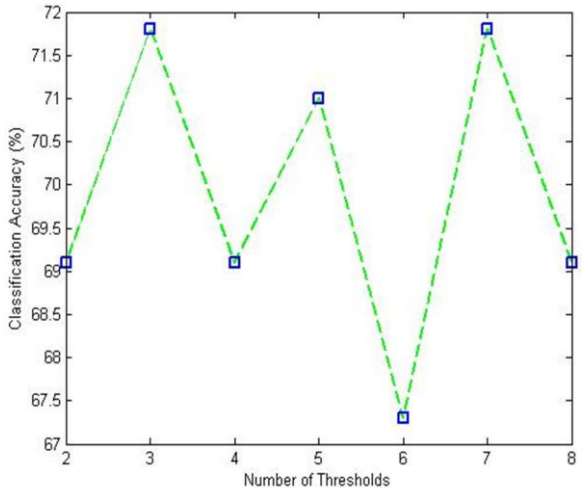


Figure (6-3): Classification accuracies of the SFTA feature vector obtained employing different number of threshold values with fixed window size (64×64)

6.6.1. Window size

Window sizes are determined in order to outperform two complementary tasks, texture feature and texture segmentation. Texture feature evaluation requires large windows in order to obtain meaningful descriptions of their content; texture segmentation requires small windows in order to locate the boundaries between different textural regions.

For normal/ abnormal classification, various window sizes are tested. As shown from results, in case of Haralick features the accuracy rates, at window sizes below (32×32) not considered, because some computed features are undefined values(NaN).

6.6.2. Effective features

From previous results, the best accuracy obtained was (75.5%), gained by using SFTA feature extraction method at window size (8×8). Moreover, in case of using Haralick features; the best accuracy is 67.3% at (32×32).window size

6.7. Effective features for classification tasks (tissue density, abnormality, class of abnormality)

Feature extraction is a primordial preprocessing task preceding classification. Methods are used, the well-known (Haralick features), and the new method (SFTA feature). The selectable methods used for three major classification tasks are:

1. Tissue classification
2. Normal/ abnormal classification
3. Benign/ malignant classification

In all classification tasks, the same criterion of classifying is followed, according to the best features from:

1. Haralick features
2. SFTA features
3. Combined features

Results in [Table (6-5)] show the superiority of SFTA features in classifying mammogram tissue types, Normal/ abnormal mammogram, and benign from malignant lesion.

Table (6-5): Best results for different classification tasks based on Haralick and SFTA feature extraction method

Classification task	**Feature extraction method**	**Acc.** (%)
Tissue types (Fatty / non-fatty)	Combined (Haralick and SFTA)	87.0
Tissue types (Glandular / Dense)	SFTA	78.0
Tissue types (Fatty / Glandular/ Dense)	Combined (Haralick and SFTA)	64.2
Normal / abnormal	SFTA	75.5
Benign / Malignant	SFTA	74.5

7. Conclusion and prospective work

The results of CAD in mammography are not yet conclusive enough to warrant a credible clinical usage. Different CAD algorithms show that the accuracy of cancer detection has indeed improved with introduction of CAD. Almost all of the existing CAD approaches are trained and tested on retrospectively collected databases that may not represent the real clinical practice. Large prospective studies are required to evaluate the performance of CAD systems in real life before employing them in clinical setting.

Researchers had been invested a lot of effort to characterize breast lesions and to investigate differentiation between lesions through different approaches. Usually, the surroundings (background) of the lesions are not including in the analysis of texture complexity. Moreover, CAD system should work as second opinion for the radiologist, so comprehensive studies must be applied to characterize the lesions, including features that are indistinguishable to the human eye.

As CAD algorithm consist of many steps, the performance of any step effect on overall performance of CAD algorithm. CAD algorithms can be stable or unstable. Stable algorithm gives substantially similar results when minor changes are made to the algorithm, features, or the training datasets. Unstable algorithm changes their output when such minor changes are made. The stability of algorithms that using automated or semi-automated methods measures by repeatedly retraining the algorithm with substantially different training sets, and then testing the algorithm on testing set. In the future, as breast cancer incidents increasing, well-designed and stable CAD algorithms should be investigated for early breast cancer detection, because breast cancer is curable if detected in early stages. CAD devices approved by the FDA are stable due to regular modification. The stability of algorithms increases as:

- The number of training cases increases.
- The number or dimensionality of initial features decreases.
- The complexity of the CAD decreases.

As there is strong correlation between breast density and breast cancer, radiologists should consider using the quantitative assessment method to provide consistent and reliable measurement of breast density. So, for the future this research may develop to provide quantitative assessment of breast density.

The need for awareness programs: time has come

As illustrated in the introduction, although the incidence of breast cancer in developing countries as Sudan is lower compared with developed countries, but the mortality rate is higher than developed countries due to many life factors, as poorness, less education, lack of continued educations and awareness programs

Since 1990s, Sudan has been on US blacklist, which led to the decline of health services. Even Khartoum Breast Cancer Centre (KBCC), which opened in 2010 (the Horn of Africa's first and only dedicated breast cancer clinic) has been hit by the sanctions with a ban on international money transfers and the restriction on imports of medical equipment and spare parts.

The Sudanese health ministry keeps no full records, because the health system focuses on communicable disease as malaria, tuberculosis and human immunodeficiency virus (HIV).

Time has come to make collaboration between universities and international organization, to get a reliable statistics of breast cancer in Sudan.

Breast cancer screening, awareness programs, and statistics should be under government responsibilities. But, researchers have a responsibility to find tracks to introduce awareness program for all society, including the simple human.

The need of SUST database

As Sudan University established a Faculty of Medicine, Hospital, and SUST also involves Engineering college (biomedical engineering department), and Radiological science, a collaboration need to establish SUST database, which will contain mammograms, ultrasounds images, MRI, CT scan images, and other medical images.

References

Abdul-Jaleel, J., Salim, S., Archana, 2014. Mammogram mass classification based on discrete wavelet transform textural features. 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI). doi: 10.1109/ICACCI.2014.6968479

Aggarwal, A., Singh, Karamjeet, Singh, Kamalpreet, 2015. Use of gradient technique for extracting features from handwritten Gurmukhi character and numerals. Procedia Computer Science, 46, 1716-1723. doi: 10.1016/j.procs.2015.02.116

Alolfe, M.A., Mohamed, W.A., Youssef, A.M., Mohamed, A.S., Kadah, Y.M., 2009. Computer aided diagnosis in digital mammography using combined support vector machine and linear discriminant analysis classification. 16th IEEE International Conference on Image Processing (ICIP), 2609-2612. IEEE. doi: 10.1109/ICIP.2009.5413992

Amendolia, S.R., Bisogni, M.G., Bottigli, U., Ceccopieri, A., Delogu, P., Dipasqale, G., Fantacci, M.E., Lorenzini, E., Marchi, A., Marzulli, V.M., Oliva, P., Palmiero, R., Reggiani, M., Rosso, V., Stefanini, A., Stumbo, S., Tangaro, S., Venier, O., 2001. The CALMA project. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 461, 428-429. doi: 10.1016/S0168-9002(00)01266-3

Antoniou, Z.C., Giannakopoulou, G.P., Andreadis, I.I., Nikita, K.S., Ligomenides, P.A., Spyrou, G.M., 2009. A web-accessible mammographic image database dedicated to combined training and evaluation of radiologists and machines. 9th International Conference on Information and Applications in Biomedicine, IEEE. doi: 10.1109/ITAB.2009.5394465

Aoki, K., Kudo, M., 2008. Feature and classifier selection in class decision trees. Structural, syntactic, and Statistical Pattern, SSPR/SPR., Lecture Notes in Computer Science 5342, 562-571. Springer. doi: 10.1007/978-3-540-89689-0_60

Arora, S., Acharya, J., Verma, A., Panigrahi, P.K., 2008. Multilevel thresholding for image segmentation through a fast statistical recursive algorithm. Pattern Recognition Letters 29, 119-125. Elsevier. doi: 10.1016/j.patrec.2007.09.005

Auweter, S.D., Herzen, J., Willner, M., Grandl, S., Scherer, K., Bamberg, F., Reiser, M.F., Hellerhoff, K., 2014. X-ray phase-contrast imaging of the breast—advances towards clinical implementation. Br J Radiol. 87. doi: 10.1259/bjr.20130606

Bai, C-L., Liu, H., 2011. Image enhancement based on wavelet transform with MATLAB. 2011 International Conference on Multimedia Technology, IEEE doi:10.1109/ICMT.2011.6001955

Bailie, R., Katzenellenbogen, J., Hoffman, M., Schierhout, G., Truter, H., Dent, D., Gudgeon, A., Van Zyl, J., Rosenberg, L., Shapiro, S., 1997. A case control study of breast cancer risk and exposure to injectable progestogen contraceptives. methods and patterns of use among controls. South African Medical Journal. 87, 302-305. doi: 10.7196/SAMJ.8262

Baker, J.A., Lo, J.Y., 2011. Breast tomosynthesis: state-of-the-art and review of the literature. Acad Radiol. 18, 1298-1310. doi: 10.1016/j.acra.2011.06.011

Balakumaran, T., Vennila, I., 2011. Detection of microcalcifications clusters in digital mammograms using multiresolution based foveal algorithm, In: World Congress on Information and Communication Technologies (WICT), 657-660. IEEE. doi: 10.1109/WICT.2011.6141323

Barlow, W.E., Lehman, C.D., Zheng, Y. Ballard-Barbash, R., Yankaskas, B.C., Cutter, G.R., Carney, P.A., Geller, B.M., Rosenberg, R., Kerlikowske, K., Weaverf, D.L., Taplin,

S.H., 2002. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. J Nati Cancer Inst. 94, 1151-1159. doi: 10.1093/jnci/94.15.1151

Bassett, L.W., Gold, R.H., 1987. Breast cancer detection: mammography and other methods in breast imaging, 2nd ed. Grune & Stratton Inc., Orlando, FL (USA)

Bassett, L.W., 2007. Standardize reporting for mammography: BI-RADS™. The Breast Journal 3, 207-210. doi: 10.1111/j.1524-4741.1997.tb00172.x

Beheshti, S.M., AhmadiNoubari, H., Fatemizadeh, E., Khalili, M., 2014. An efficient fractal method for detection and diagnosis of breast masses in mammograms. J Digit Imaging 27, 661-669. doi: 10.1007/s10278-013-9654-z

Beldalli, P., McCullagh, J., Phelan, N., Flanagan, F., 2010. Comprehensive dose survey of breast screening in Ireland. Radiation Protection Dosimetry 145, 52-60. doi: 10.1093/rpd/ncq375

Berber, T., Alpkocak, A., Balci, P., and Dicle, O., 2013. Breast mass contour segmentation algorithm in digital mammograms. Computer Methods and Programs in Biomedicine 110, 150-159, Elsevier, doi: 10.1016/j.cmpb.2012.11.003

Berry, D.A., Cronin, K.A., Plevritis, S.K., Clarke, L., Zelen, M., Mandelblatt, J.S., Yakovlev, A.Y., Habbema, J.D., Feuer, E.J., 2005. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 353, 1784-1792. doi: 10.1056/NEJMoa050518

Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computer & Geosciences 10, 191-203. doi: 10.1016/0098-3004(84)90020-7

Bharati, M.H., Liu, J.J., MacGregor, J.F., 2004. Image texture analysis: methods and comparisons. Chemometrics and Intelligent Laboratory Systems 72, 57-71. Elsevier. doi: 10.1016/j.chemolab.2004.02.005

Bhikoo, R., Srinivasa, S., Yu, T. C., Moss, D., Hill, A. G., 2011. Systematic review of breast cancer biology in developing countries (part 1): Africa, the Middle East, Eastern Europe, Mexico, the Caribbean and South America. Cancers (Basel) 3, 2358-2381. doi: 10.3390/cancers3022358

Biggelaar, F.J., Kessels, A.G., Engelshoven, J.M., Boetes, C., Flobbe, K., 2010. Computer-aided detection in full-field digital mammography in a clinical population: performance of radiologist and technologists. Breast Cancer Res Treat. 120, 499-506. doi: 10.1007/s10549-009-0409-y

Bilkova, A., Janik, V., Svoboda, B., 2010. Computed tomography laser mammography. Casopis Lekaru Ceskych. 149, 61-65. PMID: 20662467

Bindu, C.H., Prasad, K.S., 2012. An efficient medical image segmentation using conventional OTSU method. International Journal of Advanced Science and Technology 38, 67-74. doi:10.1.1.359.8532

Birdwell, R.L., Ikeda D.M., O'Shaughnessy, K.F., Sickles E.A., 2001. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. Radiology 219, 192-202. doi: 10.1148/radiology.219.1.r01ap16192

Birdwell, R.L., Bandodkar, P., Ikeda, D.M., 2005. Computer-aided detection with screening mammography in a university hospital setting. Radiology 236, 451-457. doi: 10.1148/radiol.2362040864

Bjelic-Radisic, V., Petru, E., 2010. Hormonal contraception and breast cancer risk. Wiener Medizinische Wochenschrift 160, 483-486. doi:10.1007/s10354-010-0807-0

Blackburn, H.L., Ellsworth, D.L., Shriver, C.D., Ellsworth, R.E., 2017. Breast cancer metastasis to the axillary lymph nodes: are changes to the lymph node "Soil" localized or systemic. Breast Cancer (Auckl) 11. doi: 10.1177/1178223417691246

Boeker, M., Franca, F., Bronsert, P., and Schulz, S., 2016. TNM-O: ontology support for staging of malignant tumors. Journal of Biomedical Semantics 7. doi:10.1186/s13326-016-0106-9

Bolivar, A.V., Gomez, S.S., Merino, P., Alsono-Bartolomé, P., Garcia, E.O., Cacho, P.M., Hoffmeister, J.W., 2010. Acta Radiol. 51, 1086-1092. doi: 10.3109/02841851.2010.520024

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: D. Haussler, editor, Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory. 144-152. Pittsburgh, PA, ACM Press. doi:10.1145/130385.130401

Bouyahia, S., Mbainaibeye, J., Ellouze, N., 2009. Wavelet based microcalcifications detection in digital mammograms. ICGST-GVIP 8, 23-30. ISSN: 1687-398X

Boyd, N.F., Byng, J.W., Jong, R.A., Fishell, E.K., Little, L.E., Miller, A.B., Lockwood, G.A., Tritchler, D.L., Yaffe, M.J., 1995. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian national breast screening study. Journal of the National Cancer Institute 87, 670-675. doi: 10.1093/jnci/87.9.670

Bronzino, J., 2000. The biomedical Engineering Handbook. U. S.: CRC Press, 2nd edition, vol. 1.

Burhenne, L.J., Wood, S.A., D'Orsi, C.J., Feig, S.A., Kopans, D.B., O'Shaughnessy, K.F., Sickles, E.A., Tabar, L., Vyborny, C.J., Castellino, R.A., 2000. Potential contribution of computer aided detection to the sensitivity of screening mammography. Radiology 215, 554-562. doi:10.1148/radiology.215.2.r00ma15554

Burnside, E.S., Eickles, E.A., Bassett, L.W., Rubin, D.L., Lee. C.H., Ikeda, D.M., Mendelson, E.B., Wilcox, P.A., Butler, P.F., D'Orsi, C.J., 2009. The ACR BI-RADS experience: learning from history. J Am Coll Radiol. 6, 851-860. doi: 10.1016/j.jacr.2009.07.023

Burnside, E., Rubin, D., Shachter, R., 2000. A Bayesian network for mammography. Proc AMIA Symp, 106-110. PMCID: PMC2243709

Bushberg, J.T., Seibert, J.A., Leidholdt, E.M., Boone, J.M., 2002. The essential of medical imaging. 2nd edition, by Lippincott Williams & Wilkins, pp. 191-192

Caldwell, C.B., Stapleton, S.J., Holdsworth, D.W., Jong, R.A., Weiser, W.J., Cooke, G., Yaffee, M.J., 1990. Characterization of mammographic parenchymal pattern by fractal dimension. *Phys. Med. Biol.* 35, 235-247. doi: 10.1117/12.953239

Camilus, K.S., Govindan, V.K., Sathidevi, P.S., 2010. Computer-aided identification of the pectoral muscle in digitized mammograms. Journal Digital Imaging 23, 562-580. doi: 10.1007/s10278-009-9240-6

Camilus, K.S., Govindan, V.K., Sathidevi, P.S., 2011. Pectoral muscle identification in mammograms. Journal of Applied Clinical Medical Physics 12, 215-230. doi: 10.1120/jacmp.v12i3.3285

Canny, J., 1986. A computational approach to edge detection. IEEE Transaction on Pattern Analysis and Machine Intelligence 8, 679-698. doi: 10.1109/TPAMI.1986.4767851

Caputo, B., Hayman, E., Mallikarjuna, P., 2005. Class-specific material categorization. Tenth IEEE International Conference on Computer Vision (ICCV'05) vol. 1, 1597-1604. doi: 10.1109/ICCV.2005.54

Carson, P.L., LeCarpentier, G.L., Roubidoux, M.A., Erkamp, R.Q., Fowlkes, J.B., Goodsitt, M.M., 2004. Physics and technology of breast cancer US imaging including automated three-dimensional US. RSNA Categorical Course in Diagnostic Radiology Physics: Advances in Breast Imaging-Physics, Technology, and Clinical

Applications- A. Karellas and M. L. Giger, eds. (RSNA, Oak Brook, IL, 2004), pp. 223-232

Cascio, D., Fauci, F., Iacomi, M., Raso, G., Magro, R., Castrogiovanni, D., Filosto, G., Lenzi, R., Vasile, M.S., 2014. Computer-aided diagnosis in digital mammography: comparison of two commercial systems. Imaging Med. 6, 13-20. doi:10.2217/IIM.13.68

Castelli, E., Tonutti, M., Arfelli, F., Longo, R., Quaia, E., Rigon, L., Daniela, S., Zanconati, F., Dreossi, D., Abrami, A., Quai, E., Bregant, P., Casarin, K., Chenda, V., Menk, R.H., Rokvic, T., Vascotto, A., Tromba, G., Cova, M.A., 2011. Mammography with synchrotron radiation: first clinical experience with phase-detection technique. Radiology 259. doi: 10.1148/radiol.11100745

Cernads, E., Gomez, L., Rodriguez, P.G., Casas, A., Carrio, R.G., Vidal, J.J., 1996. Design of unsharp masking filters in the frequency domain: Parameterization for breast radiographs. In Doi, K., Giger, M.L., Nishikawa, R.M., Schmidt, R.A., editors, Digital Mammography, 463-466. Elsevier, Amsterdam

Chaabani, A., Boujelben, A., Mahfoudhi, A., Abid, M., 2010. An Automatic Pre-processing Method for Mammographic Images. International Journal of Digital Content Technology and its Applications 4, 190-201, doi: 10.4156/jdcta.vol4.issue3.19

Chaki, N., Shaikh, S.H., Saeed, K., 2014. Exploring image binarization techniques. Studies in Computational Intelligence, Springer. doi: 10.1007/978-322-1907-1-2.

Chakraborty, D.P., 2013. A brief history of FROC paradigm data analysis. Acad Radiol. 20, 915-919. doi: 10.1016/j.acra.2013.03.001

Chan, H.P., Vyborny, C.J., McaMahon, H., Metz, C.E., Doi, K., Sickles, E.A., 1987. Digital mammography: ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. Invest Radiol. 22, 581-589.

Chan, H.P., Sahiner, B., Lam, K.L., Petrick, N., Helvie, M.A., Goodsitt, M.M., Adler, D.D., 1998. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. Medical Physics 25, 2007-2019. doi: 10.1118/1.598389

Chan, H.P., Doi, K., Vyborny C.J., Schmidt, R.A., Metz, C.E., Lam, K.L., Ogura, T., Wu, Y., MacMahon, H., 1990. Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis. Invest Radiol. 25, 1102-1110. doi: 10.1097/00004424-199010000-00006

Chandrasekhar, R., and Attikiouzel, Y., 1998. The need to standardize and calibrate databases of digitized mammograms. Digital Mammography: Nijmegen, 1998. Jan Hendriks and Leon van Erning (Eds), p. 503-504.

Chen, C.H., Lee, G.G., 1997. On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis. Graphical Models and Image Processing 59. 349-364. doi: 10.1006/gmip.1977.0443

Chen, H., Danielsson, M., Xu, C., Cederstrom, B., 2015. On image quality metrics and the usefulness of grids in digital mammography. J Med Imaging 2. doi: 10.1117/1.JMI.2.1.013501

Cheng, H.D., Lui, Y.M., Freimanis, R.I., 1998. A novel approach to microcalcification detection using fuzzy logic technique. IEEE Trans. Med. Imaging 17, 442-450. doi: 10.1109/42.712133

Cheng, H.D., Cai, X., Chen, X., Hu, L., Lou, X., 2003. Computer aided detection and classification of microcalcifications in mammograms: A survey. Pattern Recognition 36, 2967-2991. doi: 10.1016/s0031-3203(03)00192-4

Cheng, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P., Du, H.N., 2006. Approaches for automated detection and classification of masses in mammograms. Pattern Recognition 39, 646-668. doi: 10.1016/j.patcog.2005.07.006

Chérel, P., Hagay, C., Benaim, B., De Maulmont, C., Engerand, S., Langer, A., Talma, V., 2008. Mammographic evaluation of dense breasts: techniques and limits. Journal Radiology 89, 1156-1168. doi: JR-09-2008-89-9-C2-0221-0363-101019-200805527

Chokunonga, E., Borok, M. Z., Chirenje, Z. M., Nyakabau, A. M., Parkin,, D. M., 2013. Trends in the incidence of cancer in the black population of Harare, Zimbabwe 1991-2010, International Journal of Cancer 133, 721-729. doi: 10.1002/ijc.28063

Christ, M.C., Parvathi, R.M., 2011. Fuzzy c-means algorithm for medical image segmentation. IEEE, 3rd International Conference on Electronics Computer Technology. doi: 10.1109/ICECTECH.2011.5941851.

Ciatto, S., Bernardi, D., Calabrese, M., Durando, M., Gentilini, M.A., Mariscotti, G., Monetti, F., Moriconi, E., Pesce, B., Roselli, A., Stevanin, C., Tapparelli, M., Houssami, N., 2012. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. Breast 21, 503-506. doi: 10.1016/j.breast.2012.01.005

Ciccone, G., Vineis, P., Frigerio, A., Segnan, N., 1992. Inter-observer and intra-observer variability of mammogram interpretation: a field study. European Journal of Cancer 28, 1054-1058. doi: 10.1016/0959-8049(92)90455-B

Ciecholewski, M., 2017. Malignant and benign mass segmentation in mammogram using active contour methods. Symmetry 9, doi: 10.3390/sym9110277

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F., 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 26, 1045-1057. doi: 10.1007/s10278-013-9622-7

Constantinidis, A.S., Fairhust, M.C., Deravi, F., Hanson, M., Wells, C.P., Chapman-Jones, C., 1999. Evaluating classification strategies for detection of circumscribed masses in digital mammograms. 7th International Conference on Image Processing and Its Application, 435-439. doi: 10.1049/cp:19990359

Constantinidis, A.S., Fairhust, M.C., and Rahman, A.F.R., 2000. Detection of circumscribed masses in digital mammograms using behavior-knowledge space method. Electronic Letters 36, 302-303. doi: 10.1049/el:20000280

Corbex, M., Bouzbid, S., and Boffetta, P., 2014. Features of breast cancer in developing countries: examples from North-Africa. European Journal of Cancer 50, 1808-1818. doi: 10.1016/j.ejca.2014.03.016

Cortes, C., Vapnik, V., 1995. Support vector machine. Machine Learning 20, 273-297. doi: 10.1023/A:1022627411411

Costa, A. F., Tekli, J., Traina, A. J. M., 2011. Fast fractal stack: fractal analysis of computed tomography scan of the lung. International ACM Workshop on Medical Multimedia Analysis and Retrieval, Scottsdale, AZ, USA: ACM, pp. 13-18. doi: 10.1145/2072545.2072549

Costa, A. F., H-Mamani, G., Traina, A.J.M., 2012. An efficient algorithm for fractal analysis of textures. 25th SIBGRAP IEEE Conference, 36-46. doi: 10.1109/SIBGRAPI.2012.15

Costa, D.D., Campos, L.F., Barros, A.K., Silva, A.C., 2007. Independent component analysis in breast tissue mammograms images classification using LDA and SVM. 6th International Special Topic Conference on Information Technology Applications in Biomedicine, IEEE. doi: 10.1109/ITAB.2007.4407389

Cura, J.L., Elizagaray, E., Zabala, R., Legorburu, A., Grande, D., 2005. The use of unenhanced Doppler sonography in the evaluation of solid breast lesions. AJR AM J Roentgenol. 184, 1788-1794. doi: 10.2214/ajr.184.6.01841788

D'Orsi, C.J., 1998. Illustrated breast imaging reporting and data system: illustrated BI-RADS. 3rd ed. Reston, American College of Radiology

D'Orsi, C.J., Getty, D.J., Pickett, R.M., Sechopoulos, I., Newell, M.S., Gundry, K.R., Bates, S.R., Nishikawa, R.M., Sickles, E.A., Karellas, A., D'Orsi, E.M., 2013. Stereoscopic digital mammography: improved specificity and reduced rate of recall in a prospective clinical trial. Radiology 266, 81-88. doi: 10.1148/radiol.12120382

D'Orsi, C.J., Mendelson, E.B., Ikeda, D.M., 2013. Breast Imaging Reporting and Data system: ACR BI-RADS, 5th ed. Reston, American College of Radiology

Dehghani, S., Dezfooli, M.A., 2011. A method for improve preprocessing images mammography. International Journal of Information and Education Technology 1, 90-93. doi: 10.7763/IJIET.2011.V1.15

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum-likelihood from incomplete data via EM algorithm. Journal of the Royal Statistical Society: Series B, 39, 1-38. doi: 10.1.1.133.4884

Devi, A., Mini, M.G., 2015. Mammographic image enhancement based on SWT and high boost filtering. International Journal of Computer Theory and Engineering 7, 374-378. doi: 10.7763/IJCTE.2015.V7.988

Dheeba, J., Selvi, S.T., 2010. Screening mammogram images for abnormalities using radial basis function neural network. International Conference on Communication Control and Computing Technologies, IEEE. doi: 10.1109/ICCCCT.2010.5670778

Diamant, I., Greenspan, H., Goldberger, J., 2012. Breast tissue classification in mammograms using visual words. IEEE 27th Convention of Electrical nd Electronics Engineers in Israel, 1-4. doi: 10.1109/EEEI.2012.6377061

DICOM Standard Committee, 2004. Supplement 79: breast imaging report templates. Digital Imaging and Communications in Medicine (DICOM)

Doi, K., 2007. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Comput Med Imaging Graph. 31, 198-211. doi: 10.1016/j.compmedimag.2007.02.002

Dominguez, A.R., Nandi, A.K., 2007. Enhanced multi-level thresholding segmentation and rank based region selection for detection of masses in mammograms. IEEE Conference on Acoustics, Speech and Signal Processing. ICASSP'07, 449-452. doi: 10.1109/ICASSP.2007.366713.

Don, S., Chung, D., Revathy, K., Choi, E., Min, D., 2012. A new approach for mammogram image classification using fractal properties. Cybernetics and Information Technologies 12. doi: 10.2478/cait-2012-0013

Dromain, C., Boyer, B., Ferré, R., Canale, S., Delaloge, S., Balleyguier, C., 2013. Computed-aided diagnosis (CAD) in the detection of breast cancer. European Journal of Radiology 82, 417-423. doi: 10.1016.j.ejrad.2012.03.005

Duda, R.O., Hart, P.E., 1972. Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM 15, 11-15. doi: 10.1145/361237.361242

Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification. 2nd edition. New York, NY, USA, John Wiley & Sons.

Durate, M.A., Alvarenga, A.V., Azevedo, C.M., Infantosi, A.F., Pereira, W.C., 2011. Automatic microcalcifications segmentation procedure based on Otsu's method and morphological filters. IEEE, Pan American Health Care Exchanges, pp. 102-106. doi: 10.1109/PAHCE.2011.5871858

Egan, R.L., 1960. Experience with mammography in a tumor institution: evaluation of 1000 studies. Radiology 75, 894-900. doi: 10.1148/74.6.894

Elgaili, E.M., Abuidris, D.O., Rahman, M., Michalek, A.M., Mohammed, S.I., 2010. Breast cancer burden in central Sudan. International Journal of Women's Health 2, 77-82. doi: 10.2147/IJWH.S8447

Elmore, J.G., Jackson, S.L., Abraham, L., Miglioretive, D.L., Carney., P.A., Geller, B.M., Yankaskas, B.C., Kerlikowske, K., Onega, T., Rosenberg, R.D., Sickles, E.A., Buist, D.S., 2009. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. Radiology 253, 641-651. doi: 10.1148/radiol.2533082308

Elmore, J.G., Amstrong, K., Lehman, C., Fletcher, S.W., 2005. Screening for breast cancer. JAMA. 293, 1245-1256. doi: 10.1001/jama.293.10.1245

Elston, C.W., Ellis, I.O., 1991. Pathological prognostic factors in breast cancer I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19, 403-410, doi: 10.1111/j.1365-2559.1991.tb00229.x

Erickson, S.J., Godavarty, A., Martinez, S.L., Gonzalez, J., Romero, A., Roman, M., Nunez, A., Ge, J., Regalado, S., Kiszonas, R., Penalver, C., 2011. Hand-held optical devices for breast cancer: spectroscopy and 3D tomographic imaging. IEEE J of Selected Topics in Quantum Electronics 18, 1298-1312. doi: 10.1109/JSTQE.2011.2170664

Evans, C., Yates, K., Brady, M., 2003. Statistical characterization of normal curvilinear structures in mammograms. Peitgen HO. (eds) Digital Mammography, 285-291. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-59327-7_68

Eyk, Ng., 2009. A review of thermography as promising non-invasive detection modality for breast tumor. Int J Therm Sci. 48, 849-859. doi: 10.1016/j.ijthermalsci.2008.06.015

Fenton, J.J., Taplin, S.H., Carney, P.A., Abraham, L., Sickles, E.A., D'Orsi, C., Berns, E.A., Cutter, G., Hendrick, R.E., Barlow, W.E., Elmore, J.G., 2007. Influence of computer-aided detection on performance of screening mammography. N Engl J Med. 356, 1399-1409. doi: 10.1056/NEJMoa066099

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, DM., Forman, D., Bray, F., 2015. Cancer incidence and mortality worldwide: source, methods and major patterns in GLOBOCAN 2012. , Int J Cancer 136, 359-386. doi: 10.1002/ijc.29210.

Ferrari, R.J., Rangayyan, R.M., Desautels, J.E., Borges, R.A., Frere, A.F., 2000. Segmentation of mammograms: identification of the skin-air boundary, pectoral muscle, and fibro-glandular disc. 5th International Workshop on Digital Mammography, 573-579. YAFFE, M. J. (Ed.)

Ferrari, R.J., Rangayyan, R.M., Desautels, J.E., Borges, R.A., Frere, A.F., 2004. Automatic identification of the pectoral muscle in mammograms", IEEE Trans. on Medical Imaging 23, 232-245. doi: 10.1109/TMI.2003.823062

Franquet, T., De Miguel, C., Cozcolluela, R., Donoso, L., 1993. Spiculated lesions of the breast: Mammographic- pathologic correlation. RadioGraphics 13, 841-852. doi: 10.1148/radiographics.13.4.8356272

Freer, P.E., 2015. Mammographic breast density: impact on breast cancer risk and implications for screening. RadioGraphics 35, 302-315. doi: 10.1148/rg.352140106

Fritz, A., Percy, C., Jack, A., Shanmugaratnam, K., Sobin, L., Parkin, D.M., Whelan, S., 2013. International Classification of Diseases for Oncology, Third edition. WHO 2013. Geneva, Switzerland

Galloway, M.M., 1975. Texture analysis using gray level run lengths. Computer Graphic and Image Processing 4, 172-179. doi: 10.1016/S0146-664X(75)80008-6

Garcia-Closas, M., Brinton, L.A., Lissowska, J., Chatterjee, N., Peplonska, B., Anderson, W.F., Szeszenia, N., Bardin, A., 2006. Established breast cancer risk factors by

clinically important tumor characteristics. British Journal of Cancer 95, 123-129. doi: 10.1038/sj.bjc.6603207

Garra, B.S., Krasner, B.H., Horii, S.C., Ascher, S., Mun, S.K., Zeman, R.K., 1993. Improving the distinction between benign and malignant breast lesion: The value of sonographic texture analysis. Ultrasound Imaging 15, 267-285. doi: 10.1177/016173469301500401

Gaur, Sh., Dialani, V., Slanetz, P. J., Eisenberg, R. L., 2013. Architectural Distortion of the breast. American Journal of Roentgenology 201, 662-670. doi: 10.2214/AJR.12.10153

Ge, J., Hadjiiski, L.M., Sahiner, B., Wei, J., Helvie, C.Z., Chan, H.P., 2007. Computer aided detection system for clustered microcalcifications: comparison of performance on full field digital mammograms and digitized screen-film mammograms. Phys Med Biol. 52, 981-1000. doi: 10.1088/0031-9155/52/4/008

Ghoncheh, M., Pournamdar, Z., Salehiniya, H., 2016. Incidence and mortality and epidemiology of breast cancer in the world. Asian Pac J Cancer Prev. 17, 43-46. doi: 10.7314/APJCP.2016.17.S3.43

Giger, M.L., 2000. Computer-aided diagnosis of breast lesions in medical images. IEEE, Computing in Science & Engineering 2, 39-45. doi: 10.1109/5992.877391

Giger, M.L., Huo, Z., Kupinski, M.A., Vyborny, C.J., 2000. Computer-aided diagnosis in mammography. In: M. Sonka, JM. Fitzpatrick, Handbook of Medical Imaging. SPIE, Bellingham, 2, 915-1004.

Giger M.L., Chan, H.P., Boone, J., 2008. History and status of CAD and quantitative image analysis: the role of medical physics and AAPM. Med Phys. 35, 5799- 5820. doi: 10.1118/1.3013555

Gilbert, F.J., Astley, S.M., McGee, M.A., Gillan, M.G., Boggis, C.R., Griffiths, P.M., Duffy, S.W., 2006. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. Radiology 241, 47-53. doi: 10.1148/radiol.2411051092

Glasbey, C.A., 1993. An analysis of histogram based thresholding algorithms. CVGIP: Graphical models and image processing 55, 532-537. doi: 10.1006/cgip.1993.1040

Godavarty, A., Rodriguez, S., Jung, Y., Gonzalez, S., 2015. Optical imaging for breast cancer prescreening. Breast Cancer 7, 193-209. doi: 10.2147/BCTT.S51702

Gold, L.S., Klein, G., Carr, L., Kessler, L., Sullivan, S.D., 2012. The emergence of diagnostic imaging technologies in breast cancer: discovery, regulatory approval, reimbursement, and adoption in clinical guidelines. Cancer Imaging 12, 13-24. doi: 10.1102/1470-7330.2012.0003

Gonzalez, R. C., and Woods, R. E., 1992. Digital Image Processing, Reading, MA: Addison Wesley

Gordon, R., Rangayyan, R.M., 1984. Feature enhancement of film mammograms using fixed and adaptive neighborhood. Applied Optics 23, 560-564. doi: 10.1364/AO.23.000560

Gottlieb, S., 2000. Lumpectomy as good as mastectomy for tumor up 5 cm across. Western Journal of Medicine 173, 227-228. doi: 10.1136/bmj.321.7256.261

Guliato, D., Rangayyan, R.M., Carnielli, W.A., Zuffo, J.A., Desautels, J.E.L., 1998. Segmentation of breast tumors in mammograms by fuzzy region growing. In Proc. International Conference IEEE Engineering in Medicine and Biology Society 20, 1002-1005. doi: 10.1109/1EMBS.1998.745618

Guo, Q., Shao, J., Ruiz, V., 2009. Characterization and classifications of tumor lesions using computerized fractal based texture analysis and support vector machines in digital mammograms. Int. J. Comput. Assist. Radiol. Surg. 4, 11-25. doi: 10.1007/s11548-008-0276-8

Guohong, L., Wenming, G., 2010. Application of improved arithmetic of median filtering denoising. Computer Engineering and applications 46, 187-189.

Gupta, M.R., Chen, Y., 2011. Theory and use of the EM algorithm. Foundations and Trends in Signal Processing. 4, 223-296. doi: 10.1561/2000000034

Gur, D., Abrams, G.S., Chough, D.M., Ganott, M.A., Hakim, C.M., Perrin, R.L., Rathfon, G.Y., Sumkin, J.H., Zuley, M.L., Bandos, A.L., 2009. Digital breast tomosynthesis: observer performance study. AJR AM J Roentgenol 193, 586-591. doi: 10.2214/AJR.08.2031

Hadjiiski, L. M., Sahiner, B., Chan, H.P., 2006. Advances in CAD for diagnosis of breast cancer. Current Opinion in Obstetrics & Gynecology 18, 64-70. doi: 10.1097/01.gco.0000192965.29449.da

Hamerly, G., Elkan, C., 2002. Alternative to the K-means algorithm that find better clustering. CIKM '02 Proceedings of the 11th International Conference on information and knowledge management, p. 600-607. doi: 10.1145/584792.584890

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification, IEEE Trans. System Man Cybernetics 3, 610-621. doi: 10.1109/TSMC.1973.4309314

Haralick, R., Shapiro, L., 1992. Computer and Robot Vision, vol. 1. Addison-Wesley Publishing Company.

Heath, M.D., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K., Munishkumaran, S., 1998. Current status of the digital database for screening mammography. Digital Mammography, 457-460. Computational Imaging and Vision 13. Springer, Dordrecht. doi: 10.1007/978-94-011-5318-8_75

Helvie, M. A., Chan, H.P., Adler, D.D., Boyd, P.G., 1994. Breast thickness in routine mammograms: effect on image quality and radiation dose. Journal of Roentgenology 163, 1371-1374. doi: 10.2214/ajr.163.6.7992731

Hendrick, R.E., 2010. Radiation doses and cancer risks from breast imaging studies. Radiology 257, 246-253. doi: 10.1148/radiol.10100570

Hmida, M., Hamrouni, K., Solaiman, B., Boussetta, S., 2017. An efficient method for breast mass segmentation and classification in mammographic images. International Journal of Advanced Computer Science and Applications 8, 256-262. doi: 10.14569/IJACSA.2017.081134

Holland, K., Sechopoulos, I., Mann, R., Heeten, G., Gils, C., Karssemeijier, N., 2017. Influence of breast compression pressure on the performance of population-based mammography screening. Breast Cancer Res. 19, doi: 10.1186/s13058-017-0917-3

Hough, P.V.C., 1962. Method and means for recognizing complex patterns, US Patent 3969654.

Houssami, N., Given-Wilson, R., Ciatto, S., 2009. Early detection of breast cancer: overview of the evidence on computer aided detection in mammography screening. J Med Imaging Radiat Oncol. 53, 171-176. doi: 10.1111/j.1754-9485.2009.02062.x

Hu, J., Peng, X., Xu, Z., 2012. Study of gray image pseudo-color processing algorithms. Proc. SPIE vol. 8415, 6th International symposium on Advanced Optical Manufacturing and Testing Technologies. doi: 10.1117/12.977197

Huang, T. S., Yang, G. J., Tang, G.Y., 1979. A fast two dimensional median filtering. IEEE Transaction on Acoustics Speech, and Signal Processing 27, 13-18. doi: 10.1109/TASSP.1979.1163188

Huda, W., Abrahams, R.B., 2015. Radiographic techniques, contrast, and noise in X-ray imaging. American Journal of Roentgenology 204, 126-131. doi: 10.2214/AJR.14.13116

Huo, Z., Giger, M.L., Vyborny, C.J., Bick, U., Wolverton, D.E., Schmidt, R.A., 1995. Analysis of speculation in the computerized classification of mammographic masses. Med. Phys. 22, 1569-1579. doi: 10.1118/1.597626

Hussein, R., Engelmann, U., Schroeter, A., Meinzer, H., 2004. DICOM structured reporting. RadioGraphics 24. doi: 10.1148/rg.243035710

Intisar, E.S., Hsin-Yi, W., Kamal, H.M., Sulma, I.M., 2014. Cancer incidence in Khartoum, Sudan: first results from the Cancer Registry, 2009-2010. Cancer Medicine 3, 1075-1084. doi: 10.1002/cam4.254

Jasmine, J.L., Govardhan, A., Baskaran, S., 2009. Microcalcification detection in digital mammograms based on wavelet analysis and neural networks. International Conference on Control, Automation, Communication and Energy Conversation (INCACEC). IEEE 2009, pp 1-6.

Jeffreys, M., Harvey, J., Highnam, R., 2010. Comparing a new volumetric breast density method (Volpara ^{TM}) to cumulus. International Workshop on Digital Mammography (IWDM), 408-413. doi: 10.1007/978-3-642-13666-5_55

Juarez, C., Ponomaryov, V., Sanchez, J.L., 2006. Detection of microcalcifications in digital mammograms images using wavelet transform. Electronics, Robotics and Automotive Mechanics Conference (CERMA'06), IEEE. doi: 10.1109/CERMA.2006.36

Kabbadj, Y., Regragui, F., Himmi, M.M., 2012. Microcalcification detection using a fuzzy inference system and support vector machines. International Conference on Multimedia Computing and Systems, IEEE. doi: 10.1109/ICMCS.2012.6320216

Kalaf, J.M., 2014. Mammography: a history of success and scientific enthusiasm. Radiol Bras. 47. doi: 10.1590/0100-3984.2014.47.4e2

Kallergi, M., Woods, K., Clarke, L.P., Qian, W., Clark, R.A., 1992. Image segmentation in digital mammography: Comparison of local thresholding and region growing algorithms. Computerized Medical Imaging and Graphics 16, 323-331. doi: 10.1016/0895-6111(92)90145-Y

Kallergi, M., 2004. Computer-aided diagnosis of mammographic microcalcification clusters. Med. Phys. 31, 314-326. doi: 10.1118/1.1637972

Kaplan, S.S., 2014. Automated whole breast ultrasound. Radiol Clin North Am. 52, 539-546. doi: 10.1016/j.rcl.2014.01.002

Karellas, A., Vedantham, S., 2008. Breast cancer imaging: a perspective for the next decade. Med Phys. 35, 4878-4897. doi: 10.1118/1.2986144

Karssemeijer, N., 1993. Recognition of clustered microcalcifications using a random field model. In Proc. of SPIE. doi: 10.1117/12.148689

Karssemeijer, N., Frieling, J.T., Hendriks, JH., 1993. Spatial resolution in digital mammography. Invest Radiol. 28, 413-419. doi: 10.1097/00004424-199305000-00005

Karssemeijer, N., 1998. Automated classification of parenchymal patterns in mammograms. Phys Med Biol. 43, 365-378. doi: 10.1088/0031-9155/43/2/011

Kaur, P., Singh, G., Kaur, Pa., 2019. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. Informatics in Medicine Unlocked, Elsevier. doi: 10.1016/j.imu.2019.01.001

Ke, L., He, W., Kang, Y., 2009. Mass auto-detection in mammogram based on wavelet transform modulus maximum. Conf Proc IEEE Eng Med Biol Soc. 5760-3. doi: 10.1109/IEMBS.2009.5332615

Kegelmeyer, W., Pruneda, J., Bourland, P., Hillis, A., Riggs, M., Nipper, M., 1994. Computer-aided mammographic screening for spiculated lesion. Radiology 191, 331-337. doi: 10.1148/radiology.191.2.8153302

Keller, B., Nathan, D., Wang, Y., Zheng, Y., Gee, J., Conant, E., Kontos, D., 2011. Adaptive multi cluster fuzzy c-means segmentation of breast parenchymal tissue in digital mammography. Med Image Comput Comput Assist Interv. 14, 562-569. PMID: 22003744

Kim, S.H., Lee, E.H., Jun, J.K., Kim, Y.M., Chang, Y-W., Lee, J.H., Kim, H-W., Choi, E.J., (ABCS-K), 2019. Interpretive performance and inter-observer agreement on digital mammography test sets. Korean J Radiol. 20, 218-224. doi: 10.3348/kjr.2018.0193

Kobatake, H., Yoshinaga, Y., 1996. Detection of spicules on mammogram based on skeleton analysis. IEEE Trans. on Medical Imaging 15, 235-245. doi: 10.1109/42.500062

Kom, G., Tiedeu, A., Kom, M., 2007. Automated detection of masses in mammograms by local adaptive thresholding. Comput. in Biol. and Medicine 37, 37-48. doi: 10.1016/j.compbiomed.2005.12.004

Kopans, D.B., 1998. Breast Imaging, 2nd ed. Lippincott-Raven, Philadelphia.

Kopans, D.B., 2014. A new era in mammography screening. Radiology 271, 629-631. doi: 10.1148/radiol.14140177

Kundel, H. L., Revesz, G., 1976. Lesion conspicuity, structured noise, and film reader error. American Journal of Roentgenology 126, 1233-1238. doi: 10.2214/ajr.126.6.1233

Kupinski, M.A., Giger, M.L., 1998. Automated seed lesion segmentation on digital mammograms. IEEE Transaction on Medical Imaging 17, 510-517. doi: 10.1109/42.730396

Lacey, J.V., Kreimer, A.R., Buys, S.S., Marcus, P.M., Chang, SC., Leitzmann, M.F., Hoover, R.N., Prorok, P.C., Berg, C.D., Hartage, P., 2009. Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort. BMC Cancer 9. doi: 10.1186/1471-2407-9-84

Lai, S. M., Li, X., Biscof, W.F., 1989. On techniques for detecting circumscribed masses in mammograms. IEEE Transactions on Medical Imaging 8, 377-386. doi: 10.1109/42.41491

Langarizadeh, M., Mahmud, R., Ramli, A.R., Napis, S., Beikzadeh, M.R., Rahman, W.E., 2011. Improvement of digital mammogram images using histogram equalization, histogram stretching and median filter. J Med Eng Technol. 35, 103-108. doi: 10.3109/03091902.2010.542271.

Lawrence, W., Bassett, Conner, K., and MS, IV., 2003. The abnormal mammogram. Holland-Frei Cancer Medicine. 6th edition. Kufe DW, Pollock RE, Weichselbaum RR, et al., Editors. Hamitton (ON): BC Decker; 2003

Lazebnik, S., Schmid, C., Ponce, J., 2005. A sparse texture representation using local affine regions. Pattern Analysis and Machine Intelligence 27, 1265-1278. doi: 10.1109/TPAMI.2005.151

Lee, C.H., Dershaw, D.D., Kopans, D., Evans, P., Monsees, B., Monticciolo, D., Brenner, R.,J., Bassett, L., Berg, W., Feig, S., Hendrick, E., Mendelson, E., D'Orsi, C., Sickles, E., Burhenne, L.W., 2010. Breast cancer screening with imaging : recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer, Journal of American College of Radiology 7, 18-27. doi: 10.1016/j.jacr.2009.09.022

Lee, C.S., Bhargavan-Chatfield, M., Burnside, E.S., Nagy, P., Sickles, E.A., 2016. The national mammography database: preliminary data. AJR Am J Roentgenol. 206, 883-890. doi: 10.2214/AJR.15.14312

Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L., 2017. Acurated mammography data set for use in computer-aided detection and diagnosis research. Sci. Data 4:170177. doi: 10.1038/sdata .2017.177

Lee, Y.J., Park, J.M., Park, H.W., 2000. Mammographic mass detection by adaptive thresholding and region growing. Int. J. Imaging Systems Technol. 11, 340-346. doi: 10.1002/ima.1018

Leon, S., Brateman, L., Honeyman-Buck, J., Marshall, J., 2009. Comparison of two commercial CAD systems for digital mammography. Journal of Digital Imaging 22, 421-423. doi: 10.1007/s10278-008-9144-x

Leproux, A., Beek, M., Vries, U., Wasser, M., 2010. Automated 3D whole-breast ultrasound imaging: results of a clinical pilot study. Proceedings of SPIE- The International Society for Optical Engineering 7629. doi: 10.117/12.840391

Lewin, J.M., Hendrick, R.E., D'orsi, C.J., Isaacs, P.K., Moss, L.J., Karellas, A., Sisney, G.A., Kuni, C.C., Cutter, G.R., 2001. Comparison of full-field digital mammography with screen film mammography for cancer detection: results of 4,945 paired examinations. Radiology 218, 873-880. doi: 10.1148/radiology.218.3.r01mr29873

Lewin, J.M., D'Orsi, C.J., Hendrick, R.E., Moss, L.J., Isaacs, P.K., Karellas, A., Cutter, G.R., 2002. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. American Journal of Roentgenology 179, 671-677. doi: 10.2214/ajr.179.3.1790671

Li, H., Liu, K.J., Lo, S., 1997. Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammogram. IEEE Trans. Med. Imaging 16, 785-798. doi: 10.1109/42.650875

Li, L., Mao, F., Qian, W., Clarke, L.P., 1997. Wavelet transform for directional feature extraction in medical imaging. Proceedings of International Conference on Image Processing 3, IEEE. doi: 10.1109/ICIP.1997.632167

Li, L., Clark, R.A., Thomas, J.A., 2002. Computer-aided diagnosis of masses with full-field digital mammography. Academic Radiology 9, 4-12. doi: 10.1016/S1076-6332(03)80290-8

Li, Q., Nishikawa, R.M., 2015. Computer-aided detection and diagnosis in medical imaging. CRC Press, Taylor & Francis Group.

Li, Y., Chen, H., Yang, Y., Yang, N., 2013. Pectoral muscle segmentation in mammograms based on homogeneous texture and intensity deviation. Pattern Recognition 46, 681-691. doi: 10.1016/j.patcog.2012.09.021

Linguraru, M., Jan. 2002. Feature detection in mammographic image analysis. Ph.D. thesis, University of Oxford

Lipkin, B.S., Rosenfeld, A., 1970. Picture processing and psychopictorics. Academic Press Inc., New York, Library of Congress Catalog Card Number: 71, 127691

Liu, X., Li, B., Liu, J., Xu, X., Feng, Z., 2012. Mass diagnosis in mammography with mutual information based feature selection and support vector machine. Intelligent Computing Theories and Application, 1-8. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-31576-3_1

Luo, P., Qian, W., Romilly, P., 2005. CAD-aided mammogram training. Acad Radiol. 12, 1039-1048. doi: 10.1016/j.acra.2005.04.011

Ma, F., Bajger, M., Slavotinek, J.P., Bottema, M.J., 2007. Two graph theory based methods for identifying the pectoral muscle in mammograms. Pattern Recognition 40, 2592-2602. doi: 10.1016/j.patcog.2006.12.011

MacQueen, J.B., 1967. Some methods of classification and analysis of multivariate observations. In Proc. of the 5th Berkeley Symposium on Mathematical Statistics and

Probability 1, 281-297. University of California Press, Berkeley, Calif., 1967. http: //Projecteuclid.org/Euclid.bsmsp/1200512992

Maitra, I. K., Nag, S., Bandyopadhyay, S.K., 2012. Technique for preprocessing of digital mammogram. Comput Methods Programs Biomed. 107, 175-188. doi: 10.1016/j.cmpb.2011.05.007

Makandar, A., Halalli, B., 2016. Threshold based segmentation technique for mass detection in mammography. Journals of Computers 11, 472-478. doi: 10.17706/jcp.11.6.472-478

Malone, K.E., Daling, J.R., Thompson, J.D., O'Brien, C.A., Franciso, L.V., Ostrander, E.A., 1998. BRCA1 mutations and breast cancer in the general population: analysis in women before age 35 years and in women before age 45 years with first degree family history. Journal of the American Medical Association (JAMA), 279, 922-929. doi: 10.1001/jama.279.12.922

Mandelbrot, B.B., "Fractal geometry of nature", Freeman, New York, 1982

Marti, R., Zwiggelaar, R., Rubin, C.M.E., 2001. Tracking mammographic structures over time. In Proc. British Machine Vision Conference, 143-152. doi: 10.1.1.566.985

Marti, R., Raba, D., Oliver, A., Zwiggelaar, R., 2006. Mammographic Registration: Proposal and Evaluation of a new approach. IWDM'06 Proceedings of the 8th International Conference on Digital Mammography, 213-220. doi: 10.1007/11783237_30

Martins, L.O., Junior, G.B., Silva, A.C., Paiva, A.C., Gattass, M., 2009. Detection of masses in digital mammograms using K-Means and support vector machine. Electronic Letters on Computer Vision and Image Analysis: ELCVIA 8. doi: 10.5565/rev/elcvia.216

Marzulli, V.M., 1999. The CALMA project: computer assisted library for mammography. Neural Nets WIRN VIETRI-98, 230-235. Perspective in Neural Computing, Springer, London. doi: 10.1007/978-1-4471

Masek, M., Chandrasekhar, R., deSilva, C.J.S., Attikiouzel, Y., 2001. Spatially based application of the minimum cross entropy thresholding algorithm to segment the pectoral muscle in mammograms. IEEE, The Seventh Australian and New Zealand Intelligent Information Systems Conference, 101-106. doi: 10.1109/ANZIIS.2001.974058

Matheus, B.R., Schiabel, H., 2011. Online mammographic images database and comparison of CAD schemes. Journal of Digital Imaging 24, 500-506. doi: 10.1007/s10278-010-9297-2

Matsubara, T., Fujita, H., Kasai, S., Goto, M., Tani, Y., Hara, T., Endo, T., 1997. Development of new schemes for detection and analysis of mammographic masses. Proceeding Intelligent Information Systems, IIS'97, 63-66. doi: 10.1109/IIS.1997.645180

McKenzie, P., and Alder, M., 1994. Initializing the EM algorithm for use in Gaussian mixture modeling. Machine Intelligence and Pattern Recognition 16, 91-105. doi: 10.1016/B978-0-444-81892-8.50013-4

Méndez, A.J., Souto, M., Tahoces, P. G., and Vidal, J. J., 2003. Computer aided diagnosis for breast mass detection on a telemammography system. Computerized Medical Imaging and Graphics 27, 497-502. doi: 10.1016/S0895-6111(03)00035-1

Mini, M.G., Devassia, V.P., Thomas, T., 2004. Multiplexed wavelet transform technique for detection of microcalcification in digitized mammograms. J Digit Imaging 17, 285-291. doi: 10.1007/s10278-004-1020-8

Mohan, S., and Ravishankar, M., 2013. Modified contrast limited adaptive histogram equalization based on local contrast enhancement for mammogram images. In: Das V.V., Chaba Y. (eds) Mobile Communication and Power Engineering. AIM 2012.

Communications in Computer and Information Science 296. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-35864-7_60.
Moore, K.L., Agur, A. M, Dalley., F., 2010. Essential clinical anatomy, 4th edition, published by Lippincott Williams and Wilkin
Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S., 2012. INbreast: Toward a Full-field Digital Mammographic Database. Technical Report, Acad Radiol. 19, 236-248. doi: 10.1016/j.acra.2011.09.014
Morrow, M., Krontiras, H., 2001. Who should not receive chemotherapy? data from American databases and trials. JNCI Monographs 30, 109-113. doi: 10.1093/oxfordjournals.jncimonographs.a003446
Morrow, W.M., Paranjape, R.B., Rangayyan, R.M., Desautels, J.E., 1992. Region-based contrast enhancement of mammograms. IEEE Transaction on Medical Imaging 11, 392-406.
Morton, M.J., Whaley, D.H., Brandt, K.R., Amrami, K.K., 2006. Screening mammograms: Interpretation with computer-aided detection-Prospective evaluation. Radiology 239, 375-383. doi: 10.1148/radiol.2392042121
Mudigonda, N.R., Rangayyan, R.M., Desautels, J.E., 2000. Gradient and texture analysis for the classification of mammographic masses. IEEE Transactions on Medical Imaging 19, 1032-1043. doi: 10.1109/42.887618
Mudigonda N.R., Rangayyan, R.M., Desautels, J.E., 2001. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. IEEE Transactions on Medical Imaging 20, 1215-1227. doi: 10.1109/42.974917
Muhimmah, I., Oliver, A., Denton, E.R.E., Pont, J., Pérez, E., Zwiggelaar, R., 2006. Comparison between Wolfe, Boyd, BI-RADS and Tabar based mammographic risk assessment. IWDM'06 Proceedings of the 8th International Conference on Digital Mammography, 407-415. doi: 10.1007/11783237_55
Munshi, S., 2008. Scintimammography- an emerging technique for breast imaging. In: Hayat MA, editor. Lung and breast carcinomas. Elsevier. Available from: http://www.frost.com/sublib/display-market-insight.do?id=137248277
Muralidhar, G. S., Markey, M. K., and Bovik, A. C., 2001. Snakules for automatic classification of candidate spiculated mass locations on mammography. Southwest Symposium on Image Analysis & Interpretation. 197-200. doi: 10.1109/SSIAI.2010.5483885
Muralidhar, G.S., Haygood, T.M., Stephens, T.W., Whitman, G.J., Bovic, A.C., Markey, M.K., 2008. Computer-aided detection of breast cancer – Have all bases covered?. Breast Cancer (Auckl) 2, 5-9. PMC3085409
Mustra, M., Grgic, M., Delac, K., 2012. Breast density classification using multiple feature selection. Automatika 53. doi: 10.7305/automatika.53-4.281
Nailon, W. H., 2010. Texture analysis methods for medical image characterization. Biomedical Imaging, Youxin Mao, IntechOpen. doi: 10.5772/8912
Nakagawa, T., Hara, T., Fujita, H., Iwase, T., Endo, T., Horita, K., 2004. Automated contour extraction of mammographic mass shadow using an improved active contour model. International Congress Series 1268, 882-885. doi: 10.1016/j.ics.2004.03.172
Nalawade, Y, V., 2009. Evaluation of breast calcifications. Indian J Radiol Imaging 19, 282-286. doi: 10.4103/0971-3026.57208
Narod, S.A., Igbal, J., Jakubowska, A., Huzarski, T., Sun, P., Cybulski, C., Gronwald, J., Byrski, T., Lubinski, J., 2013. Are two-centimeter breast cancer large or small?. Curr Oncol. 20, 205-211. doi: 10.3747/co.20.1364
Nelson, H.D., Zakher, B., Cantor, A., Fu, R., Griffin, J., O'Meara, E.S., Buist, D., Kerlikowske, K., Ravesteyn, N.T., Dietz, A.T., Mandelblatt, J., Miglioretti,, D., 2012.

Risk factors for breast cancer for women age 40 to 49: A systematic review and meta-analysis. Annual International Med 156, 635-648. doi: 10.7326/0003-4819-156-9-201205010-00006

Netter, F. H, 2006. Atlas of human anatomy, 4th edition. Philadelphia, Pa: Elsevier, 2006.

Nixon, M., and Aguado, A., 2008. Feature extraction and image processing. 2nd edition, Academic Press in an imprint Elsevier.

Nover, A.B., Jaqtap, S., Anjum, W., Yegingil, H., Shih, W.Y., Shih, W-H., Brooks, A.D., 2009. Modern breast cancer detection: A technological review. Int J Biomed Imaging. doi: 10.1155/2009/902326

Ogundiran, T.O., Huo, D., Adenipekun, A., Campbell, O., Oyesegun, R., Akang, E., Adebamowo, C., Olopade, O.I., 2010. Case-control study of body size and breast cancer risk in Nigerian women. American Journal of Epidemiology 172, 682- 690. doi: 10.1093/aje/kwq180

Ogundiran, T. O. Huo, D., Adenipekun, A., Campbell, O., Oyesegun, R., Akang, E., Adebamowo, C., Olopade, O. I., 2012. Body fat distribution and breast cancer risk: findings from the Nigerian breast cancer study. Cancer Causes Control 23, 565-574. doi: 10.1007/s10552-012-9916-y

Okobia, M.N., Bunker, C. H., Okonofua, F.E., Osime, U., 2006. Knowledge, attitude and practice of Nigerian women towards breast cancer: a cross-sectional study. World Journal of Surgical Oncology 4, p. 11. doi: 10.1186/1477-7819-4-11

Oliveira, J.E.E., Gueld, M.O., Araujo, A.A., Ott, B., Deserno, T.M., 2008. Toward a standard reference database for computer-aided mammography. Proceedings of SPIE. 6915, Paper ID 69151Y. doi: 10.1117/12.770325

Oliver, A., Freixenet, J., Bosch, A., Raba, D., Zwiggelaar, R., 2005. Automatic classification of breast tissue. In: Marques J.S., Pérez de la Blanca N., Pina P. (eds) Pattern Recognition and Image AnalysisibPRIA 2005. Lecture Notes in Computer Csienc, 3523. Springer, Berlin, Heidelberg. doi: 10.1007/11492542_53

Oliver, A., Freixenet, J., Marti, R., Pont, J., Péréz, E., Denton, ER., Zwiggelaar, R., 2008. A novel breast tissue density classification methodology. IEEE Transactions on Information Technology in Biomedicine 12, 55-65. doi: 10.1109/TITB.2007.903514

Oliver, A., Llado, X., Péréz, E., Pont, J., Denton, E.R., Freixenet, J., Marti, J., 2010. A statistical approach for breast density segmentation", Journal of Digital Imaging 23, 527-537. doi: 10.1007/s10278-009-9217-5

Otsu, N., 1979. A threshold selection method from gray-level histogram. IEEE Trans. Syst. Man Cybern 9, 62-66. doi: 10.1109/TSMC.1979.4310076

Ozekes, S., Osman, O., Camurcu, Y., 2005. Mammographic mass detection using a mass template. Korean Journal of Radiology 6, 221-228. doi: 10.3348/kjr.2005.6.4.221

Patel, B.C., Sinha, G.R., 2010. An adaptive k-means clustering algorithm for breast image segmentation. International Journal of Computer Applications 10, 35-38. doi: 10.5120/1467-1982

Patel, M.N., Young, K., Halling-Brown, M., 2017. Optimam mammography imaging database (OMI-DB): a valuable dataset to fuel machine learning research. SIM 2017 Scientific Session, Analytics & Deep Learning Part 3.

Pavan, L.M., Oliveira, M., Alvarez, M., Sampaio, J.M., Trindade, A.P., Duarte, S.B., Pina, D.R., 2016. Breast tissue segmentation by fuzzy C-means. Physics Medica 32, p. 336. doi: 10.1016/j.ejmp.2016.07.253

Peitgen, H., Jurgen, H., Saupe, D., 1992. Chaos and fractals: new frontiers of science. Springer-Verlag, ISBN 0-387-97903-4, NewYork.

Peng, H.C., Long, F., Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. on Pattern Analysis and Machine Intelligence 27, 1226-1238. doi: 10.1109/TPAMI.2005.159

Pentland, P., 1984. Fractal-based description of natural scenes. IEEE Transaction on Pattern Analysis and Machine Intelligence 6, 661-674. doi: 10.1109/TPAMI.1984.4767591

Petrick, N., Chan, H.P., Sahiner, B., Wei, D., 1996. An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. IEEE Trans. on Medical Imaging 15, 59-67. doi: 10.1109/42.481441

Petrick, N., Chan, H.P., Sahiner, B., Helvie, M.A., 1999. Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms. Med Phys. 26, 1642-1654. doi: 10.1118/1.598658

Petroudi, S., Kadir, T., Brady, M., 2003. Automatic classification of mammographic parenchymal patterns: a statistical approach. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 798-801. doi: 10.1109/IEMBS.2003.1279885

Pierobon, M., Frankenfeld, C.L., 2013. Obesity as a risk factor for triple-negative breast cancers: a systematic review and meta-analysis. Breast Cancer Res Treat. 137, 307-314. doi: 10.1007/s10549-012-2339-3

Pietikainen, M.K., 2000. Texture analysis in machine vision. World Scientific Publishing, 981-02-4373-1, Singapore

Pisano, E.D., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J.K., Acharyya, S., Conant, E.F., Fajardo, L.L., Bassett, L., D'Orsi, C., Jong, R., Rebner, M., 2005. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 353, 1773-1783. doi: 10.1056/NEJMoa052911

Pizer, S.M., Amburn, E.P., Austin, D.D.,Cromartie, R., Geselowitz, A., Geer, T., Remeny, B., Zimmerman, J.B., Zuiderveld, K., 1987. Adaptive histogram equalization and its variations. Comput. Vision, Graph. Image Processing 39, 355-368. doi: 10.1016/S0734-189X(87)80186-X

Ponraj, D.N., Jenifer, M.E., Poongodi, P., Manoharan, J.S., 2011. A survey on the preprocessing techniques of mammogram for the detection of breast cancer. Journal of Emerging Trends in Computing and Information Science 2, 656–664. doi: 10.1.1.651.592

Pradeep, N., Girisha, H., Sreepathi, B., Karibasappa, K., 2012. Feature extraction of mammograms. International Journal of Bioinformatics research 4, 241-244.

Price, E.R., Joe, B.N., Sickles, A., 2015. The developing asymmetry: Revisiting a perceptual and revising challenge. Radiology 274. doi: 10.1148/radiol.14132759

Pu, J., Zheng, B., Leader, J.K., Gur, D., 2008. An ellipse-fitting based method for efficient registration of breast masses on two mammographic views. Med. Phys. 35, 487-494. doi: 10.1118/1.2828188

Puig, D., Garcia, M.A., 2001. Determining optimal window size for texture feature extraction methods. IX Spanish Symposium on Pattern Recognition and Image Analysis 2, 237-242. ISBN: 84-8021-351-5

Qian, W., Li, L., Clarke, L., Clark, R. A., Thomas, J., 1999. Digital mammography: comparison of adaptive and nonadoptive CAD methods for mass detection. Acad. Radiol 6, 471-480. doi: 10.1016/S1076-6332(99)80166-4

Raba, D., Oliver, A., Marti, J., Peracaula, M., and Espunya, J., 2005. Breast segmentation with pectoral muscle suppression on digital mammogram. Pattern Recognition and Image Analysis 3523, 471-478. doi: 10.1007/11492542_58

Rafael, C.G., and Richard, E.W.,2002 Digital Image Processing, Prentice Hall, Second Edition, 2002

Ramani, R., Valarmathy, S., Vanitha, N.S., 2013. Breast cancer detection in mammograms based on clustering techniques- A survey. International Journal of Computer Applications 62, 17-21. doi: 10.1.1.303.5047

Rangayyan, R.M., Shen, L., Shen, Y., Desautels, J.E.L., Bryant, H., Terry, T.J., Horeczko, N., Rose, M.S., Sep. 1997 . Improvement of sensitivity of breast cancer diagnosis with adaptive neighborhood contrast enhancement of mammograms. IEEE Transactions on Information Technology in Biomedicine 1, 161-170. doi: 10.1109/4233.654859

Rangayyan, RM., El-Faramawy, NM., Desautels, JE., and Alim, OA., Dec. 1997 Measure of acutance and shape for classification of breast tumors. IEEE Trans. on Medical Imaging 16, 799-810. doi: 10.1109/42.650876

Rangayyan, R.M., Ayres, F., 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. doi: 10.1016/j.jfranklin.2006.09.003

Rangayyan, R.M., Nguyen, T.M., 2007. Fractal analysis of contours of breast masses in mammograms. J Digit Imaging 20, 223-237. doi: 10.1007/s10278-006-0860-9

Raza, S., Odulate, A., Ong, E.M., Chikarmane, S., Harston, C.W., 2010. Using real-time tissue elastography for breast lesion evaluation: our initial experience. J Ultrasound Med. 29, 551-563. doi: 10.7863/jum.2010.29.4.551

Rejani, Y.A., Selvi, S.T., 2009. Early detection of breast cancer using SVM classifier technique. International Journzl on Computer Science and Engineering 3, 127-130.

Ren, J., 2012. ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. Knowledge-Based System 26, 144-153. doi: 10.1016/j.knosys.2011.07.016

Renehan, A.G., Tyson, M., Egger, M., Heller, R.F., Zwahlen, M., 2008. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet. 371, 569-578. doi: 10.1016/S0140-6736(08)60269-X

Roberts, L.G., 1963. Machine perception of three-dimensional solid. PhD Thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering. URL: http://hdl.handle.net/1721.1/11589

Robottino, G., Mencattini, A., Salmeri, M., Caselli, F., Lojacono, R., 2008. Mass contour extraction in mammographic image for breast cancer identification. 16th IMEKO TC4 Symposium, Exploring New Frontiers of Instrumentation and Methods for Electrical and Electronic Measurements, Florence, Italy, 11-15.

Ross, S., Ejofodomi, O., Jendoubi, A., Kinnard, L., Chouika, M., Lo, B., Wang, P., Zeng, J., 2008. A mammography database and view system for African-American patients. J Digit Imaging 21, 18-26. doi: 10.1007/s10278-007-9019-6

Saha, P. K., Udupa, J. K., Conant, E. F., Chakraborty, P., Sullivan, D., 2001. Breast tissue density quantification via digitized mammograms. IEEE Trans. Med. Imag. 20, 792-803. doi: 10.1109/42.938247

Sahiner, B., Chan, H.P., Petrick, N., Helvie, M. A., and Goodsit, M. M., 1998. Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. Medical Physics 25, 516-526

Sampat, M., Markey, M., Bovic, A., 2005. Computer-aided detection and diagnosis in mammography. In A. Bovic, Handbook of Image and Video Processing, 1195-1217. Elsevier.

Sampat, M., Whitman, G., Bovic, A., Markey, M., 2008. Comparison of algorithms to enhance spiculated masses on mammography. Journal of Digital Imaging 21, 9-17. doi: 10.1007/s10278-007-9015-x

Sankar, D., Thomas, T., 2010. A new fast fractal modeling approach for the detection of microcalcifications in mammograms. J Digit Imaging 23, 538-546. doi: 10.1007:S10278-009-9224-6

Scaranelo, A.M., Crystal, P., Bukhanov, K., Helbich, T.H., 2010. Sensitivity of a direct computer-aided detection system in full-field digital mammography for detection of microcalcifications not associated with mass or architectural distortion. Can Assoc Radiol J. 61, 162-169. doi: 10.1016/j.carj.2009.11.010

Schroeder, M., 1992. Fractals, Chaos, Power Laws: Minutes from an infinite paradise. New York, NY, USA: W. H. Freeman

Schuster, D.M., 2015. Clinical utility of PET scanning in breast cancer management. The American J of Hemat/Oncol. 11, 20-25. Available from: https://www.gotoper.com/publications/ajho/2015/2015june/clinical-utility-of-pet-scanning-in-breast-cancer-management

Seely, R., Stephens, T., Tate, P., 2004. Anatomy and Physiology. The McGraw-Hill Companies

Selvan, S.E., Xavier, C.C., Karssemeijer, N., Sequeira, J., Cherian, R.A., Dhala, B.Y., 2006. Parameter estimation in stochastic mammogram model by heuristic optimization techniques. IEEE Trans. Inf. Technol. Biomed. 10, 685-695. doi: 10.1109/TITB.2006.874197

Sengal, A. T., Haj-Mukhtar, N. S., Elhaj, A. M., Bedri, S., Kantelhardt, E. J., and Mohamedani, A. A., 2017. Immunohistochemistry defined subtypes of breast cancer in 678 Sudanese and Eritrean women; hospitals based care series. BMC Cancer 17. doi: 10.1186/s12885-017-3805-4

Shanmugavadivu, P., Sivakumar, V., Sudhir, R., 2016. Fractal dimension-bound spatio-temporal analysis of digital mammograms. The European Physical Journal Special Topics 225, 137-146. doi: 10.1140/epjst/e2016-02615-x

Shanmugavadivu, P., Sivakumar, V., 2013. Fractal-based detection of microcalcification clusters in digital mammograms. ICECIT-2012, Elsevier, 58-63. arXiv: 1304.8092v1

Shanthi, S., Bhaskaran, V.M., 2014. A novel approach for classification of abnormalities in digitized mammograms. Sadhana 39, 1141-1150. doi: 10.1007/s12046-014-0278-x

Shapiro S., Rosenberg, L., Hoffman, M., Truter, H., Cooper, D., Rao, S., Dent, D., Gudgeon, A., van Zyl, J., Katzenellenbogen, J., Bailie, R., 2000. Risk of breast cancer in relation to the use of injectable progestogen contraceptives and combined estrogen/progestogen contraceptives. Am J Epidemiol 151, 396-403.

Sharma, V., Singh, S., 2014. CFS-SMO based classification of breast density using multiple texture models. Medical and Biological Engineering and Computing 52, 521-529. doi: 10.1007/s11517-014-1158-6

Sharma,S., Khanna P., 2015. Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM. J Digit Imaging 28, pp. 77-90, doi: 10.1007/s10278-014-9719-7

Sickles, E.A., 2007. Wolfe mammographic parenchymal patterns and breast cancer risk. American Journal of Roentgenology 188, 301-303. doi: 10.2214/AJR.06.0635

Siddharth, Gupta, R., Bhateja, V., 2012. A new unsharp masking algorithm for mammography using non-linear enhancement function. Advances in Intelligent and Soft Computing 132, 779-786. doi: 10.1007/978-3-642-27443-5_89

Silva, W. R., Menotti, D., 2012. Classification of Mammograms by the breast composition. International Conference on Image Processing, Computer Vision and Pattern Recognition(IPCV)

Singh, S., Kumar, V., Verma, H.K., Singh, D., 2006. SVM based system for classification of microcalcifications in digital mammogram. Conf Proc IEEE Eng Med Biol Soc., 1, 4747-4750. doi: 10.1109/EMBS.2006.259320

Sivaramakrishna, R., Obuchowski, N., Chilcote, W., Cardenosa, G., Powell, K., 2000. Comparing the performance of mammographic enhancement. Am J Roentgenol. 175, 45-51. doi: 10.2214/ajr.175.1.1750045

Song, E., Jiang, L., Jin, R., Zhang, L., Yuan, Y., Li, Q., 2009. Breast mass segmentation in mammography using plane fitting and dynamic programming. Academic Radiology 16, 826-835. doi: 10.1016/j.acra.2008.11.014

Sorenson, J.A., Niklason, L.T., Knutti, D.F., 1980. Performance characteristics of improved antiscatter grids. Med Phys. 7, 525-528. doi: 10.1118/1.594752

Souza FH., Wendland EM., Rosa MI., Polanczyk CA., 2013. Is full-field digital mammography more accurate than screen-film mammography in overall population screening? A systematic review and meta-analysis. Breast 22, 217-224. doi: 10.1016/j.breast.2013.02.013

Sree, S.V., Ng, E.Y-K., Acharya, R.U., Faust, O., 2011. Breast imaging: A survey. Worl J Clin Oncol. 2, 171-178. doi: 10.5306/wjco.v2.i4.171

Sreedhar, K., Panlal, B., 2012. Enhancement of Image Using Morphological Transformation. International Journal of Computer Science & Information Technology 4, 33-50. doi: 10.5121/ijcsit.2012.4103

Srinivasan, G. N., Shobha, G., 2008. Statistical texture analysis. Proceeding of World Academy of Science Engineering and Technology 36

Subashini, T., Ramalingam V., Palanivel, S., 2010. Automated assessment of breast tissue density in digital mammograms. Computer Vision and Image Understanding 114, 33-43. doi: 10.1016/j.cviu.2009.09.009

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C.. Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D., Savage, J., 1994. The mammographic images analysis society digital mammogram database. Exerpta Medica. International Congress Series 1069: 375-378, available in: http:\\ peipa.essex.ac.uk/info/mias.html

Suckling, J., Dance, DR., Moskovic, E., Lewis, DJ., Blacker., SG., 1995. Segmentation of mammograms using multiple linked self-organization neural networks. Med Phys. 22, 145-152. doi: 10.1118/1.597464

Sun, X., Qian, W., Song, D., 2004. Ipsilateral-mammogram computer-aided detection of breast cancer. Computerized Medical Imaging and Graphics 28, 151-158. doi: 10.1016/j.compmedimag.2003.11.004

Sundaram, M., Ramar, K., Arumugam, N., Parbin, G., Jul. 2011. Histogram based contrast enhancement for mammogram images. 2011 IEEE International Conference on Signal Processing, Communication, Computing and Networking Technologies. doi: 10.1109/ICSCCN.2011.6024667

Sundaram, M., Ramar, K., Arumugam, N., Prabin, G., Dec. 2011. Histogram modified local contrast enhancement for mammogram images. Applied Soft Computing 11, 5809-5816. doi: 10.1016/j.asoc.2011.05.003

Surendiran, B., Vadivel, A., 2012. Mammogram mass classification using various geometric shape and margin features for early detection of breast cancer. Int. J. Medical Engineering and Informatics 4, 36-54. doi: 10.1504/IJMEI.2012.045302

Tabár, L., Tot, T., Dean, P.B., 2004. Breast cancer: The art and science of early detection with mammography: Perception, Interpretation, Histopathologic Correlation. 1st edition,Georg Thieme Velag

Tahmasbi, A., Saki, F., Shokouhi, S.B., 2011. Classification of benign and malignant masses based on Zernike moments. Comput Biol Med. 41, 726-735. doi: 10.1016/j.compbiomed.2011.06.009

Tao, Z., Shi, A., Lu, C., Song, T., Zhang, Z., Zhao, J. J., 2015. Breast cancer: epidemiology and etiology. Cell Biochemistry and Biophysics 72, 333-338. doi: 10.1007/s12013-014-0459-6

Tariq, N., 2017. Breast cancer detection using artificial neural networks. Journal of Molecular Biomarkers and Diagnosis 9. doi: 10.4172/2155-9929.1000371

Teuhola, J., Nevalainen, O., 1982. Two efficient algorithms for random sampling without replacement. IJCM 11, 127-140. doi: 10.1080/00207168208803304

Thangaraju, B., Vennila, I., Chinnasmy, G., 2012. Detection of microcalcification clusters using Hessian matrix and foveal segmentation method on multiscale analysis in digital mammograms. J Digit Imaging 25, 607-619. doi: 10.1007/s10278-012-9489-z

Thangavel, K., Karnan, M., Sivakumar, M., Mohideen, A., 2005. Automatic detection of microcalcification in mammogram- A Review. ICGST International Journal of Graphics, Vision and Image Processing, 31-61

Timp, S., Engeland, S., Karssemeijer, S., 2005. A regional registration method to find corresponding mass lesions in temporal mammogram pairs. Med Phys. 32, 2629-2638. doi: 10.1118/1.1984323

Timp, S., Karssemeijer, N., 2006. Interval change analysis to improve computer aided detection in mammography. Med. Image Anal. 10, 82-95. doi:

Torrent, A., Bardera, A., Oliver, A., Freixenet, J., Boada, I., Feixes, M., Marti, R., Llado, X., Pont, J., Pérez, E., Pedraza, S., Marti, J., 2008. Breast density segmentation: a comparison of clustering and region based techniques. International Workshop on Digital Mammography (IWDM), Digital Mammography, 9-16, LNCS 5116. doi: 10.1007/978-3-540-70538-3_2

Tourassi, G.D., Vargas-Vorecek, R., Catarious, D.M., Floyd, C.E., 2003. Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information. Med Phys. 30, 2123-2130. doi: 10.1118/1.1589494

Tzikopoulos, S., Georgiou, H., Mavroforakis, M. , Theodoridis, S., 2009. A fully automated scheme for breast density estimation and asymmetry detection of mammograms. 17th European Signal Processing Conference, IEEE, 1869-1873.

Urban, M., Banks, E., Egger, S., Canfell, K., O'Connell, D., Beral, V., Sitas, F. 2012. Injectable and oral contraceptive use and cancers of the breast, cervix, ovary, and endometrium in black South Africa women: case-control study. PLoS Med. 9. doi: 10.1371/journal.pmed.1001182

Ursin, G., Larcen, L., Parisk, Y. R., Pike, M. C., Wu, A. H., 2005. Greatly increased occurrence of breast cancer in areas of mammographically dense tissue. Breast Cancer Res. 7, 605-608. doi: 10.1186/bcr1260

Valarmathie, P., Sivakrithika, V., Dinakaran, K., 2016. Classification of mammogram masses using selected texture, shape and margin features with multilayer perceptron classifier. Biomed Res-India 2016 Special Issue: 310-313. ISSN: 0970-938X

Vanderpuye, V. S., Grover, N., PoojaPrabhakar, H., Simonds, H., Olopade, F., Stefan, D. C., 2017. An update on the management of breast cancer in Africa. Infectious Agents and Cancer 12, 13. doi: 1O.1186/s13027-017-0124-y

Vapnik, V., 1998. Statistical learning theory. New York: Wiely, 1998.

Vedanarayanan, V., Nandhitha, NM., 2017. Advanced image segmentation techniques for accurate isolation of abnormality to enhance breast cancer detection in digital mammographs. Biomedical Research 28. ISSN: 0970-938X

Velthuizen, R. P., 2000. Computer diagnosis of mammographic masses. Proceedings 29th Applied Imagery Pattern Recognition Workshop, 166-172. doi: 10.1109/AIPRW.2000.953621

Vibha, L., Harshavardhan, G.M., Pranaw, K., Shenoy, P.D., Venugopal, K.R., Patnaik, L.M., 2006. Classification of mammogram using decision trees. 10th International Database Engineering and Applications Symposium (IDEAS'06), IEEE. doi: 10.1109/IDEAS.2006.14

Vijayalakshmi, S., Nair, P. S., Nithyalakshmi, S., 2016. A FCM based approach for automated segmentation of breast masses in mammograms. International Journal of Computational Science and Information Technology (IJCSITY) 4, 39-48. doi: 10.5121/ijcsity.2016.4104

Viton, J. L., Rasigni, M., Rasigni, G., Llebaria, A., 1996. Method for characterizing masses in digital mammograms. Opt. Eng. 35, 3453-3459. doi: 10.1117/1.601107

Wagner, R.F., Beiden, S.V., Campbell, G., Metz, C.E., Sacks, W.M., 2002. Assessment of medical imaging and computer-assisted systems: lessons from recent experience. Acad Radiol. 9, 1264-1277. PMID: 12449359

Wai, L. C. C., Brady, M., 2005. Curvilinear structure based mammographic registration. Computer Vision for Biomedical Image Applications.: Lecture Notes in Computer Science (LNCS 3765) 261-270. doi: 10.1007/11569541_27

Wang, L., Zhu, M-L., Deng, L-P., Yuan, X., 2010. Automatic pectoral muscle boundary detection in mammograms based on Markov chain and active contour model. Journal of Zhejiang University 11, 111-118. doi: 10.1631/jzus.C0910025

Webster, J., 2006. Encyclopedia of Medical Devices and Instrumentation. 2nd ed., vol. 4, U.S.: Wiley Insterscience,

Wedegartne, U., Bick, U., Wortler, K., Rommeny, E., Bongartz, G., 2001. Differentiation between benign and malignant findings on MR-mammography: usefulness of morphological criteria. European Radiology 11, 1645-1650. Springer-Verlag. doi: 10.1007/s00330010088

Wei, J., Hadjiiski, L. M., Sahiner, B., Chan, H-P., Ge, J., Roubidoux, M. A., Helvie, M. A., Zhou, Ch., Wu, Y-T., Paramagul, Ch., Zhang, Y., 2007. Computer aided detection system for breast masses: comparison of performances on full-field digital mammograms and digitized screen film mammograms. Acad Radiol 14, 659-669. doi: 10.1016/j.acra.2007.02.017

Wei, J., Chan, HP., Zhou, C., Wu, Y. T., Sahiner, B., Hadjiiski, L. M., Roubidoux, M. A., Helvie, M. A., 2011. Computer-aided detection of breast masses: Four view strategy for screen mammography. Med Phys. 38, 1867-1876. doi: 10.1118/1.3560462

Weidong, X., Shunren, X., 2003. A model based algorithm to segment the pectoral muscle in mammograms. Proceedings of IEEE Int. Conf. Neural Networks & Signal processing, Nanjing, China. Piscataway, NJ: IEEE, 1163-1169. doi: 10.1109/ICNNSP.2003.1281076

Wikipedia, 2012. Demographics of Sudan. Retrieved from: https://en.wikipedia.org/wiki/Demographics of Sudan. Last update on 28 August 2018

Winkler, NS., Raza, S., Mackesy, M., Birdwell, RL., 2015. Breast density: clinical implications and assessment methods. RadioGraphics 35, 316-324. doi: 10.1148/rg.352140134

Winsberg, F., Elkin, M., Macy, J., Bordaz, V., Weymouth, W., 1967. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. Radiology 89, 211-215. doi: 10.1148/89.2.211

Wolfe, JN., 1976. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer 37, 2486-2492. PMID: 1260729

Xu, S., Liu, H., Xu, X., Song, E., Zeng, J., 2010. Bilateral asymmetry detection in mammograms using non-rigid registration and Pseudo-color coding. International Conference on Electrical and Control Eng. (ICECE2010), 544-547. doi: 10.1109/iCECE.2010.140

Yaffe, M. J., 1990. AAPM tutorial: Physics of mammography: image recording process. RadioGraphics 10, 341-363. doi: 10.1148/radiographics.10.2.2183301

Yaffe , M.J., 1994. Syllabus: a categorical course in physics: technical aspects of breast imaging. Digital mammography, 275-286, In: Haus, AG. , Yaffe, MJ., editors, Oak Brook

Yaffe, M. J., Bunch, P. C., Desponds, L., Jong, R. A., Nishikawa, R. M., Tapiovaara, M. J., Young, K. C., 2009. Mammography- assessment of image quality: Noise. Journal of the International Commission on Radiation Units and Measurements 9, 36-38. doi: 10.1093/jicru/ndp029

Yang, Y., Zheng, Ch., Lin, P., 2005. Fuzzy c-means clustering algorithm with a novel penalty term for image segmentation. Opto-Electronics Review 13, 309-315.

Yankaskas, BC., Cleveland, RJ., Schell, MJ., Kozar, R., 2001. Association of recall rates with sensitivity and positive predictive values of screening mammography. American Journal of Roentgenology (AJR) 177, 543-549. doi: 10.2214/ajr.177.3.1770543

Yapa, R. D., Harade, K., 2007. A Connected Component Labeling Algorithm for Grayscale Images and Application of the Algorithm on Mammograms. ¨roceedings of the 2007 ACM symposium on Applied Computing, 146-152. doi: 10.1145/1244002.1244040

Yoon, WB., Oh, JE., Chae, EY., Kim, HH., Lee, SY., Kim, KG., 2016. Automatic detection of pectoral muscle region for computer aided diagnosis using MIAS mammograms. BioMed Research International, Article ID 5967580. doi: 10.1155/2016/5967580

Yousefi, P., 2015. Mammographic image enhancement for breast cancer detection applying wavelet transform. IEEE Student Symposium in Biomedical Engineering & Sciences (ISSBES). doi: 10.1109/ISSBES.2015.7435919

Yuan, Y., Giger, ML., Li, H., Suzuki, K., Sennett, C., 2007. A dual-stage method for lesion segmentation on digital mammograms. Med. Phys. 34, 4180-4193. doi: 10.1118/2790837

Zadeh, H.S., Nezhad, S.P., Radc, F.R., 2001. Shape-based and texture-based feature extraction for classification of microcalcifications in mammograms. In Proceedings of SPIE- The International Society for Optical Engineering 4322, 301-310. doi: 10.1117/12.431100

Zahedi, Z., Sadri, S., Soltani, M., Tehrani, M. K., 2011. Breast diseases detection and pseduo-coloring presentation for gray infrared breast images. IEEE, Asia Communications and Photonics Conference and Exhibition, ACP. doi: 10.1117/12.905604

Zhang, W., Yoshida, H., Nishikawa, R., Doi, K., 1998. Optimally weighted wavelet transform based on supervised training for detection of microcalcifications in digital mammograms. Med. Phys. 25, 949-956. doi: 10.1118/1.598273

Zheng, B., Good, W. F., Armfield, D. R., Cohen, C., Hertzberg, T., Sumkin, J. H., Gur, D., 2003. Performance change of mammographic CAD schemes optimized with most-recent and prior image database. Academic Radiology 10, 283-288. doi: 10.1016/S1076-6332(03)80102-2

Zheng, L., Chan, A.K., 2001. An artificial intelligent algorithm for tumor detection in screening mammogram. IEEE Trans. on Medical Imaging 20, 559-567. doi: 10.1109/42.932741

Zheng, Y., Keller, B. M., Ray, S., Wang, Y., Conant, E. F., Gee, J. C., Kontos, D., 2015. Parenchymal texture analysis in digital mammography: A fully automated pipeline for

breast cancer risk assessment. Medical Physics 42, 4149-4160. doi: 10.1118/1.4921996

Zhu, Y., Huang, Ch., 2012. An improved median filtering algorithm for image noise reduction. Physics Procedia 25, 609-616. doi: 10.1016/j.phpro.2012.03.133

Ziskin, M.C., 1993. Fundamental physics of ultrasound and its propagation in tissue. RadioGraphics 13, 705-709. doi: 10.1148/radiographics.13.3.8316679

Zimmerman, J.B., Pizer, S.M., Staab, E.V., Perry, J.R., McCartney, W., Brenton, B.C., 1988. An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement. IEEE Trans. Med. Imaging 7, 304-312. doi: 10.1109/42.14513

Zou, Y., Guo, Z., 2003. A review of electrical impedance techniques for breast cancer detection. Med Eng Phys. 25, 79-90. doi: 10.1016/S1350-4533(02)00194-7

Zucker, S. W., 1976. Region growing: Childhood and adolescence. Computer Graphics and Image Processing 5, 382-399. doi: 10.1016/S0146-664X(76)80014-7

Zuiderveld, K., 1994. Contrast Limited Adaptive Histogram Equalization. Graphics Gems IV, 474-485. Academic Press Professional, Inc. San Diego, CA, USA. ISBN: 0-12-336155-9. doi: 10.1016/B978-0-12-336156-1.50061-6

Appendix

Program Codes

```
%% Preprocessing algorithm, 13Nov013, work a
lhamdo'LLAH;
% applied on Learning dataset 3Apr014
%function Pect1
%remove pectoral muscle

% Select Image
[FileName,PathName]=uigetfile('*.pgm','Select the Image');
im=strcat(PathName,FileName);


%Stage1

% cropping img , function test
im=imread(im);
imd=im2double(im);
imshow(imd)
figure, imhist(imd)

d=double(im);
imdd=imd*d;
imdd=im2double(imdd);
th=graythresh(imdd);
thimdd=im2bw(imdd,th);
thimdd = imopen(thimdd,strel('disk',1)); %image like 0019, not cropped if this statement
abscent, cause nonzeros pixel
BinImge = thimdd;
[nonZeroRows nonZeroColumns] = find(BinImge);
topRow = min(nonZeroRows(:));
bottomRow = max(nonZeroRows(:));
leftColumn = min(nonZeroColumns(:));
rightColumn = max(nonZeroColumns(:));
croppedImage = BinImge(topRow:bottomRow, leftColumn:rightColumn);
imo=im(topRow:bottomRow , leftColumn:rightColumn);

im=imo;
figure,imshow(imo);
figure,imhist(imo);

%%
%Stage2
% Label Omitting Function test_jm.m
x=im2double(im);
t=graythresh(x);
```

```
bi=im2bw(x,(t-0.2));  % decreasing threshold,let breast outer soft tissue appear; omit zigzag, soft tissue inside breast appear like in mdb001-002,(img190,not work if (t-0.25)so become(t-0.2)
bi = imopen(bi,strel('disk',1)); % solve problem that label connect with breast by upper border of image, clear in binary image, eg: img003
c= bwconncomp(bi);
d = c.PixelIdxList;
max_vox = 0;
biggest_object = 0;
for i = 1:c.NumObjects
   if (size(d{1,i},1) > max_vox)
      max_vox = size(d{1,i},1);
      biggest_object = i;
   end
end
biggest_object;
v = bi;
v(:)= 0;
v(d{1,biggest_object}) = 1;
m=immultiply(v,x);

figure,imshow(m);
figure,imhist(m);
%%
%Step3
% Just flipping odd files numbers, 'right breast'
im=m;
t=graythresh(im);
bi=im2bw(im,t);
[m n]=size(bi);
 %check the breast position

 if  bi((1:m),1)==0

            im=fliplr(im);

         else
           im=im;
 end


   %figure, imshow(im) %,title('Original Image')
   %figure,imhist(im) %, title('Orig hist');
   %%
   %%  SPnoise
%
newim=zeros(size(im)+2); %mask image
B=zeros(size(im)); %output image

for i=1:size(im,1)
```

```
    for j=1:size(im,2)
       newim(i+1,j+1)=im(i,j);
    end
end

for x=1:size(newim,1)-2
    for y=1:size(newim,2)-2
       window=zeros(3);
       inc=1;

       for i=1:3
       for j=1:3
          window(inc)=newim(x+i-1,y+j-1);
          inc=inc+1;
       end
       end

       med=sort(window);
    B(x,y)=med(3);
   end
end

denois=mat2gray(B);
im=denois;
%figure,imshow(newim),figure,imshow(B),figure,imshow(im);
figure,imshow(im); figure,imhist(im);
%im=adapthisteq(im); % not suitable here, will use in breast only as ROI
%%


   % piecewise linear

   %im=imread(im);
%im=im2double(im);
a=[0 1];
b=[1 0];
N=length(a);
out=ones(size(im));

for i=1:N-1
    pix=find (im>a(i) & im<a(i+1));
    out(pix)=(im(pix)-a(i))*(b(i+1)-b(i))/(a(i+1)-a(i))+b(i);
end

pix=find(im==a(N));
out(pix)=b(N);
m=imsubtract(im,out);

pect=m;
pect=imopen(pect,strel('disk',2));
```

```
figure,imshow(pect);
figure,imhist(pect);
%  %

  %%
  % Step4-1 %not give require output
  % pectoral muscle removal
 % t=graythresh(im);
 %bi=im2bw(im,t);
  %l=bwlabel(bi,8);
  %x=find(l==1);
  %x=find(l~=1);
  %bwl=im;
  %bwl(x)=0;

  %imshow(bwl)
  %  select suitable threshold method??? (not yet)
  %%
  %figure, imhist(im);
  %Step4-2
thresh=multithresh(im,3); %the 3rd value give thresh for pectoral muscle, it just after peak,
unique threshold for same person eg:(im1;im2),(3,4),....etc
 %three=(thresh(2)+thresh(3))*(2/3);
 %thresh=thresh+0.3;
 pect =im;
[x,y] = size(im);
for i=1:x
    for j=1:y
        if (pect(i,j)<= thresh(3))
            pect(i,j)=0;
        end
    end
end

 figure,imshow(pect);
 pw=im2bw(pect);
 figure,imshow(pw);
 %title('segmentation');


sub=imsubtract(im,pect);
figure,imshow(sub)
     imp=pect;
impd=im2double(imp);
t=graythresh(impd);
bi=im2bw(impd,t);
bi = imopen(bi,strel('disk',2)); %because some pectoral muscle connect with breast, eg:im
c=bwconncomp(bi);
d = c.PixelIdxList;
max_vox = 0;
```

```
biggest_object = 0;
for i = 1:c.NumObjects
    if (size(d{1,i},1) > max_vox)      % && (d(1,i)==findobj(gca,'Type','line') )) % and shape
is traingle then
       max_vox = size(d{1,i},1);
       biggest_object = i;
   end
end
%biggest_object;
 v = imp;
 v(:)= 1;
 v(d{1,biggest_object}) =0;
c1=bwconncomp(v);
d1=c1.PixelIdxList;
max_vox1=0;
biggest_object1=0;
for j=1:c1.NumObjects
   if (size(d1{1,j},1) > max_vox1)
       max_vox1=size(d1{1,j},1);
       biggest_object1=j;
   end
end

v1=imp;
v1(:)=1;
v1(d{1,biggest_object}) =0;
v1(d{1,biggest_object1})=0;
%figure,imshow(v1);


vc=bwconncomp(v);
vc1=bwconncomp(v1);
if (((vc.NumObjects) ~= (vc1.NumObjects))||((bwarea(v)) ~= (bwarea(v1))))
  pe=imsubtract(v,v1);                                    % _correct by add OR statement , it
work good
  pe=~pe;
  %figure,imshow(pe)
  m=immultiply(pe,im);
  %figure,imshow(f);
else
   %figure,imshow(v)
   m=immultiply(v,im);

end
m=im2double(m);
m=adapthisteq(m);
figure,imshow(m)
figure,imhist(m)
```

```
Segmentation
ROI selection based on image quantization; manual selection
% initiate 24Apr014
% last modify 14May014
%%
fontSize = 20;

%%
clinee
%%
subplot(2,3,1);
imshow(imd)
title('Original Image','FontSize',fontSize);

%%
threseg=multithresh(im,7);
seg1=imquantize(im,threseg);
RGB=label2rgb(seg1);
%figure, imshow(RGB)

%%
% ROI manually selection
Pseudo = RGB;
subplot(2, 3, 2);
imshow(Pseudo, []);
title('Quantization', 'FontSize', fontSize);
set(gcf, 'Position', get(0,'Screensize')); % Maximize figure.

%Nroi=sprintf('Enter Numbers of ROI.\n ');
%uiwait(msgbox(Nroi));
%n=input(Nroi);

message = sprintf ('Left click and hold to begin drawing.\nSimply liftthe mouse button to
finish');
uiwait(msgbox(message));
%hFH = imfreehand();  % free hand selection
hFH=imrect(); % to drag rectangle
position=wait(hFH);
                % double click on rectangle to resume execute matlab command

% Create a binary image ("mask") from the ROI object.
binaryImage= hFH.createMask();
% Display the freehand mask.
subplot(2, 3, 3);
imshow(binaryImage);
title('ROI selection', 'FontSize', fontSize);
% Burn line into image by setting it to 255 wherever the mask is true.
burnedImage = im;
burnedImage(binaryImage) = 255;
% Display the image with the mask "burned in."
```

```
subplot(2, 3, 4);
imshow(burnedImage);
title('Mask burned into GrayImage', 'FontSize', fontSize);

% Mask the image and display it.
% Will keep only the part of the image that's inside the mask, zero outside mask.

maskedImage = im;
maskedImage(~binaryImage) = 0;
subplot(2, 3, 5);
imshow(maskedImage);
title('Masked Image', 'FontSize', fontSize);
%
%obtain mask image
[nonZeroRows nonZeroColumns] = find(binaryImage);
topRow = min(nonZeroRows(:));
bottomRow = max(nonZeroRows(:));
leftColumn = min(nonZeroColumns(:));
rightColumn = max(nonZeroColumns(:));
cm=maskedImage(topRow:bottomRow , leftColumn:rightColumn);
figure,imshow(cm);
%figure,imhist(cm);

%%
h=adapthisteq(cm);
figure,imshow(h);
%figure,imhist(h);
%%
%binary image
mhb=graythresh(h);
mhbb=im2bw(h,mhb);
figure,imshow(mhbb);
%%
% multiply enhanced image with mask to obtain segmented mass
mhbb=im2double(mhbb);
mem= immultiply(h,mhbb);
figure,imshow(mem);
%figure,imhist(mem);
% test area of binary
memt=graythresh(mem);
memb=im2bw(mem,memt);
% memb=im2double(memb); % if not convert to double it will be logical and if applied
regionprops(memb,'area','perimeter'), it give struct array
%% select the big mass
%x=im2double(im);
%t=graythresh(x);
%memb=im2bw(memb,(memt-0.2));
bii = imopen(memb,strel('disk',2));
c= bwconncomp(memb);
d = c.PixelIdxList;
```

```
max_vox = 0;
biggest_object = 0;
for i = 1:c.NumObjects
   if (size(d{1,i},1) > max_vox)
      max_vox = size(d{1,i},1);
      biggest_object = i;
   end
end
biggest_object;
v = bii;
v(:)= 0;
v(d{1,biggest_object}) = 1;
ml=immultiply(v,memb);
figure,imshow(ml)
m11=immultiply(ml,mem);
figure,imshow(m11)
```

MIAS Coordinates

```
function a = MiasCoordinates( TextFilename,DatasetNumber)
TextFilename =('Abnorm.txt');
DatasetNumber = 1;

   f = fopen(TextFilename);
   StructFile = textscan(f,'%s %s %s %s %d %d %d');

   ImageName = strcat(StructFile{1,1}{DatasetNumber,1},'.pgm');
   ImageVol = imread(ImageName);


   xCoord = double(StructFile{1,5}(DatasetNumber));
   yCoord = double(size(ImageVol,2) - StructFile{1,6}(DatasetNumber));
   Radius = double(StructFile{1,7}(DatasetNumber));

   imagesc(ImageVol);
   colormap(gray(256));
   a = circle(xCoord,yCoord,Radius);
   fclose(f)
end
```

SFTA code

```
function [ D ] = sfta( I, nt )
  % If necessary, convert I to a grayscale image with bit-depth of 8.
   I=imread('mdb005.pgm');
  nt=4;

   I = im2double(I);
  % if size(I,3) ~= 1
   %    I = rgb2gray(I);
```

```
%end;

T = otsurec( I, nt );
dSize = (numel(T) * 6) - 3;
D = zeros(1, dSize);
pos = 1;
for t = 1 : numel(T)
   thresh = T(t);

   Ib = im2bw(I, thresh);  figure,imshow(Ib)
   Ib = findBorders(Ib);

   vals = double(I(Ib));

   D(pos) = hausDim(Ib);
   pos = pos + 1;

   D(pos) = mean(vals);
   pos = pos + 1;

   D(pos) = numel(vals);
   pos = pos + 1;

end;

T = [T; 1.0];
range = getrangefromclass(I);
range = range(2);

for t = 1 : (numel(T) - 2)
   lowerThresh = T(t);
   upperThresh = T(t + 1);

   Ib = I > (lowerThresh * range) & I < (upperThresh * range);   figure,imshow(Ib)
   Ib = findBorders(Ib);

   vals = double(I(Ib));

   D(pos) = hausDim(Ib);
   pos = pos + 1;

   D(pos) = mean(vals);
   pos = pos + 1;

   D(pos) = numel(vals);
   pos = pos + 1;
```

Haralick features

```
% init 25Apr014, UPMC
```

```
% Haralick 14 Feature Extraction from GLCM
% GLCM (glcm), calculated in segm code.
%%
 % HaralickFeatures =
        % {1 'AngularSecondMoment',...
         % 2 'Contrast',... = Gray level differences
          %3 'Correlation',...
          %4 'Variance',...
          %5 'InverseDifferenceMoment',...
          %6 'SumAverage',...
          %7 'SumVariance',...
          %8 'SumEntropy',...
          %9 'Entropy',...
          %10 'DifferenceVariance',...
          %11 'DifferenceEntropy',...
          %12 'InfoMeas1',...
          %13 'InfoMeas2'
          %14 'Max.Correlation coeff}
%%
% AutoSeg;   when stract masses
%%
function ha = har(I);

% 24April014
% create Gray Level Co=occurence Matrix from image
%ROI=m11;  % if use segm.m
ROI=I;
%ROI=BWH{ij};
%glcm=graycomatrix(ROI);
glcm = graycomatrix(ROI,'Offset',[-1 1]); % offset angle45Â°, Distance=1

% kernel = [0 1; 0 -1; 1 0; -1 0]
% glcm = graycomatrix( image, 'NumLevels', 8, 'GrayLimits', [],'offset', kernel );
% stats = graycoprops( glcm, 'Contrast Correlation Energy Homogeneity');
% c = mean( stats.Contrast );
% cor = mean( stats.Correlation );
% e = mean( stats.Energy );
% h = mean( stats.Homogeneity );

%%
% matlab function graycoprops give 4features from 14Haralick features
FourProp=graycoprops(glcm,{'Contrast','Correlation','Energy','Homogeneity'});
%%

SGLD= glcm;

%%
% GLCM requirements (1,2,3):
% 1- Decide the relation between the refrence and neighbour-
% pixels(angle,offet)
```

```
% 2- GLCM be a semmetrical matrix
% 3- Normalize the matrix
% make the SGLD matrix symmetric by adding it's transpose to it
SGLD=SGLD+SGLD';

% normalize the SGLD matrix to values between 0 and 1
SGLD=SGLD/sum(sum(SGLD));

% ******************************************************
% Calculating the texture features from the SGLD matrix
% ******************************************************
foo=SGLD;

% Entropy
entropy=sum(sum(-((full(spfun(@log2,foo))).*foo)));

% Energy:
energy=sum(sum(foo.*foo));

% Inertia:
[i,j,v]=find(foo);
inertia=sum((((i-1)-(j-1)).*((i-1)-(j-1))).*v);

% Variance
% variance=var(foo);

% Inverse differnece moment:
inverse_diff=sum((1./(1+(((i-1)-(j-1)).*((i-1)-(j-1))))).*v);

% Correlation:
[m,n]=size(foo);

px=sum(foo,2);
[i,j,v]=find(px);
mu_x=sum((i-1).*v);
sigma_x=sum((((i-1)-mu_x).^2).*v);
h_x=sum(sum(-((full(spfun(@log2,px))).*px)));
temp1=repmat(px,[1 m]);

py=sum(foo,1);
[i,j,v]=find(py);
mu_y=sum((j-1).*v);
sigma_y=sum((((j-1)-mu_y).^2).*v);
h_y=sum(sum(-((full(spfun(@log2,py))).*py)));
temp2=repmat(py,[n 1]);

[i,j,v]=find(foo);
correlation=(sum(((i-1)-mu_x).*((j-1)-mu_y).*v))/sqrt(sigma_x*sigma_y);

% Information measures of correlation 1 and 2:
```

```
foo1=-(foo.*(((temp1.*temp2)==0)-1));
foo2=-((temp1.*temp2).*((foo1==0)-1));
[i1,j1,v1]=find(foo1);
[i2,j2,v2]=find(foo2);
h1=sum((sum(-(v1.*(log2(v2))))));
info_corr_1=(entropy-h1)/max(h_x,h_y);
[i,j,v]=find(temp1.*temp2);
h2=sum((sum(-(v.*(log2(v))))));
info_corr_2=sqrt((1-exp(-2*(h2-entropy))));

% Sum average, variance and entropy:
[i,j,v]=find(foo);
k=i+j-1;
pk_sum=zeros(max(k),1);
for l=min(k):max(k)
pk_sum(l)=sum(v(find(k==l)));
end

[i,j,v]=find(pk_sum);
sum_avg=sum((i-1).*v);

sum_var=sum((((i-1)-sum_avg).^2).*v);
sum_entropy=sum(-((full(spfun(@log2,pk_sum))).*pk_sum));
% Difference average, variance and entropy:
[i,j,v]=find(foo);
k=abs(i-j);
pk_diff=zeros(max(k)+1,1);
for l=min(k):max(k)
pk_diff(l+1)=sum(v(find(k==l)));
end

[i,j,v]=find(pk_diff);
diff_avg=sum((i-1).*v);
diff_var=sum((((i-1)-diff_avg).^2).*v);
diff_entropy=sum(-((full(spfun(@log2,pk_diff))).*pk_diff));

%********
%%
ha= [energy correlation inertia entropy inverse_diff sum_avg sum_var sum_entropy diff_avg
diff_var diff_entropy info_corr_1 info_corr_2];
% I=inertia;
% IC = info_corr_2;
% CO= correlation;
% SE= sum_entropy;
% ha=[I,IC, CO, SE];
%fprintf (fileID,'%1.4f\t %1.4f\t %1.4f\t %1.4f\t %1.4f\t %1.4f\t %1.4f\t %1.4f\t
%1.4f\t %1.4f\t %1.4f\t %1.4f\t\n',F);
end
```