
ADDIS ABABA UNIVERSITY

COLLEGE OF TECHNOLOGY AND BUILT ENVIRONMENT

SCHOOL OF BIOMEDICAL ENGINEERING

# SVM Based Detection and Classification of Breast Abnormalities in Mammography Images

**BY**

**RAHEL BEYENE SHEWA**

In Partial Fulfillment of the Requirements for the Degree of Master of engineering in Biomedical Engineering

**Advisor:** Dawit Assefa Haile (PhD)

November 2024

Addis Ababa, Ethiopia

## Declaration

I, the undersigned, declare that this MSc Project is my original work, has not been presented for fulfillment of a degree in this or any other University, and all sources and materials used for the project have been acknowledged.

Name: Rahel Beyene Shewa

Signature: _____

Date: _____

This MSc. Project has been submitted for examination with my approval as an advisor.

_____

Dawit Assefa Haile (PhD)

# Certificate of Examination

## Addis Ababa University

## School of Graduate Studies

This is to certify that the project prepared by Rahel Beyene Shewa entitled ‘‘*SVM based Detection and Classification of Breast Abnormalities in Mammography Images*’’ submitted in partial fulfillment of the requirements for the Degree of Master of Engineering in Biomedical Engineering (Bioinstrumentation and Imaging) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the examining committee:

Examiner: _____ Signature: _____ Date: _____

Examiner: _____ Signature: _____ Date: _____

Advisor: _____ Signature: _____ Date: _____

Chairman: _____ Signature _____ Date _____

_____

Chief of Department or Graduate program coordinator

# Abstract

Breast cancer is the second leading cause of death among females. Digital mammography is the most effective screening technique for early detection of breast cancer and other abnormalities. However, interpreting digital mammograms can be challenging due to the small differences in attenuation among various soft tissue structures in the female breast. This results in low-quality X-ray images, high background noise, and artifacts, making diagnosis difficult. Therefore, it is essential to develop an effective computer-aided diagnosis system to improve the detection and classification of abnormalities in mammogram images. The proposed Support Vector Machine (SVM)-based mammography image detection and classification method involves four stages: preprocessing, segmentation, feature extraction, and classification. During preprocessing, the images undergo denoising with median and winner filters, pectoral muscle removal using the seeded region grow algorithm, background removal through morphological operations, and image enhancement with range contrast adjustment and Contrast Limited Adaptive Histogram Equalization (CLAHE). Segmentation is performed using global thresholding. In the feature extraction stage, 24 features were calculated from locally computed Gray Co-occurrence Matrices (GLCM). Finally, SVM was employed to classify the mammogram images. The study utilized 274 mammograms, of which 102 were abnormal-samples (malignant or benign) and 172 were normal images taken from control subjects. Of these, 80% were used for training and 20% for testing the SVM classification model. Based on the performance metrics, the system has an 84.82% probability of detecting abnormalities in patients breast tumor with positive predictive value (PPV) of 94.06%. It correctly classifies mammogram images as normal in 96.3% of patients without breast tumor (specificity = 96.3%), with an overall accuracy of 91.6% with negative predictive value (NPV) of 90.17%. Among the various feature combinations, dissimilarity and inverse difference normalized features provided the highest AUC value of 0.92.



**Keywords:** Mammogram, Mammography, Support Vector Machine, Pectoral Muscle, Classification.

# Acknowledgment

To begin with, I express my gratitude to God, the Almighty, for bestowing upon me numerous blessings, knowledge, strength, and opportunities, enabling me to successfully complete this project.

I extend my deepest gratitude to my advisor, Dawit Assefa Haile (PhD.), for his unwavering guidance, ongoing support, encouragement, enthusiasm, patience, insightful feedback, and extensive knowledge throughout the research and writing of this project.

I am also immensely thankful to my peers and colleagues for their valuable feedback and exceptional collaboration, which greatly enhanced the intellectual dialogue of this research project.

Lastly, I express my heartfelt appreciation to my parents for their unending support and unwavering belief throughout my educational journey. This achievement would not have been attainable without their help.

Thank you to everyone who has contributed to the completion of this project, for their willingness to share their experience and insights, without which this research project would not have been possible.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| ACS | **A**merican **C**ancer **S**ociety |
| AEC | **A**utomatic **E**xposure **C**ontrol |
| AHE | **A**daptive **H**istogram **E**qualization |
| BIRADS | **B**reast **I**maging **R**eporting and **D**ata **S**ystem |
| CAD | **C**omputer **A**ided **D**etection |
| CLAHE | **C**ontrast **L**imited **A**daptive **H**istogram **E**qualization |
| MLO | **M**ediolateral **O**blique |
| CC | **C**ranio **C**audal |
| CCD | **C**harge-**C**oupled **D**evice |
| DCIS | **D**uctal **C**arcinoma **I**n **S**itu |
| DDSM | **D**igital **D**atabase **F**or **S**creening **M**ammography |
| FP | **F**alse **P**ositive |
| FN | **F**alse **N**egative |
| GLCM | **G**ray **L**evel **C**o-occurrence **M**atrix |
| IBC | **I**nflammatory **B**reast **C**arcinoma |
| INN | **I**nverse **D**ifference **N**ormalized |
| LCIS | **L**obular **C**arcinoma **I**n **S**itu |
| MIAS | **M**ammographic **I**mage **A**nalysis **S**ociety |
| MRI | **M**agnetic **R**esonance **I**maging |
| NCI | **N**ational **C**ancer **I**nstitute |
| ROI | **R**egion **O**f **I**nterest |
| SVM | **S**upport **V**ector **M**achine |
| TP | **T**rue **P**ositive |
| TN | **T**rue **N**egative |
| US | **U**ltra**s**ound |

# Chapter 1

# Introduction

## 1.1 Background

The word cancer implies the abnormal reproduction of cells from a specific location that spreads to other parts of the body via the bloodstream and lymphatic system, although the root cause is still unknown [1]. Cancer is a global health issue that affects individuals and communities worldwide. Research in this field commenced in the 1900s and was once considered incurable[2][3][4]. It remains a primary cause of mortality around the world, leading to approximately 10 million fatalities in 2020, which is nearly one in every six deaths. Roughly 70% of these fatalities occur in low- and middle-income countries. Globally, breast cancer is the second most common form of cancer and ranks among the top five causes of death for women worldwide [4][5][6][7].

Breastcancer.org reported that breast cancer has become the most common diagnosed cancer among women worldwide, surpassing lung cancer. In 2020, there were approximately 2,261,419 new cases of breast cancer diagnosed in women globally. In the United States, breast cancer is the second most diagnosed cancer among women, following skin cancer. It represents one-third of all new cancer cases among women each year. In 2023, estimates suggested that 297,790 women in the United States would be diagnosed with invasive breast cancer, while 55,720 women would be diagnosed with non-invasive (in situ) breast cancer. The annual rate of invasive breast cancer in women has been increasing by about 0.5% since the mid-2000s. In 2023, invasive breast cancer diagnoses in men in the United States were anticipated to reach approximately 2,800 cases [8] [9]. At present, over four million women in the U.S. are living with breast cancer, which was estimated to have caused around 43,700 deaths in 2023.

In 2020, approximately 684,996 women globally lost their lives to breast cancer [8][9]. According to the 2022 statistics from the World Health Organization (WHO), mortality rates are increasing, particularly in areas lacking early detection programs. Early detection plays a crucial role in reducing the death rate, facilitating effective treatment, and improving both survival rates and

prognoses [10]. However, achieving early detection necessitates an accurate and reliable diagnosis to distinguish between malignant and benign tumors [11].

Women are more prone to developing breast cancer compared to men. The cancer can occur in various parts of the breast, such as the ducts, lobules, and surrounding tissue. Typically, breast cancer usually begins in the cells of the lobules, which are responsible for producing milk, or in the ducts that transport milk from the lobules to the nipple. It can also less commonly originate in the stromal tissues, which are made up of the fat and fibrous connective tissue of the breast. A self-exam can help detect the early signs of breast cancer, such as small lumps in the breast tissue. Other methods for identifying abnormal tissues include clinical breast examinations and breast imaging techniques like mammography, ultrasound (US), magnetic resonance imaging (MRI), and scintimammography. Mammography is regarded as the primary effective approach for the early detection of breast cancer and other breast-related conditions [12].

Mammography is a type of low-dose X-ray imaging technique employed for diagnosing and screening breast conditions. Images obtained from a mammogram are utilized to detect unusual tissues in the breast that could indicate the presence of cancer or other abnormalities [13][14][15]. The first attempts to utilize radiography for identifying breast abnormalities occurred in the late 1920s, but the mammography we recognize today, featuring specialized X-ray units, emerged in the 1960s [3]. During that decade, radiologists conducted mammography examinations with standard X-ray tubes and without compression, recording images on direct-exposure films similar to those used for chest X-rays. These resulting images had low contrast, and the tissue area adjacent to the chest wall appeared "white" due to inadequate exposure [3] [16].

Mammography underwent significant changes with the introduction of screen-film mammography which enhanced imaging speed, decreased radiation exposure, and improved contrast, allowing for better visualization of breast tissues. In the 1980s and 1990s, advancements in screen-film technology and the establishment of specialized mammography units resulted in higher-quality images. As technology continued to evolve, mammography screening for breast cancer became more common for two primary reasons: a) various randomized controlled studies demonstrated that mammography screening significantly lowered breast cancer mortality rates, b) the

development of effective pre-operative image-guided wire localization techniques made it easier to diagnose tissue from suspicious lesions identified through mammography [15].

Nowadays, MRI and US imaging are used as alternatives to mammography in the detection of breast cancers**.** Despite the limitations of these modalities, which include low contrast and resolution, sensitivity difficulties, false positive results, blurry tissue boundaries, and speckle noise , there is a growing interest in using MRI and US imaging for the detection of breast cancer [16]. Ultrasound (US) generates images with good contrast but lacks the detailed information provided by digital mammograms. While Magnetic Resonance Imaging (MRI) is more sensitive than digital mammograms, it can result in false positive diagnoses, which can lead to unnecessary follow-up tests, biopsies, and increased anxiety for patients. The American Cancer Society advises the use of MRIs for women whose lifetime risk of breast cancer is 20–25% or greater, particularly those with a significant family history of breast or ovarian cancer. The benefits of digital mammograms for the early detection of breast cancer clearly surpass those of the other methods mentioned. This aligns with the results of numerous studies indicating that digital mammograms are more effective in identifying early-stage breast cancer [17].

Even though digital mammography is recognized as an effective technique for identifying breast cancer, the interpretation of these mammograms necessitates the expertise and proficiency of a qualified radiologist. It is observed that approximately 10-30% of breast lesions may be missed during routine screenings. Having two radiologists independently review the mammograms can enhance sensitivity but also raises the cost of the screening process. Therefore, computer-aided detection (CAD) can serve as a supplementary reviewer, providing additional analysis while leaving the final decision to the radiologist [12].

Over the years, many scientists have used different approaches and attempts to improve image enhancement such as low contrast, and poor visualization of abnormalities, artifact detection, and noise removal, to assist analysts in automatic breast cancer detection in mammography imaging. But there are still some limitations on their level of accuracy and computational efficiency as well. The current study aimed to improve image enhancement and suggest a better image quality improvement approach in mammography image CAD technique. The study aims to develop a methodology for detecting and classifying breast tumors using Support Vector Machine (SVM)

classification of mammography images. The goal is to create an automated system that enhances the accuracy of breast cancer diagnoses by effectively analyzing mammography images to distinguish between normal and abnormal tumors.

## 1.2 Statement of the Problem

The imaging method and technology utilized to create the radiography image are the key factors in determining the characteristics and quality of the result. Mammography is the main imaging modality for early detection and assessment of breast abnormalities. The difference in attenuation of soft tissue structures in the breast are minimal, requiring the use of X-rays with low photon energies to attain sufficiently high contrast in mammography images. However, the overall contrast remains low, particularly in dense and glandular breast tissues.

Improving contrast, especially for dense breasts, is the primary enhancement needed in mammography. For instance, the contrast between malignant and normal dense tissue may be present but below the threshold of human perception. Similarly microcalcifications within dense masses often lack sufficient contrast, making it challenging to define their characteristics. Additionally, mammogram images frequently have high background noise and artifacts, complicating diagnosis. In such cases, radiologists may miss diagnostically important findings.

In response to those problems, the current project aims to develop a novel mammogram image detection and classification method by exploring various image enhancement techniques proven by scholars. It is crucial to employ an effective image enhancement method to enhance the quality of the image and assist radiologists in more accurate breast cancer detection. The goal of the project is to improve the accuracy and reliability of mammography image analysis, leading to early detection of breast cancer, which is crucial for successful treatment and higher survival rates.

## 1.3 Objective of the Project

### 1.3.1 General Objective

The main objective of this project is to produce a better mammography image detection and classification approach for use in breast cancer diagnosis.

### 1.3.2 Specific Objectives

- To develop a pre-processing technique to smoothing of image, contrast enhancement and noise removal.
- To investigate the application of SVM in mammography image classification (normal and abnormal) and detection.
- To carry out qualitative and quantitative assessment of the proposed algorithm with respect to gold standards provided by experts and show directions of their clinical implications.
- To indicate potential clinical implications.

## 1.4 Significance of the Study

Since the introduction of mammography for early breast abnormality detection, numerous studies have been published highlighting its benefits compared to other imaging techniques. Mammography is especially valuable for early detection and diagnosis of breast cancer. The findings of the current study could contribute to improving mammographic cancer detection and classification. By incorporating advanced machine learning techniques, as later discussed in this study, the number of missed diagnoses (false negatives) can be reduced, ensuring more cases are caught at an early, treatable stage. The proposed detection scheme can aid radiologists by providing an additional tool to verify their findings and make better-informed decisions. It is also believed that the study will further open doors for future researchers to refine and expand studies in relation to machine learning based mammography image detection and classification techniques. As such algorithms continue to develop and integrate into clinical practice, they hold the promise of making a significant impact on both individual patient care and the broader healthcare system. Its success can result in more accurate, efficient, and accessible diagnostic tools, ultimately making a profound impact on public health.

## 1.5 Scope and Delimitations of the Study

The methodology developed in this project for detecting breast abnormalities has been tested on mammography images sourced from research databases. Only breast images of female subjects were included in this study. The entire methodology was implemented on the MATLAB platform. However, the actual clinical validation of the work falls outside the scope of the current study.

## 1.6 Organization of the Project Document

The remaining sections of the project document is organized into four chapters. Chapter 2 covers the anatomy of the breast and various breast pathologies. Also, in this chapter technical aspects and physics behind mammography exam and equipment are highlighted including state of the art technologies. Chapter 3 presents the proposed methodology detailing the pre-processing, segmentation, feature extraction and classification steps. Chapter 4 presents the experimental results along with some insightful discussions. Chapter 5 presents the conclusions, recommendations as well as possible future directions of this project work.

# Chapter 2

# Breast Imaging

## 2.1 Introduction

The breast is an endocrine gland located at the anterior part of the thorax. It contains glandular units called acini or alveoli, which are lined with cells that can produce milk when stimulated by hormones like prolactin [18]. Both men and women have breasts, but the male breast does not have the specialized lobules found in females, as there is no physiological requirement for milk production in males [19].

In an adult woman, the breast (see Figure 2.1) consists of glandular tissue, connective tissue, and fat, which together contribute to its size, shape, and texture. It also includes lobes, lobules, ducts, lymph nodes, blood vessels, and ligaments. Each breast contains several sections called lobules that extend from the nipple. These lobules contain small, hollow sacs called alveoli. A network of slender tubes, referred to as ducts connects the lobules, transports milk from the alveoli to the darker skin area around the nipple, known as the areola. The ducts merge into larger ducts that lead to the nipple. The area around the lobules and ducts contain fat, ligaments, and connective tissue, with the volume of fat having a major impact on breast size. Beneath the breasts are muscles that separate them from the ribs. Blood vessels, including arteries and capillaries, supply oxygen and nutrients to the breast tissue [18] [20].

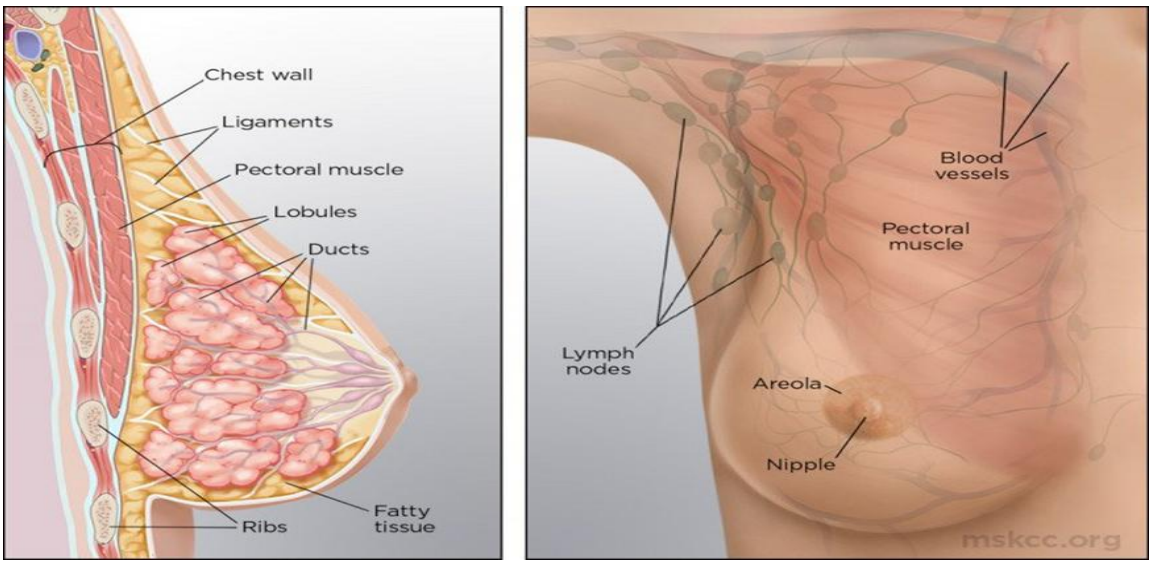



*Figure 0.1: Breast anatomy of adult woman.*

The milk-producing structures in women are quite similar across all individuals. Female breast tissue is responsive to cyclical hormonal changes. As women age, their breast tissue generally shifts to contain more fat compared to dense tissue. The breast itself does not contain muscle tissue. Lymph nodes situated under the armpit, above the collarbone, behind the breastbone, and in other body areas help to trap harmful substances within the lymphatic system and ensure their safe removal from the body [18][21].The majority of tumors arise at the terminal ends of the lobular units within the breast tissue. The glandular tissue is more prevalent in the outer upper segment of the breast; consequently, around 50% of tumors occur in this part of the gland. [18].

## 2.2 Development of the Breast

Human breast tissue starts to develop around the sixth week of fetal development [19]. At first, it grows from the armpits to the groin area by the 9th week, forming two breast buds on the upper chest, which later move back to the chest region. From each bud, columns of tissue grow inward, creating separate sweat glands with ducts that lead to the nipple. At birth, the breasts of both males and females appear similar and are underdeveloped. During early puberty, breasts start to enlarge, and the areola becomes a noticeable bud. In late puberty, the areola becomes more glandular, and the breast accumulates more fatty tissue. Male breasts contain fat and glandular tissue, while female breasts include glandular tissue, acini, ducts, Cooper's ligaments, and Montgomery's glands. The hormone estrogen and progesterone stimulate the development of glandular breast tissue. Factors influencing breast size in women include the amount of breast tissue, age, family history, weight changes, hormonal influences, and pregnancy history.

## 2.3 Breast Pathologies

Breast diseases encompass a variety of conditions, including infections, cysts, and growths, which may or may not be accompanied by pain. Some growths are benign (non-cancerous), while others are malignant (cancerous and capable of spreading) [22][23]. The most common breast conditions consist of breast pain, gynecomastia, cysts, fibroadenomas, fat necrosis, sclerosing adenosis, generalized breast lumpiness, breast tenderness, and breast cancer.

## 2.4 Breast Density

Breast density refers to the ratio of dense tissue to fatty tissue seen on a mammogram. Women with dense breasts possess a greater amount of glandular and fibrous tissue relative to fat, which

can make mammogram interpretation more difficult. This is because both dense tissue and certain abnormal changes, like calcifications and tumors, appear as white areas on the mammogram. Consequently, mammography is less effective in women with dense breasts, increasing the likelihood of missed cancer diagnoses. Additionally, women with dense breasts face a greater risk of developing, but not necessarily dying from breast cancer [24] [25].

Breast density is a significant marker for breast cancer, indicating the potential for detecting abnormalities in mammograms. Detecting tumors in their early stages in dense breasts is challenging because the dense tissue tends to obscure abnormal tissue. To address this issue, regular examinations are highly recommended. Breast density decreases with age, as fibroglandular tissue is gradually replaced by fat tissue. [25].

Byrne et al. studied the mammographic characteristics and breast cancer risk, analyzing nearly 2000 cases. The researchers found that the risk almost doubled in women with some breast density. For women with a density above 75%, the risk increased more than fourfold. Numerous studies have demonstrated that breast density should be considered a criterion for selecting women for screening programs and determining the frequency of exams [25].
Radiologists classify breast density into four categories, as illustrated in Figure 2.2: entirely fatty, scattered fibro-glandular density, heterogeneously dense, and extremely dense [26].

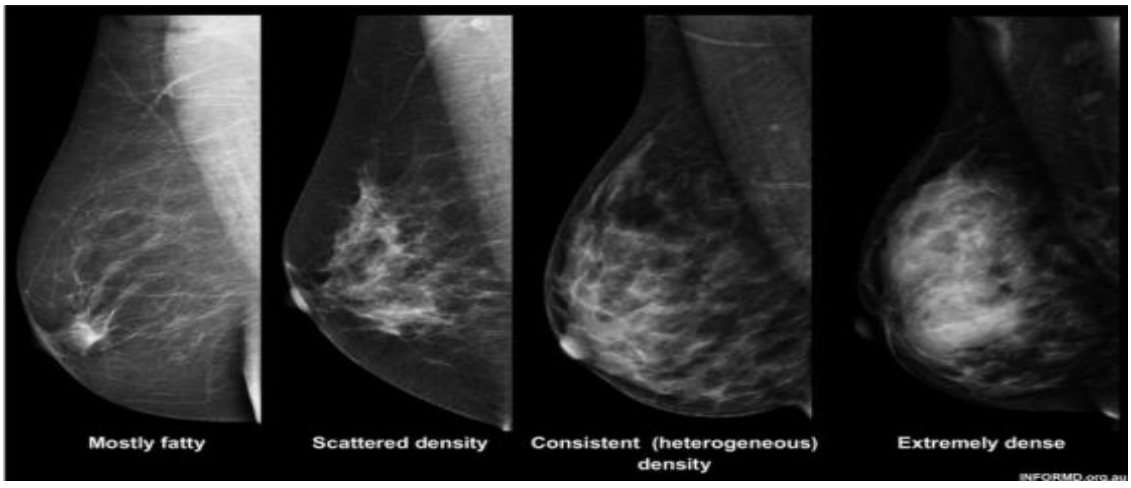



*Figure 0.2: Breast density classification using BIRADS (reproduced from http://www.informd.org.au).*

Classifying and validating breast density includes components such as skin, blood vessels, ductal structures, and stromal elements of the glands, which show up as radio-opaque or white on a mammogram, along with fat that appears radiolucent or black. Entirely fatty breasts are primarily

composed of fatty tissue. In breasts with scattered of fibro-glandular density, most of the breast tissue is non-dense, but some areas are dense. Heterogeneously dense breasts have more than half of the tissue being dense, while the rest is fatty. Extremely dense breasts are highly dense, making it difficult to detect tumors within the tissue [27][26].

## 2.5 Breast Cancer

Breast cancer is a malignancy of breast tissues, most commonly originating from the milk ducts. It occurs when breast cells grow and divide uncontrollably, forming a mass of tissue known as a tumor [26] [27]. Tumors can be classified as benign, which means that the cells do not spread to other tissues in the body, or malignant, where the cells invade surrounding tissues and can disseminate through the bloodstream to other areas of the body through a phenomena called metastasis [25]. As shown in Figure 2.3, some symptoms of breast cancer involve a detectable lump in the breast, changes in the shape or size of the breast, dimpling of the skin, fluid discharge from the nipple, or the appearance of a red patch of skin [6]. Diagnosis can also occur when the cancer is asymptomatic, through breast cancer screening programs like mammograms. Breast cancer outcomes vary based on factors such as cancer type, disease stage, family history, genetics, sex, and the individual's age [7].

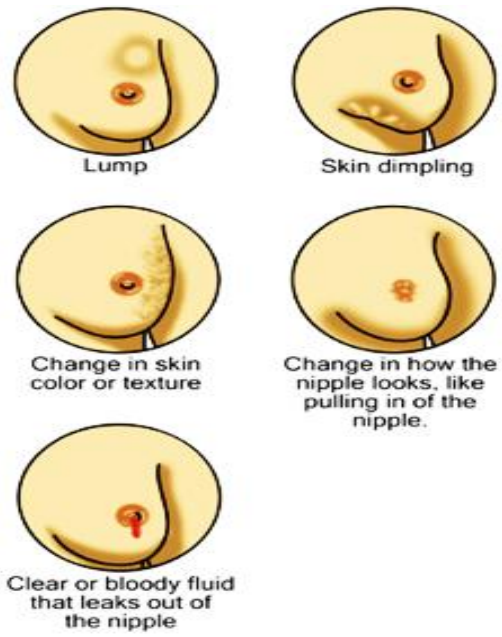



*Figure 0.3 : Symptoms of breast cancer.*

Breast cancer is the most prevalent type of cancer in women, representing 25% of all cases worldwide. It primarily affects women over 40, with incidence rates increasing significantly until menopause[4]. Additional risk factors associated with the development of breast cancer include the age at which menstruation begins and ends, age at first childbirth, family medical history, prior benign breast conditions, and exposure to radiation [25].

The damage caused by cancerous tumors to vital organs can result in serious illness. Therefore, early detection and diagnosis are essential for successful treatment and recovery. Early detection of breast cancer significantly improves prognosis, with five year survival rates reaching 80-90% when diagnosed at initial stages [28]. The American Cancer Society (ACS) states that 1 in 8 women will develop breast cancer during their lifetime, making it the most common cancer among women. Furthermore, research indicates that over 8% of women will face breast cancer in their lifetime, making it the most widespread female cancer in both developed and developing nations.

The rates of breast cancer incidence, mortality, and survival vary across different areas of the world, but the incidence is rising globally. In the absence of early detection programs, mortality rates are on the rise. Early diagnosis and detection are crucial for controlling breast cancer, as the cause of the disease is still unclear. Early cancer detection improves treatment success, saves lives, and reduces costs. Routine screening for breast cancer is vital because many women exhibit no symptoms in the early stages. Most cancers detected during screening are small [29].

The current standard for breast cancer screening and diagnosis method is mammography imaging, which utilizes low energy X-rays in the range of 20–30 keV. Screening mammography is the leading imaging modality for early detection of breast cancer, as it is the only breast imaging approach that has consistently demonstrated a reduction in mortality associated with breast cancer. Mammography is capable of identifying malignancies 1.5 to 4 years before clinical symptoms manifest. Even though mammography is still the gold standard in breast cancer screening, it is still associated with false positive result and low sensitivity. This issue arises from the intermingling of normal fibro-glandular tissues in 2D imaging and the emergence of abnormalities, as shown in Figure 2.4, leading to unnecessary biopsies [33][34].

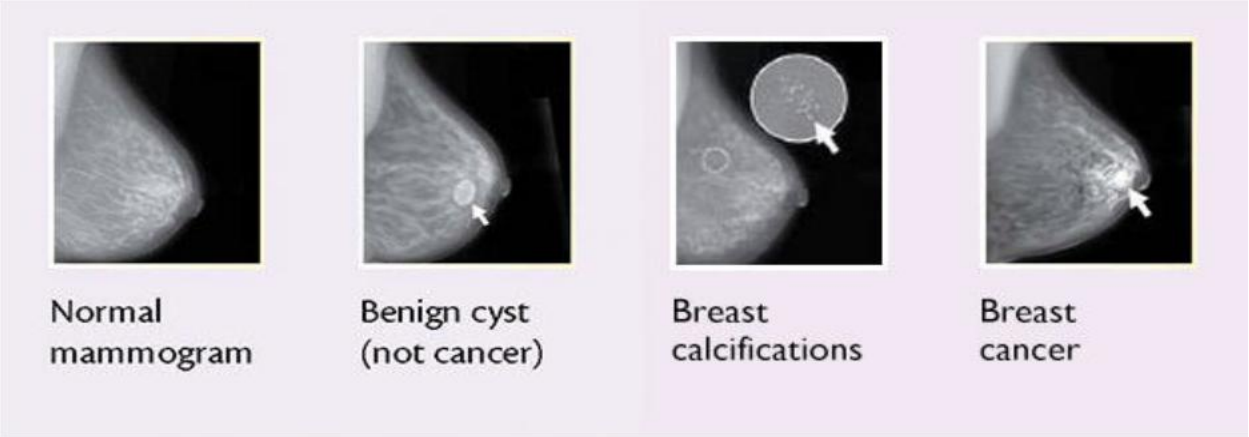



*Figure 0.4 : Normal and abnormal breast images.*

### 2.5.1 Breast Cancer Lesions

Breast cancer is defined by distinct abnormalities, including microcalcifications, tumors, changes in architectural structure and asymmetry between the breasts.

**a) Masses**

Lesions classified as masses are typically harder to identify in mammograms compared to microcalcifications, as their characteristics can resemble those of normal breast tissues (see Figure 2.5). The size, shape, and edges of a mass provide the radiologist with vital information. Generally, mass shapes can be oval, round, irregular, and the edges can vary from well-defined to spiculated.

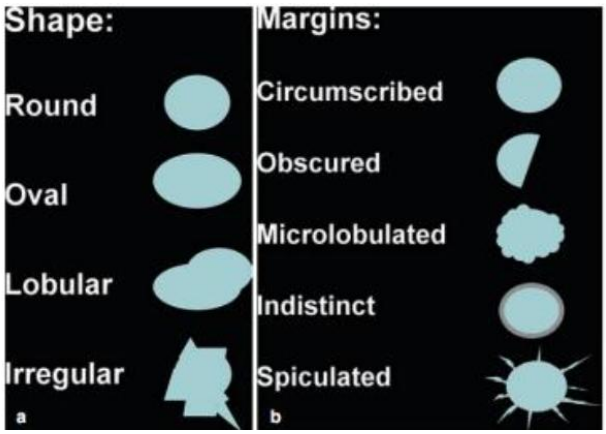



*Figure 0.5 : BI-RADS standardized characterization of masses (Source: Jatoi, 2010).*

When a mass is detected, distinguishing between benign and malignant can be challenging. However, differences in shape and texture characteristics can help. Benign masses typically exhibit

smooth, distinct edges and often have round shapes. Conversely, as illustrated in Figure 2.6, malignant masses have jagged outlines, irregular shapes, and often blurry boundaries. A mass with a regular shape is more likely to be benign, whereas one with an irregular form is more likely to be malignant [26] [34] [35].

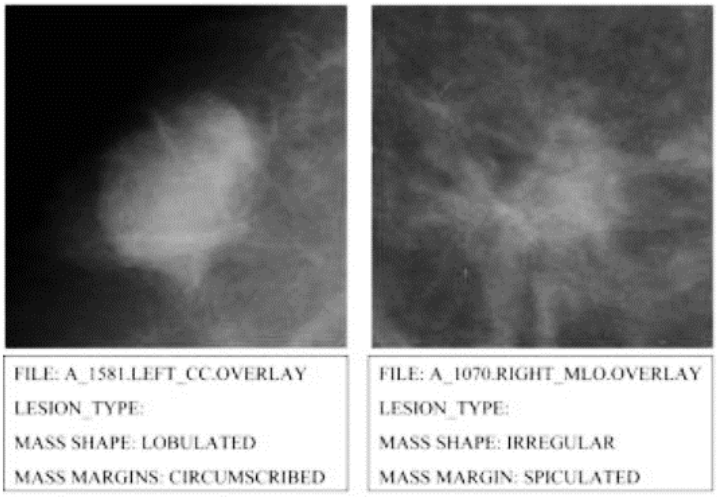



*Figure 0.6 : Two ROI containing malignant masses, extracted from full-sized mammograms with their pathological features (Source: DDSM).*

**b) Microcalcification**

Calcifications refer to small calcium deposits found in breast tissue. They can be categorized into two types: Macrocalcification and Microcalcification. Macrocalcifications appear as tiny white dots on a mammogram. They are typically the result of aging, previous injuries, or inflammation and are generally non-cancerous. In contrast, microcalcifications are smaller deposits of calcium that measure between 0.33 and 0.7 mm and appear slightly brighter than the surrounding tissues. Detecting these lesions through mammography can be difficult because they have low contrast and are small in size, even though they possess high inherent attenuation characteristics. Microcalcifications, which are associated with heightened cellular activity in breast tissue, can present in clusters or specific patterns [34][35][36]. An example of a typical mammogram illustrating microcalcifications from the MIAS database can be seen in Figure 2.7a, with the corresponding ground truth shown in Figure 2.7b.

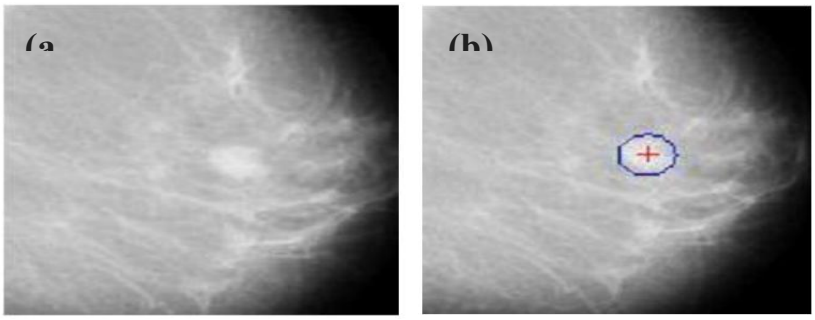



*Figure 0.7: a) Original mammography image b) Ground truth (Source: MIAS Database).*

**c) Asymmetries**

Asymmetries appear as white areas on a mammogram that differ from the usual pattern of breast tissue. Bilateral breast asymmetry signifies a variation between matching areas in the left and right breasts, which can be divided into global and focal asymmetry [33]. Global asymmetry is identified when one breast has a larger amount of fibroglandular tissue than the other within the same area. In contrast, focal asymmetry refers to a distinct area of asymmetry that can be visible in two views and indicates a patch of healthy fibroglandular tissue surrounded by fatty tissue [25].

**d) Architectural Distortions**

An area of breast tissue may appear distorted or drawn toward a specific point, which can result from the positioning of the breast during the mammogram or could be linked to a past injury or procedure involving the breast [33]. The breast's anatomy features various linear structures that create a directional texture visible in mammograms; therefore, changes in the texture of the breast serve as indicators for detecting architectural distortions. These distortions can be classified as either malignant or benign; malignant cases include cancer, while benign cases involve scarring and soft-tissue damage resulting from trauma. Because of its fine presentation and variability in appearance, it is the most frequently overlooked abnormality in instances of false negatives [25].

## 2.6 Types of Breast Cancer

Breast cancer is categorized into two majors types: invasive and non-invasive, each having distinct patterns for classification. Non-invasive breast cancer remains contained and does not infiltrate or spread to surrounding cells. In contrast, invasive breast cancer breaches the tissue barrier and disseminates throughout the breast [34]. Non-invasive breast cancer is further split into ductal and

lobular . As implied by their names, ductal carcinomas develop from the ducts, whereas lobular carcinomas originate from lobules and do not penetrate adjacent tissues [35].

**Ductal Carcinoma in Situ (DCIS)**: is the most prevalent type of noninvasive cancer. The appearance of DCIS on mammograms often features microcalcifications, which are generally easier to detect than masses. In cases where detecting DCIS through mammography proves challenging, research indicates that alternative imaging methods like MRI, which is more sensitive than mammography, should be considered [36].

**Lobular Carcinoma in Situ (LCIS)**: originates in the milk-producing glands and typically appears prior to menopause. This form of carcinoma is often challenging to identify through mammography and during self-examination. Because diagnosing LCIS can be difficult, approximately 15% to 20% of women with this tumor may later develop invasive breast cancer [36].

Breast cancer also includes less prevalent types like inflammatory breast carcinoma and Paget's disease. Inflammatory breast carcinoma (refer to Figure 2.8) is rare and is marked by symptoms such as a brawny appearance, redness, heat, swelling, and changes in the skin that are characterized by erythema, and it has the most unfavorable prognosis among locally advanced breast cancers [25].

*Figure 0.8 : Inflammatory breast carcinoma (Source: Dixon, 2006).*

## 2.7 Breast Mammography Imaging

Mammography is an advanced imaging technique employed to examine the breast, offering comprehensive details regarding its morphology, anatomy, and potential pathologies. It plays a crucial role in the detection and diagnosis of breast cancer and in evaluating mass lesions within

the breast. Early breast cancer detection greatly increases the chances of effective treatment. Although similar to other X-ray procedures, mammography utilizes low doses of radiation, thereby producing high-quality images with excellent contrast and resolution, and minimal noise. Due to the breast's sensitivity to ionizing radiation, it is essential to utilize the lowest radiation dose possible while still achieving excellent image quality [26][42].

Digital mammography is a modern breast imaging technique that employs the same system as conventional mammogram however, replaces the screen-film system with a digital detector. This detector converts the image into an electronic signal, which is then digitized and stored. Studies indicate that digital mammography is particularly more effective than film mammography for patients with dense breast tissue, such as younger women, who typically have denser breasts than older women. Additionally, the technique offers the advantage of a lower radiation dose compared to traditional analog mammography [15].

The National Cancer Institute (NCI) states that there are two categories of mammography: screening and diagnostic mammograms. A screening mammogram produces X-ray images of the breast to identify any changes in women who do not exhibit signs or symptoms of breast cancer. Typically, it consists of two X-rays taken of each breast, making it possible to discover a tumor that may not be palpable [30]. Conversely, a diagnostic mammogram involves X-ray images of breast changes, including lumps, pain, nipple thickening or discharge, or alterations in breast size or shape. Additionally, a diagnostic mammogram is utilized to investigate anomalies found during a screening mammogram. It serves as a fundamental medical tool and is suitable for assessing breast changes, irrespective of a woman's age [38].

Mammography is more accurate in evaluating fatty breast tissue compared to dense breast tissue. Assessing dense breast tissue can be particularly challenging in younger women. Additionally, mammography plays a role in aiding needle core biopsies and helps with the localization of lesions that are not easily felt. In screening mammography, the consistent compression of the breast is crucial for ensuring image clarity, which means these tools need to be highly sensitive to correctly identify potentially malignant tumors. Conversely, diagnostic mammography typically requires more time, costs more, and exposes the patient to a higher radiation dose than screening mammography; therefore, diagnostic tools must have excellent specificity to effectively detect malignant tumors.

To evaluate variations in breast tissue density, mammography is performed using two standard views: craniocaudal (CC) and mediolateral oblique (MLO). The MLO view typically captures more breast tissue compared to the CC view due to the chest wall's slope and curvature. In the CC projection (refer to Figure 2.9a), it is essential to include all breast tissue, excluding the axillary portion. When positioning, ensure the upper and posterior parts of the breast are included by elevating the breast within its natural range of motion. The MLO projection (refer to Figure 2.9b) should maximize breast tissue visualization, capturing the free edge of the pectoral muscle. It is also crucial to position the patient as far over the detector as possible to include the breast's tail in the image.

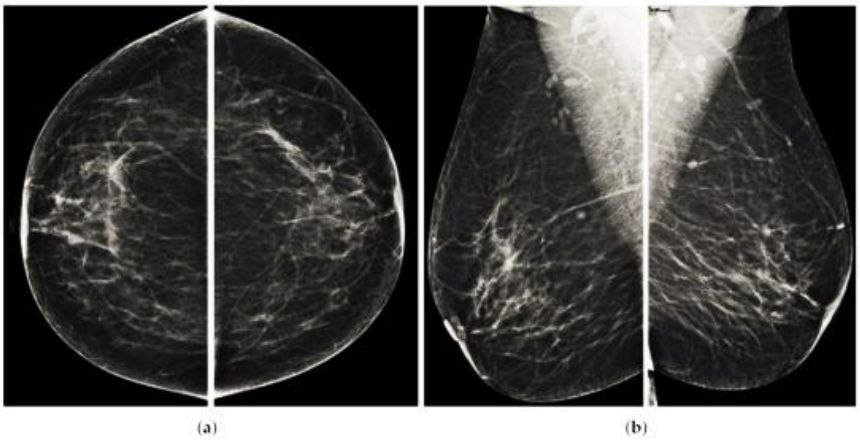



*Figure 0.9 : Two mammographic projections: (a) Cranio Caudal view; (b) Medio Lateral Oblique vie (Source: Kopans, 2007).*

When localized lesions are necessary, techniques for segmenting the breast into smaller regions are essential. The quadrant method utilizes the nipple as a point of reference for dividing the breast. Figure 2.10 illustrates two variations of this technique.

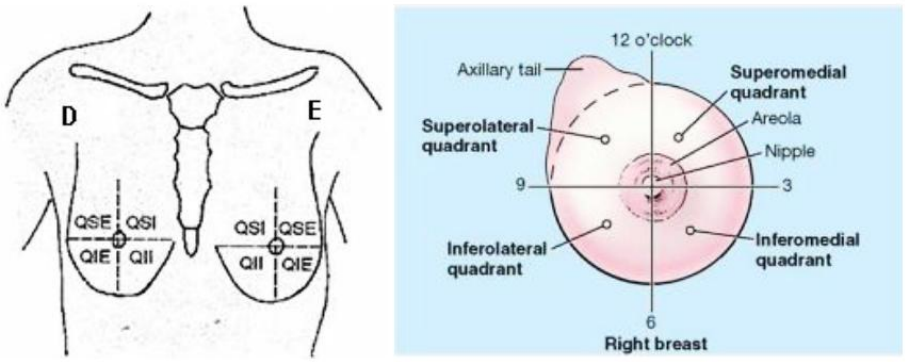



*Figure 0.10 :(a) quadrants method (b) naming of each quadrant (Source: Kopans, 2007; Moore and Agur, 2003)*

Mammography has significantly progressed from the era of direct-exposure films and widespread quality issues. Today, breast imaging radiologists have access to a wider array of tools and are more involved in patient care than ever. With ongoing technological advancements and skill development among radiologists, mammography is anticipated to maintain a key position in early detection initiatives [15].

## 2.8 Mammography System: Components, Technical Principles & Procedures

The fundamental parts of a mammography machine include a X-ray tube, an auxiliary filter, a breast compression device, an anti-scatter grid, image receptor, automatic exposure control (AEC) and detector system (see also Figure 2.11) [39].

The mammographic examination begins with positioning the patient's breast on a specialized imaging platform, where it is gently compressed between a fixed support plate and a movable paddle. A precisely calibrated X-ray tube then emits a controlled burst of low-dose radiation that penetrates the breast tissue and is captured by a detector system on the opposite side. This detector then converts the X-ray patterns into electronic data for computer processing and display. The resulting images are referred to as mammograms, provide radiologists with detailed visualization of breast architecture for diagnostic evaluation [39].

***X-ray Tube:*** The mammography equipment, like all other X-ray units, consists of two main components: a generator and an X-ray tube. The high voltage generator converts low voltage (200-240v) into high voltage (20-40kv) to power the X-ray tube. This high voltage is essential for the

X-ray tube to produce X-ray photons (X-radiation), as illustrated in Figure 2.12 [39]. A significant distinction between mammographic and traditional radiographic tube functions is the use of a lower operating voltage, typically under 35 kVp, in mammography. Mammography specific tubes consist of a vacuum-sealed glass enclosure, containing a heated filament cathode for electron emission and an anode target for X-ray production, along with a window that allows the beam to pass through. The characteristics of the emitted radiation are influenced by the materials used for the anode. The applied kilovoltage potential between cathode and anode governs the maximum photon energy capable of penetrating breast tissue. This optimized low-energy photon production is crucial for achieving the high soft-tissue contrast resolution required in breast imaging [40].

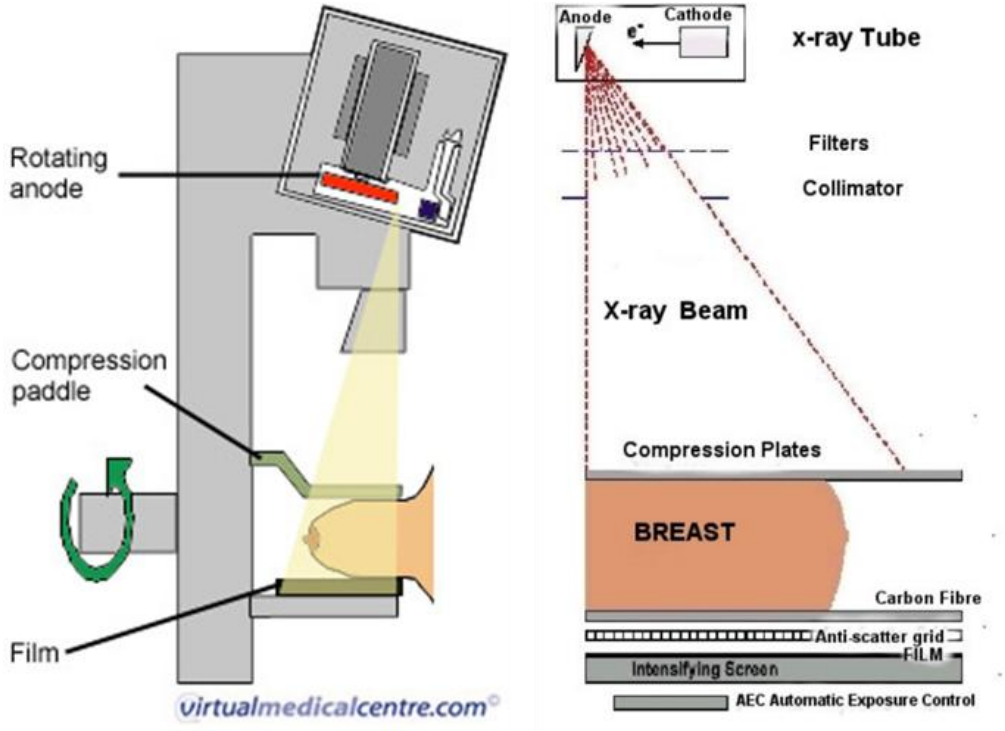



*Figure 0.11: Components of the mammographic imaging system.*

Collimators and filters are essential for controlling and directing the radiation output. The application of low voltages in the tube decreases the beam's energy, thereby reducing the radiation dose to the patient. Image focus is attained by using dual-filament design in the X-ray tubes: one with a large focus of 0.3 mm and another with a small focus of 0.1 mm. These filaments are typically made from high melting point materials like tungsten, molybdenum, or rhodium.

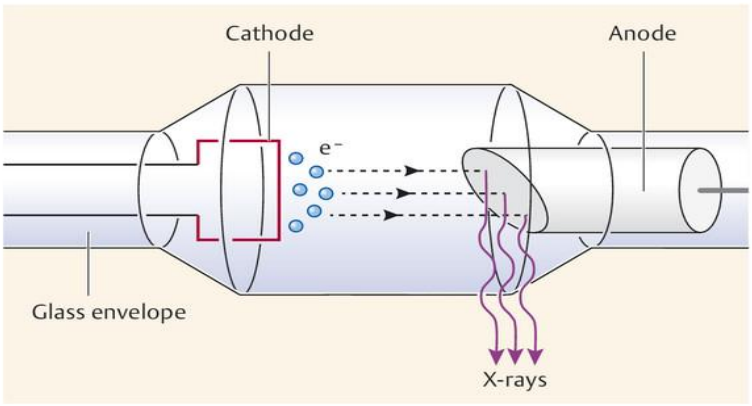



*Figure 0.12: X-ray tube generating X-ray photons.*

***Auxiliary Filter:*** These filters are strategically positioned based on individual breast characteristics, such as thickness and density, to improve contrast resolution while reducing patient radiation dose. Tailoring the anode-filter combination for each individual can enhance image contrast while minimizing radiation exposure (refer to Figure 2.13). Molybdenum and rhodium are the filter materials most frequently utilized in mammography [40].

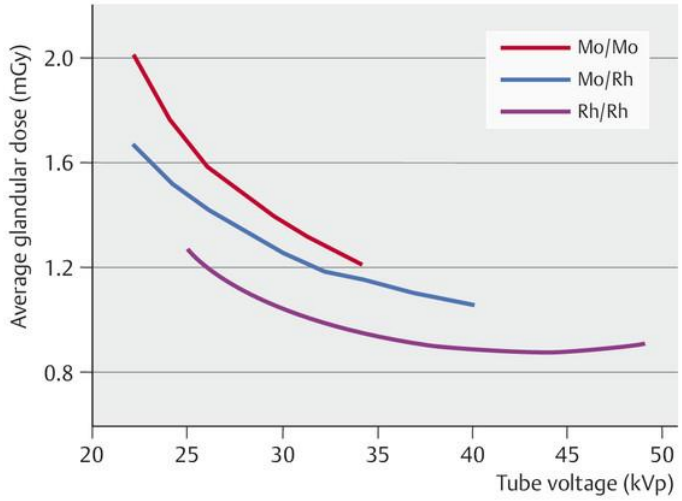



*Figure 0.13:The effect of various anode-filter pairings on radiation exposure (molybdenum–molybdenum, molybdenum–rhodium, rhodium–rhodium; with the first term indicating the anode material and the second representing the filter material).*

***Compression Plates:*** are designed to prevent resolution loss due to patient movement, using the shortest possible exposure time. By applying uniform compression, they reduce breast thickness, which allows for better X-ray beam penetration. This technique helps minimize superimposition

and decreased geometric unsharpness, while also improving resolution by decreasing the distance to the target area. Additionally, it enhances image contrast and reduces superimposition from different planes.

***Anti-Scatter Grid:*** Consisting of precisely aligned lead (Pb) strips separated by a stiff interspace material, are positioned between the breast compression plate and image receptor. The design of the anti-scatter grid includes suitable widths and spacings to mainly absorb scattered radiation while allowing primary radiation to pass through unimpeded. Anti-scatter grids help prevent a decrease in image contrast caused by scattered radiation reaching the image receptor. Static grids are composed of thin slices that have a lower resolution than the system, ensuring that they remain invisible in the image.

***Image Receptor:*** The image receptor in mammography can be a fluorescent screen, a photosensitive phosphor plate, or a charged electron device, depending on the type of mammography being performed. Digital mammography utilizes an image receptor that converts X-ray information into a digital picture, while screen-film mammography employs a fluorescent screen held in a light-tight cassette with a single-emulsion photographic film. The fluorescent screens, along with single-coated radiographic film, enable X-rays to penetrate the cassette cover and interact with the screen. Maintaining optimal screen-film contact is critical for preserving spatial resolution. The phosphor crystals in the screens absorb the energy and isotropically emit visible light, which exposes the adjacent film. Higher film sensitization velocities correlate with reduced radiation doses to the patient. However, films that are faster often exhibit more noise. The sensitization speed depends on the thickness of the film; thinner films are slower but provide superior resolution. Recent mammography films offer enhanced resolution, greater speed, and reduced noise.

***Automatic Exposure Control (AEC):*** The AEC system is employed to prevent incorrect exposure. Due to significant variations in breast composition among individuals, the radiation dose required can differ from individual to individual. In analog mammography, an AEC device is situated beneath the detector to measure the incoming dose and calculate the optimal individual dosage. Upon reaching the desired dosage, the exposure is brought to an end. The AEC device automatically compensates for absorption variations caused by breast anatomy, thereby optimizing image quality and minimizing radiation dose. Its sensors regulate exposure time based on tissue density and thickness, allowing for precise and consistent imaging.

***Detector:*** For many years, film or film-foil systems have been the standard for detecting and visualizing X-radiation in conventional mammography. In the analog method, the image-capture system consists of a cassette that contains the mammographic film and an intensifying screen. The quality and processing of the film play a significant role in influencing the character and quality of the image. The films utilized in this technique are specialized orthochromatic films known for their high contrast and sensitivity [40]. The first fully digital mammography systems employed charge-coupled device (CCD) detectors, which offered quantum efficiency improvements but were limited by their small field of view. The early 21st century saw two parallel advancements, such as CR cassettes compatible with existing analog mammography units and the development of direct digital detectors utilizing amorphous selenium (a-Se) that converted X-rays directly into electrical signals without light conversion. Modern mammography systems now utilize flat panel detector technology, where the digital detector is fully integrated into the system.

## 2.9 Other Technologies for Breast Imaging

Advanced breast imaging leverages complementary modalities to address diagnostic challenges. Ultrasound (US) represents another promising modality in breast imaging, notably advantageous in distinguishing between cystic formations and cancerous masses that lack calcification. Additionally, Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) has demonstrated exceptional sensitivity in detecting invasive breast cancer and is not constrained by breast tissue density. Nonetheless, this technique falls short in providing the resolution necessary for identifying susceptible masses and calcifications, as seen with mammography. When appropriately combined with other methods, MRI can offer significant benefits. Regarding molecular imaging of the breast, scintimammography emerges as the superior technique. It involves the use of a gamma camera to capture breast images of patients injected with radioisotopes. This method proves particularly useful for patients with dense breast tissue and those who have undergone mammary surgery. Another technique, galactography/ductography, entails the use of contrast radiologic agents to image the breast ducts, aiding in diagnosing abnormal nipple discharge and intraductal papillomas.

### 2.9.1 Breast Ultrasound

The clinical application of breast ultrasound traces its origins to 1951, when Wild and Neal pioneered A-mode sonography to differentiate benign and malignant breast masses [41]. This landmark study sparked subsequent technological innovations, including the development of automated systems and multi-transducer arrays designed for comprehensive breast evaluation. These advancements aimed to address two key clinical needs such as detection of occult breast pathology (screening applications) and improved diagnostic characterization of palpable abnormalities. As the resolution and quality of ultrasonographic imaging have improved, the application of ultrasonography for breast scanning has increased [42].

Breast US utilizes high-frequency sound waves to generate a computerized detailed cross-sectional image of breast anatomy, as illustrated in Figure 2.14. It can reveal specific alterations in the breast, such as fluid-filled cysts, particularly in identifying fluid-filled cysts that often appear indeterminate on mammography [43]. It is used when:

- It can effectively examine palpable lumps that are detectable by touch but not clearly visible on mammogram images, as well as detect subtle changes in women with dense breast tissue that might obscure potential issues
- To reliably distinguish between fluid-filled cysts (which are typically benign) and solid lesions (which may require additional evaluation to rule out malignancy)
- To assist in directing a biopsy needle toward a specific location to extract and examine cells for cancer. This procedure can also be performed on swollen lymph nodes located beneath the arm.

  US is readily accessible, simple to obtain, and does not subject an individual to radiation. Additionally, it is more affordable compared to many other alternatives.

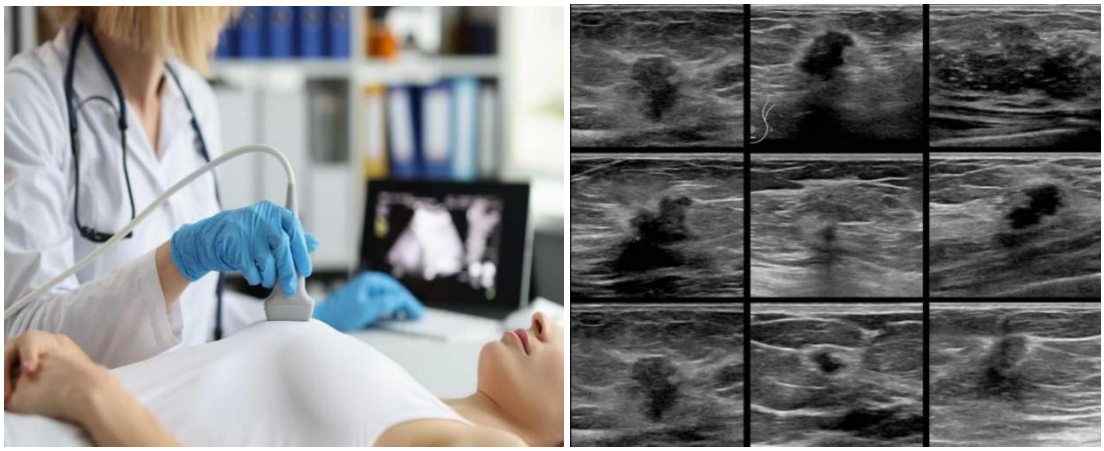


*Figure 0.14 : Breast ultrasound imaging (https://radiologyassistant.nl/breast/ultrasound/ultrasound-of-the-breast).*

### 2.9.2 Breast MRI

MRI is a technique primarily used in medical settings to generates high-resolution soft tissue visualization making it particularly valuable for breast assessment (see Figure 2.15). Breast MRI employs radio waves and powerful magnets to create detailed images of the interior of the breast. The ACS recommends MRI as a supplemental screening tool for high-risk populations for developing breast cancer [48][49][50]. It is used:

- ***To help determine the extent of breast cancer:*** Breast MRI is sometimes used in women who already have been diagnosed with breast cancer, to help measure the size of the cancer, look for other tumors in the breast, and to check for tumors in the opposite breast. But not every woman who has been diagnosed with breast cancer needs a breast MRI.
- ***To screen for breast cancer:*** For certain women at elevated risk for breast cancer, particularly those with a significant family history of the disease or those who carry genetic mutations linked to hereditary breast cancer, an annual mammogram accompanied by a screening MRI is advised. An MRI is not advised as a standalone screening method since it may overlook some cancers that a mammogram would uncover.

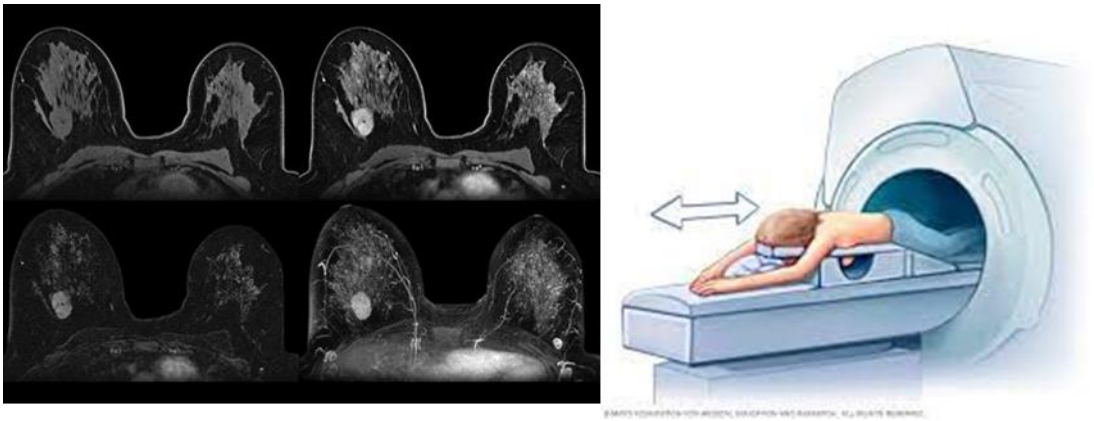


*Figure 0.15: Breast MRI imaging (Source: https://www.mayoclinic.org/tests-procedures/breast-mri/about/pac-20384809)*

While MRI can detect some cancers not visible on a mammogram, it is also more likely to identify false positives, which can lead to unnecessary tests and biopsies. Due to these limitations, current guidelines reserve MRI as an adjunct screening tool specifically for high-risk populations not as a standalone replacement for mammography in average-risk women. Despite being a highly sensitive test, a breast MRI can still overlook some breast cancers that a mammogram might catch [45].

### 2.9.3 Scintimammography

Scintimammography is a new and advanced imaging technique used to detect cancerous tissue in the breast. It involves administering an intravenous administration of Tc-99m Sestamibi, a tumor-avid radiopharmaceutical followed by imaging with a Gamma Camera. Unlike mammography, this method does not require breast compression. The radiopharmaceutical accumulates in cancerous tissue, making it easily visible in the images. Cancerous tissue has up to nine times greater affinity for this agent compared to normal breast tissue [51][52].

Research indicates that scintimammography boasts a sensitivity and specificity exceeding 85% and 90% [49] respectively while X-ray mammography offers only 60% and 80% [50]. As Nuclear Medicine facilities become increasingly accessible, a growing number of patients are anticipated to gain advantages from this innovative imaging technique [53][54].

Scintimammography is not a substitute for X-ray mammography in the standard screening of breast cancer patients. Nevertheless, scintimammography can serve as a valuable method in circumstances such as uncertain or non-conclusive mammograms, dense or irregular breast tissue,

breasts with scarring from past surgeries, breasts with implants where mammography poses challenges due to painful compression. It is also useful during hormone replacement therapy (HRT), for detecting suspected multicentric or multifocal disease, evaluating palpable masses with negative or inconclusive biopsy results, assessing dense masses without calcifications, resolving discrepancies between clinical findings and mammographic results, and guiding biopsy when ultrasound-guided biopsy is not an option [55] [56].

### 2.9.4 Breast Biopsy

When other tests indicate the presence of breast cancer, a biopsy is performed. However, needing a breast biopsy, as illustrated in Figure 2.16, does not necessarily mean there is cancer. Most biopsy results are benign, but a biopsy is the only definitive way to determine if cancer cells are present. During the procedure, a doctor removes small tissue samples from the suspicious area for laboratory examination to check for cancer cells [57] [58].

There are various types of breast biopsies. Some procedures involve a hollow needle, while others require making an incision in the skin. Each method presents unique benefits and limitations.

- A fine-needle aspiration (FNA) biopsy, utilizes an exceptionally thin, hollow needle (even finer than standard phlebotomy needles) connected to a syringe to extract cellular samples from suspicious lesions [57].
- A core needle biopsy utilizes a larger-gauge needle compared to FNA to obtain tissue samples from suspicious breast abnormalities detected through physical examination or imaging modalities (ultrasound, mammography, or MRI). This diagnostic technique is particularly valuable when malignancy is clinically suspected.
- In limited cases where less invasive methods are inconclusive; a surgical biopsy may be performed to excise the suspicious lesion entirely. Usually, the surgeon removes the entire mass or abnormal area along with a surrounding margin of normal breast tissue.

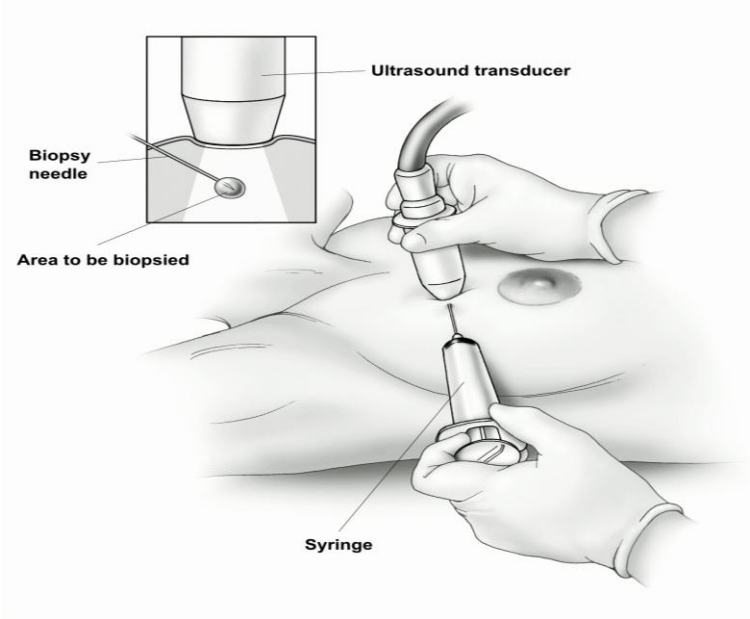



*Figure 0.16: Fine needle aspiration using ultrasound.*

Regardless of the type of biopsy is performed, the samples will be sent to a laboratory where a pathologist, a specialized physician, will examine them. Generally, it takes a minimum of several days to receive the results [56].

# Chapter 3

# Methodology

## 3.1 Introduction

Research on breast cancer screening is often centered on mammography since it is the most useful imaging tool for early tumor detection. Nevertheless, in order to diagnose patients more accurately, radiologists must continually examine images in great detail, which emphasizes the value of computer-aided diagnosis (CAD), which acts as an additional reviewer of mammographic images. CAD systems typically rely on a number of techniques, such as pretreatment phases, preprocessing, image enhancement, segmentation, extraction of classification parameters, and, in the end, the interpretation and categorization of possible anomalies. In this chapter the method used for mammogram classification will be presented.

## 3.2 Data Set

In this study, mammography image data set is obtained from mini-MIAS database, which is freely available. The dataset comprises a total of 322 clinical MLO view mammogram images. Of these, 202 were normal cases, while the remaining 120 exhibited abnormal breast tissue (lesions). This lesion data set includes 69 benign and 51 malignant masses. These mammograms are digitized with resolution of 1024×1024 pixels and each pixel is stored as 8 bits. The proposed algorithm is implemented on a MATLAB 2020 platform, Lenovo computer (2.66GHz/4GB RAM). Based on some inclusion-exclusion criteria (mainly based on visual data quality assessment), a total of 274 breast mammogram images with 172 normal and 102 abnormal images were used.

## 3.3 Preprocessing

Preprocessing is a crucial step in image processing as it helps with region localization for anomaly detection. The accuracy of this step significantly influences the likelihood of success for subsequent steps, such as segmentation and classification. Clinical images often exhibit characteristics such as unrecognized noise, low contrast, inhomogeneity, weak boundaries, and unrelated parts, as shown in Figure 3.1. These challenges can be addressed by pre-processing techniques.

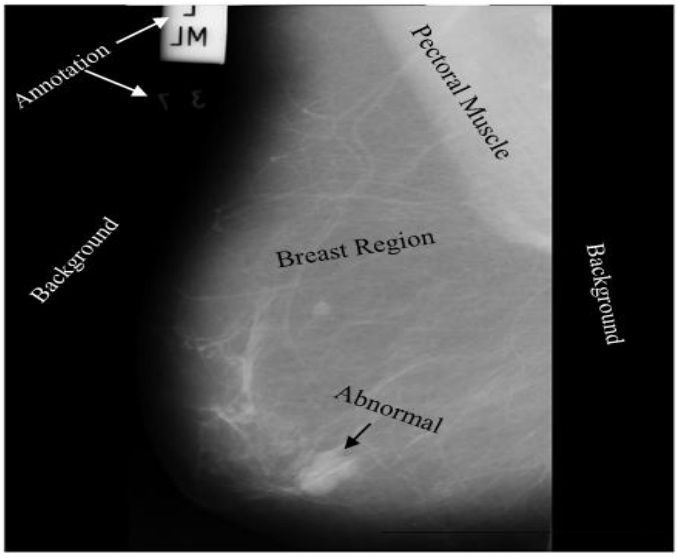



*Figure 0.1 : A typical mammogram image of the breast.*

The flowchart presented in Figure 3.2 shows the different pre-processing steps utilized in the current study to reduce the effect of some unwanted objects within the mammogram images before further image analysis. The steps include background removal (unnecessary black area and special mark removal), removal of pectoral muscle, filtering (noise removal) and image enhancement (improve the visual effect) [58]. Pre-processing is believed to aid in region localization for anomaly search and improves the accuracy of detection of early signs of breast cancer.

### 3.3.1 Background Removal

In mammogram images, there are different kinds of unwanted background objects including annotations, labels, and background noises outside the breast region, which affects the abnormality detection process. Moreover, there is an increased risk of noise-induced image distortion when processing images for enhancement and when thresholding is done. There are therefore flaws in the image's structure. Thus, in this study morphological operations are used in order to remove these unwanted background artifacts [59] [60].

Morphology encompasses a broad range of image processing techniques that focus on shapes. Within image processing, morphology pertains to the characterization of an object's shape and structure. Morphological techniques utilize a structuring element on an input image, resulting in an output image of the same dimension. When morphological operations are performed, the value of every pixel in the final image is determined by comparing it to its neighbors in the corresponding

input image pixel. It is possible to create a morphological operation that is responsive to particular shapes in the input image by selecting the neighborhood's size and form.

Two major morphological processes are of interest in the current study: dilation and erosion. These two fundamental processes are the foundation for the other two morphological operations known as opening and closing that are employed in the current study [61].

**Dilation:** The dilation operation (Eq. 3.1) increases the size of an object. Generally, the dilation equation of set A by B is given by [59]:

$$F \oplus B = \left\{ y | \left(\hat{B}\right)_y \cap F \neq \emptyset \right\} \quad \text{(Eq. 3.1)}$$

Given a structuring element B, if we reflect B about its origin and shifted by $y$, then the dilation of $F$ by $B$ is the set of all displacements $z$ such that $\hat{B} = \{-b : b \in B\}$, which represents the mirrored version of B, has at least one overlapping element with F. Dilation merely increases the number of pixels to the edges of boundary elements.

**Erosion:** The erosion operation (Eq. 3.2) reduces the size of an object. The erosion of set $F$ by set $B$ is given by [59]:

$$F \ominus B = \left\{ y | \left(\hat{B}\right)_y \subseteq F \right\} \quad \text{(Eq. 3.2)}$$

The process of eroding image F with structuring element B involves identifying all points y for which the translated version of structuring element B fits entirely within the image. This operation leads to the reduction of boundary pixels associated with the object.

**Opening:** is a combined operation involving both erosion and dilation. The opening of an image (Eq. 3.3) F using the structuring element B is defined as [59] [61]:

$$F \circ B = (F \ominus B) \oplus B \quad \text{(Eq. 3.3)}$$

This definition indicates that the opening operation involves eroding the image and then dilating the result. This process creates new boundaries where the structuring element B makes contact with the outermost edges of image F as B traverses along F perimeter. The final opened image is constructed through union operations that aggregate all valid contact points between B and F boundaries.

**Closing:** The closing of an image (Eq. 3.4) entails a mix of dilation and erosion, contrasting with opening in the sequence of these operations, which is also a composite function of dilation and erosion. The key distinction between the closing and opening operations lies in the order in which erosion and dilation occur. The closing of an image F with respect to the structuring element B is given by [59] [61]:

$$F \cdot B = (F \oplus B) \ominus B \qquad \text{(Eq. 3.4)}$$

This equation shows that the closing operation starts with dilating the image FF by structuring element BB, followed by eroding the result with the same structuring element. The boundary of the closed image consists of points in BB that reach the extreme points of FF's boundary as BB moves around the outside of FF's boundary. In this study, morphological reconstruction (opening followed by closing) is used to enhance the image, making the foreground, which includes breast tissue, more recognizable.

### 3.3.2 Removal of Pectoral Muscle

It is crucial to separate the breast region from the pectoral muscle region in order to facilitate further processing and ensure an accurate computer-aided diagnosis. The pectoral muscle is visible in the upper left corner of the mammography's MLO view, exhibiting a somewhat greater intensity relative to the adjacent breast tissue. In the current study, muscle region on the mammograph is segmented and removed using region (seed) growing algorithm. Region growing is a process that uses preset criteria to group pixels or sub-regions into larger regions. The fundamental tenet of region-growing techniques is that adjacent pixels within a region have comparable values i.e. regions on the breast with similar intensity value. Each region in the mammogram has a different intensity value. The muscle, fatty tissues, glands, lobules and the ducts display different intensity values and thus can be segregated into different regions. A typical step involves comparing a single pixel with its surrounding pixels. Pixels can be designated as belonging to the cluster if one or more of their neighbors meet a set of similarity criteria. In every case, noise affects the results, and the choice of the similarity criterion is important [64][72][73].

This method accepts the image as input along with a set of seeds. Every object that is to be segmented is marked by a seed. Through comparison with all unallocated neighboring pixels, the regions are grown iteratively. Similarity is measured as the difference between the intensity value

of a pixel and the mean of the region. The pixel belonging to the corresponding region is the one with the least difference, as determined by this method. Until all pixels are assigned to a region, this process is repeated. The growing of a region ceases once there are no additional pixels that meet the criteria for being included. Since seeded region growing relies on seed points as extra input, the segmentation outcomes are influenced by the initial choice of seeds, and noise within the image can lead to poorly positioned seeds [64].

**Algorithm to Detect the Pectoral Muscle:**

Step 1: Locate starting seeds location; choose left-top-most pixel of the breast region.

Step 2: Mark objects to be segmented using the seed.

Step 3: Measure similarity by calculating the difference between the neighbor pixel value and region mean value.

Step 4: Pixels with similarity are assigned to the respective region.

Step 5: Repeat the above steps until all pixels are assigned to a region.

### 3.3.3 Filtering

There are different types of noises on the images obtained from miniMIAS datasets. High intensity noise is marked by elevated optical density values, which can include labeling or scanning artifacts. Tape artifacts consist of marks left by tape or other shadows appearing as horizontal streaks. These types of noise should be substituted with black pixels [64]. In the current study, two known filters have been used to deal with possible noises in the mammograms.

**Median Filter:** is a nonlinear technique commonly used in image processing to reduce salt-and-pepper (impulse) noise and spot noise [65]. It is commonly used in digital image processing because it can preserve edges while reducing noise under certain conditions. The filtering process involves sliding a window over the image, with the median value from the window replacing the center pixel in the output image. When dealing with Laplacian noise distribution, the median serves as the maximum likelihood estimator of the location. In areas that are relatively uniform, the median filter successfully estimates the gray-level value, particularly in the presence of long-tailed noise. As an edge is crossed, one side of the window becomes more dominant than the other,

causing the output to switch sharply between values, thereby preventing any blurring of the edge [66].

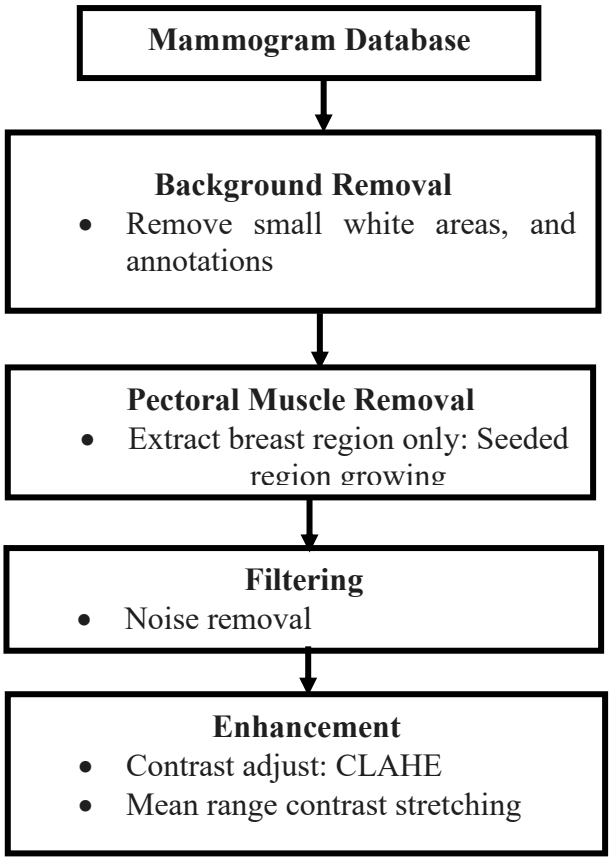



*Figure 0.2: Mammogram pre-processing flow chart.*

The primary drawbacks of median filters include the creation of new flag points that do not exist in the original signal, which may be undesirable in certain applications. Additionally, the median filter tends to remove both noise and fine detail. Smaller elements within the neighborhood size have minimal influence on the median value and will be filtered out. Consequently, the median filter cannot effectively differentiate fine details from noise. In Matlab, the syntax J = medfilt2(I, [m n]) performs median filtering, whereby each output pixel represents the median value in the m-by-n neighborhood surrounding the corresponding pixel in the input image.

**Wiener Filter:** The wiener filter is a spatial domain linear filter, which uses statistical approach. It compares the image's noise level to the desired noiseless signal estimate in order to limit the amount of noise present. The wiener filter does not smooth images much in areas with high local image variance [67][68]. Wiener filter estimates the local mean and variance at each pixel during

operation. For an image $I$ over a window $W$ of size X-by-Y, the local mean (Eq. 3.5) is calculated as:

$$\mu = \frac{1}{XY} \sum_{x,y \epsilon W} I(x,y) \quad \text{(Eq. 3.5)}$$

and the local variance (Eq. 3.6) is calculated as:

$$\sigma^2 = \frac{1}{XY} \left( \sum_{x,y} (I(x,y) - \mu)^2 \right) \quad \text{(Eq. 3.6)}$$

Now the weiner filterd (Eq. 3.7) image is given by:

$$f(x,y) = \mu + \frac{\sigma_W^2 - \upsilon^2}{\sigma^2} (I(x,y) - \mu) \quad \text{(Eq. 3.7)}$$

where $f(x,y)$ represents the filtered image, $\mu$ is the local mean, $\sigma_W^2$ is the local variance, $\sigma^2$ is the global variance and $\upsilon^2$ is noise variance [69]. The MATLB 'wiener2' function designed to perform Wiener filtering based on a specified window size.

In the current study, a filtered image was created by multiplying the median and wiener filtered outputs of that specific image.

### 3.3.4 Image Enhancement

Digital mammographic image preprocessing frequently involves enhancing the image to improve its quality and visual appeal while retaining important details. Using suitable image enhancement techniques on digital mammograms can aid in better detection of breast cancer, thereby increasing survival rates.

One way of enhancing mammogram images is by adjusting the image contrast, which remaps the intensity values to cover the full range of the data type. A well-contrasted grayscale image displays distinct variations between black and white. Therefore, increasing contrast deepens dark areas while brightening light regions, improving the visibility of lesions against the background [70]. In the current study, two popular enhancement techniques have been used for the purpose of image enhancement: mean range contrast stretching and contrast limited adaptive histogram equalization.

**Mean Range Contrast Stretching**: Contrast stretching is employed to enhance the range of gray levels in an image. For instance, in an 8-bit system, the display can represent a maximum of 256 shades of gray. If the gray levels in the captured image are confined to a narrower range, the images can be improved by broadening the range of gray levels. This technique is referred to as contrast stretching.

In order to carry out the mean range contrast stretching, the formula indicated in Eq. 3.8 has been used, which adjusts intensity values with poor contrast due to glare. The normalization maps the intensity levels of the mamograms into the range $[g_{min}, g_{max}]$. The gray level normalization formula given below.

$$g(i,j) = g_{min} + \frac{(g_{max} - g_{min})(g_0(i,j) - g_{0\,min})}{(g_{0\,max} - g_{0\,min})} \qquad \text{(Eq. 3.8)}$$

where $g_{0\,min}$ and $g_{0\,max}$ denote the lowest and highest intensity levels of the original image, $g_{min}$ and $g_{max}$ represent the lowest and highest intensity levels of the normalized image, and $g_o(i,j)$ and $g(i,j)$ are the gray levels at the coordinates $(i,j)$ prior to and following normalization. In this study, $g_{min}$ is the average image intensity of the given image and $g_{max} = g_{0\,max}$.

**Contrast Limited Adaptive Histogram Equalization (CLAHE):** In this study, to further enhance the images, CLAHE was used. This method refines Adaptive Histogram Equalization (AHE), a specific version of histogram equalization that adaptively processes the images for enhancement. The histograms of the sub-images are clipped to control the level of enhancement for each pixel before being equalized. As a result, the image details become clearer in relation to the background. Simultaneously, the background is enhanced in harmony with the foreground, resulting in a high-contrast output image. The enhancement is minimized in highly uniform areas of the image, which helps prevent excessive noise enhancement and mitigates the edge-shadowing effect associated with unrestricted AHE [70][71]. MATLAB's adapthisteq(I) was used to implement the CLAHE operation.

In the current study, the final enhanced image was taken by multiplying the mean contrast stretched image and the CLAHE applied image.

## 3.4 Segmentation for Breast Region of Interest Extraction

Intuitively speaking, segmentation divides an image into its constituent regions or objects. After removal of the pectoral muscle, breast region of interest (BROI) is segmented. Moreover, mammogram images have a lot of dark areas that aren't breast tissues. These areas have been masked/removed to speed up processing and minimize artifact. The mask template is a binary matrix of size equal to the original image and this mask preserves breast tissue region. The pre-processed mammogram images are inputs to the segmentation algorithm. A mammogram showcases two distinct areas: the breast area that is exposed and the non-breast area filled with air. The primary characteristic of a mammogram is the outline of the breast, commonly referred to as the skin-air interface or breast border, along with the pectoral muscle. The breast area can be extracted by dividing the mammogram into the breast region, the non-breast area, and the pectoral muscle. By isolating both the non-breast area and the pectoral muscle, we can delineate the breast region for subsequent analysis [64] [74].

When applied to breast images, image segmentation techniques extract the region of interest (ROI), effectively separating breast masses from the adjacent tissue. Traditional segmentation methods such as simple thresholding, edge-based methods like the Canny filter, and region growing have been widely employed in various medical image segmentation applications. In the current study, four segmentation methods have been utilized and compared for their efficacy in accurately segmenting the region of interest by comparing their performance against the available gold standard/ground truth information which comes with the breast mammogram images acquired from the mini-MIAS database. The four methods were global thresholding, Otsu segmentation, watershed segmentation and morphological segmentation methods [73].

Note that, in the current study, global thresholding was carried out by calculating a single threshold value which is equal to the total mean of the input image. A single threshold value is applied to the entire image, classifying pixels with higher values as foreground and pixels with lower values as background. In this technique a chosen threshold value $T$ is compared to the intensity value of each pixel in the image. If the pixel intensity is greater than or equal to $T$, then the pixel is classified as part of the foreground (usually assigned a value of 1 or white). If the pixel intensity is less than $T$, then the pixel is classified as part of the background (usually assigned a value of 0 or black). Mathematically, this operation can be expressed as:

$$I(x,y) = \begin{cases} 1, if\ f(x,y) \geq T \\ 0, if\ f(x,y) < T \end{cases} \quad \text{(Eq. 3.9)}$$

where $I(x,y)$ is the output at pixel location$(x,y)$, $f(x,y)$ is the intensity of the pixel at location $(x,y)$ and $T$ is the global threshold value. In the current study, the global threshold value is calculated by multiplying the mean of the muscle removed image, the matlab function is: tresh = mean (mean (muscle_removed_image, 1)).

## 3.5 Feature Extraction

After the pectoral muscle is detached from the breast muscle, the resulting image is filtered, enhanced and the clean breast ROI is obtained, features are extracted from the ROI to differentiate normal and abnormal cases. Various methods for feature extraction have been proposed as described in the literature review. In the current study, feature extraction was carried out using gray level co-occurrence matrix (GLCM). GLCM examines how different combinations of pixel intensities occur within an image. The GLCM process includes several steps. The probability of co-occurrence between two grey levels $i$ and $j$ at a specified relative orientation $(y)$ and distance (d) is calculated for all possible pairs of co-occurring grey levels within an image window [75][76].

Mammograms possess a high degree of texture and complexity, which can complicate their interpretation. Consequently, to improve diagnostic precision and reliability, robust feature extraction from mammographic images is critical. The process of feature extraction and selection from an image is crucial for the effectiveness of any classifier. Improving the classifier's accuracy can be attained through by selecting an optimal set of features [76].

In the current study, 4 GLCM matrices were calculated at four different orientations/angles $\theta = \{90°, 135°, 180°, 225°\}$and A total of 24 features were derived from the GLCM, including, autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares: variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation 1, information measure of correlation 2, inverse difference (INV), inverse difference normalized (INN), inverse difference moment normalized, skewness, kortosis and smoothness. These features were extracted from both normal and abnormal subjects, resulting in a feature vector of length 1x96 for each GLCM matrix.

## 3.6 Classification

Following feature extraction, the GLCM features obtained in the previous step (computed from the segmented images) are now classified using support vector machine (SVM). SVM employs a linear discriminant function to classify the features into two or more distinct classes using a linear separator in the feature space [77]. The standard SVM takes a set of given input and predicts one of two possible classes, functioning as non-probabilistic binary linear classifier [78]. We should note here that the SVM is also applicable in multi-class classification, though the interest in the current study is binary classification. When given a labeled set of training samples from two distinct categories, the SVM training algorithm constructs a model that classifies new samples into one of these categories.

The process of classifying images involves two phases: training phase and testing phase [79]. In the training phase, from the training images, the GLCM features are extracted and its model file is created from the features for specific kernel types (Linear, Polynomial and RBF). On the second phase, testing phase, the trained SVM is tested with the features of test set. Features from the test set of the Mammographic Image Analysis Society (MIAS) database are extracted, and these test features are used directly to assess the accuracy of the classification.

A mathematical description of a binary SVM classification scheme is shown in Eq. 3.10. The 24 texture features obtained from the segmented mammogram images of both normal and abnormal breast were used to determine a maximum margin hyper plane between the two classes [77], [80].

$$f_{SVM}(X) = w^T \varphi(X) + b \qquad \text{(Eq. 3.10)}$$

where $w_t$ is the weight vector and $\varphi$ is the mapping function used to map any input function to another dimension space.

The hyperplane is constructed so that it measures the distance to the closest data points on either side, known as support vectors, called support vectors, is determined as also illustrated in Figure 3.3. While Figure 3.4 presents the overall flow chart of the classification process. SVM utilize the data from training images to effectively classify the test images. In a binary classification scenario where two classes labeled as +1 and -1, we possess a training dataset that includes input feature

vectors X alongside their respective class labels $Y$. The formula for the linear hyperplane can be expressed as the set of point's $x$ satisfying the equation given in Eq. 3.11 [77], [80].

$$w^T x + b = 0 \qquad \text{(Eq. 3.11)}$$

The vector $w$ indicates the normal vector of the hyperplane, it points in the direction that is perpendicular to the hyperplane and $T$ stands for transpose. The parameter $b$ in the equation signifies the offset or distance from the origin to the hyperplane along the direction of the normal vector $w$.

The distance between a data point $x_i$ and the decision boundary can be determined as :

$$d_i = \frac{w^T x_i + b}{\|w\|} \qquad \text{(Eq. 3.12)}$$

where $\|w\|$ denotes the Euclidean norm of the weight vector w. For a linear SVM classifier, the Euclidean norm of the normal vector w is defined as follows:

$$\hat{y} = \begin{cases} +1 : w^T x + b \geq 0 \\ -1 : w^T x + b < 0 \end{cases} \qquad \text{(Eq. 3.13)}$$

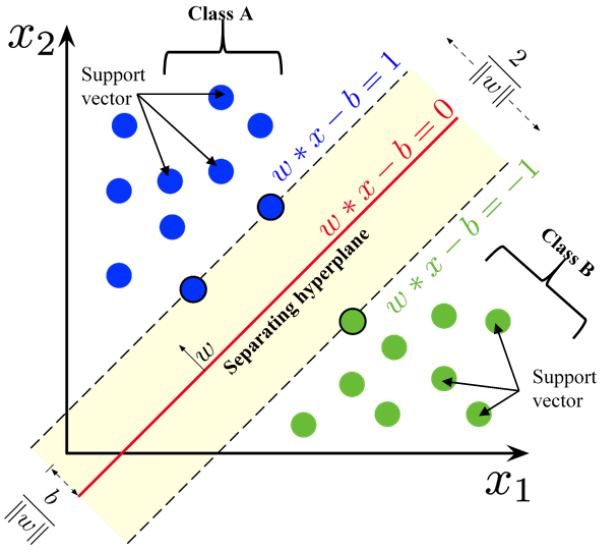



*Figure 0.3: SVM maximum-margin and margins hyperplane trained with two class samples.*

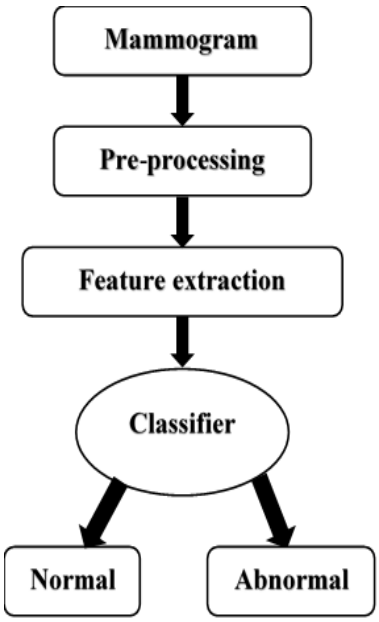



*Figure 0.4: Mammogram image classification flowchart.*

## 3.7 Performance Evaluation

The efficacy of the proposed segmentation approach has been entirely assessed qualitatively by comparing it to the available ground truth data (manual delineation) from the mini-MIAS database. While the performance of the proposed classification system was measured using popular quantitative metrics: sensitivity, specificity and overall accuracy as given in Eq. (14)-(16) [76], [81].

$$Sensitivity = TP / (TP + FN) \quad \text{(Eq. 3.14)}$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad \text{(Eq. 3.15)}$$

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad \text{(Eq. 3.16)}$$

In this context, TP (True Positive) refers to the positive cases that have been correctly classified, TN (True Negative) indicates the negative cases that have been accurately identified, FP (False Positive) indicates negative cases wrongly labeled as positive, and FN (False Negative) indicates the positive cases that have been incorrectly categorized as negative.

# Chapter 4

# Results and Discussion

## 4.1 Introduction

This chapter presents and discusses the results obtained in the current study starting from preprocessing, pictorial muscle removal, abnormal region segmentation up to classification of normal and abnormal mammograms. The discussion includes description of image processing results and the SVM classification results.

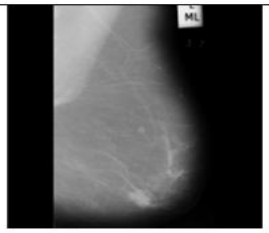
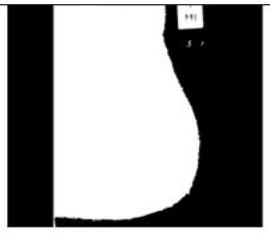
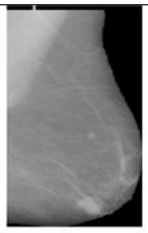
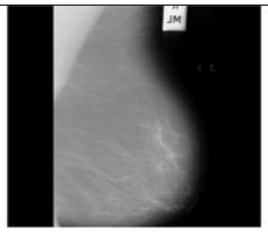
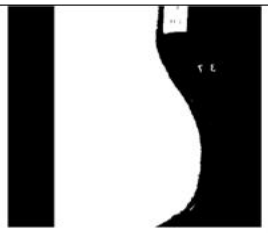
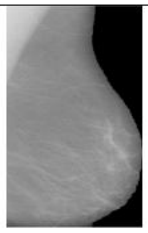
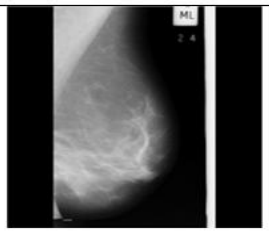
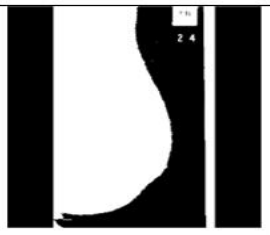
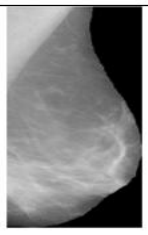
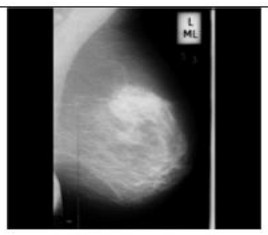
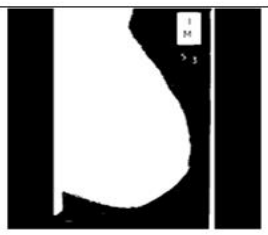
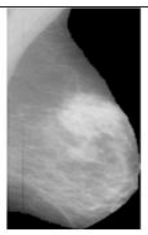
| Image Id | Original image With artifact | Binary image | Binary Image of the Breast Region | Breast Region after re-moving artifact |
|---|---|---|---|---|
| Mdb005 |  |  |  |  |
| Mdb007 |  |  |  |  |
| Mdb053 |  |  |  |  |
| Mdb111 |  |  |  |  |

*Figure 0.1: Preprocessing results: raw image (left) and back ground removed image (right).*

## 4.2 Preprocessing Results

### 4.2.1 Background Removal

Figure 4.1 presents sample results obtained after removing background from four mammogram images (Mdb005, Mdb007, Mdb053, Mdb111). The background removal was carried out in such a way that only the breast region is carried on to the next level of the image processing. The step also removes annotations and other unwanted texts on the mammograms before further image processing is applied.

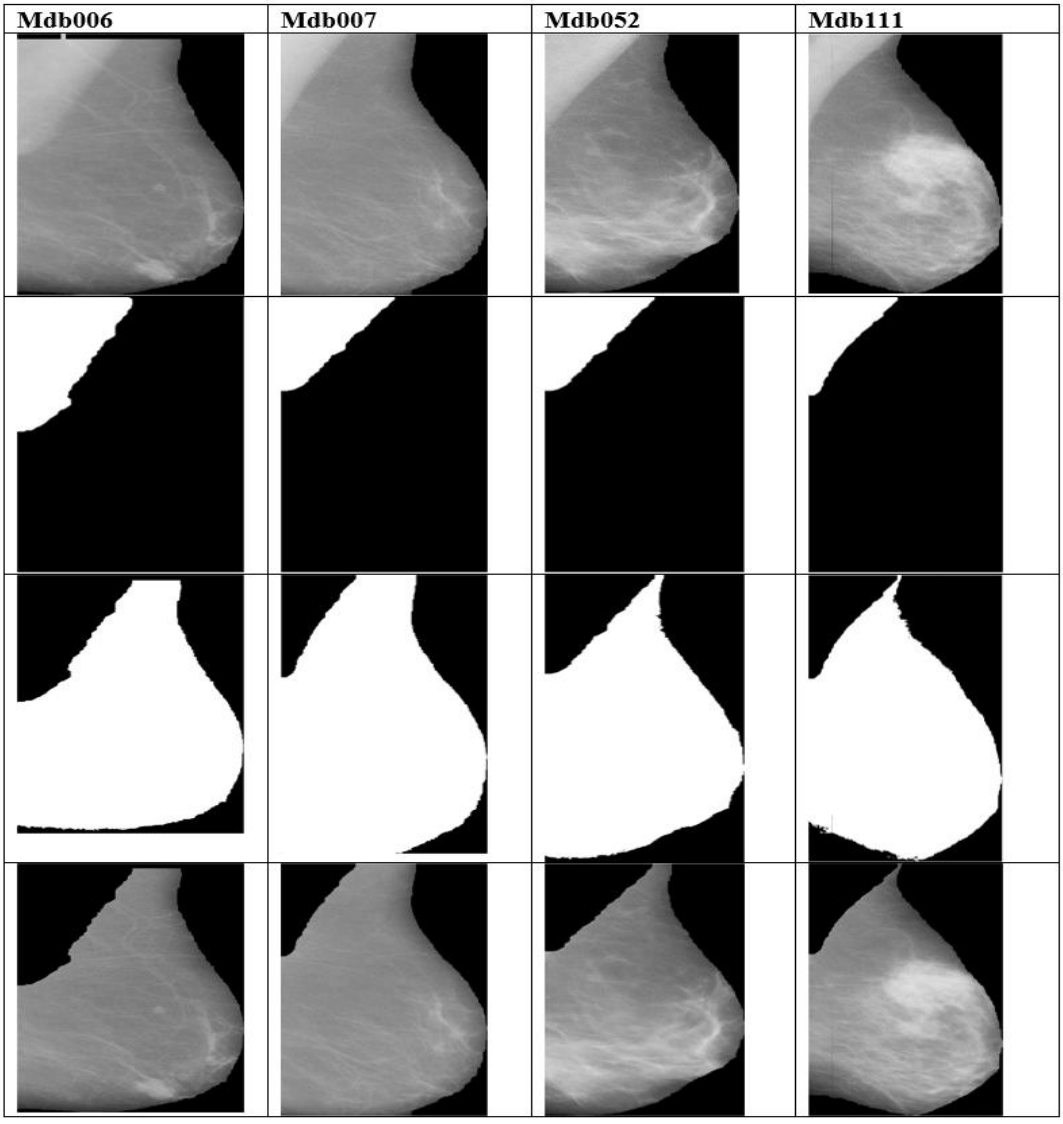



*Figure 0.2: Pictorial muscle removal process: background removed image (top row) and pectoral muscle removed image (bottom row).*

### 4.2.2 Pectoral Muscle Removal

As explained in Chapter 3, pectoral muscle is extracted using seeded region growing method. Distinct intensity values in muscle tissue, fatty tissues, glands, lobules, and ducts allow for their separation into different areas. On a typical mammogram, muscle tissue at the top left has higher intensity region than breast region boundary. Hence, after seeding the top left, a region growing algorithm is applied and the algorithm stops when it reaches breast region which has relatively less intensity. Figure 4.2 demonstrates a typical result that is generated after the pectoral muscle is removed.

### 4.2.3 Image Enhancement

Enhancing the mammogram images involves eliminating noise and adjusting image contrast to facilitate the identification of important features. In this study image denoising is done using combination of median filter and winner filter. Mammograms frequently show low-intensity noise at the skin-air interface of the breast area. Additionally, scratch artifacts may also appear in mammograms. These types of noise and artifacts contain high-frequency components. Thus, it is advisable to apply low-pass filtering to eliminate noise and artifacts . Consequently, a median filter with 3 x 3 window is used for noise suppression. Furthermore, a Wiener spatial domain filter is employed for the removal of speckle noise.

Further image contrast enhancement is achieved using combination of mean contrast stretching and Contrast Limited Adaptive Histogram Equalization (CLAHE). Contrast stretching is used to increase the dynamic range of the gray levels in the image. The range is stretched from the lowest mean gray level to the highest gray level value of the image. Therefore, by distributing the number of gray levels in the captured image over a smaller range, the images can be improved by broadening the number of gray levels to a more extensive range. Thus, in the current study contrast stretching is used first and then CLAHE is used next to enhance the mammogram images. CLAHE improves the contrast of a low-contrast mammogram image by dividing the image into sections and applying histogram equalization to improve the contrast. As can be seen from Figure 4.3, the image contrast enhancement method applied on the pectoral muscle removed images enhances the visibility of abnormal structures (the tumors) in the images. This is considered an important step towards effective segmentation of the abnormalities in the images.

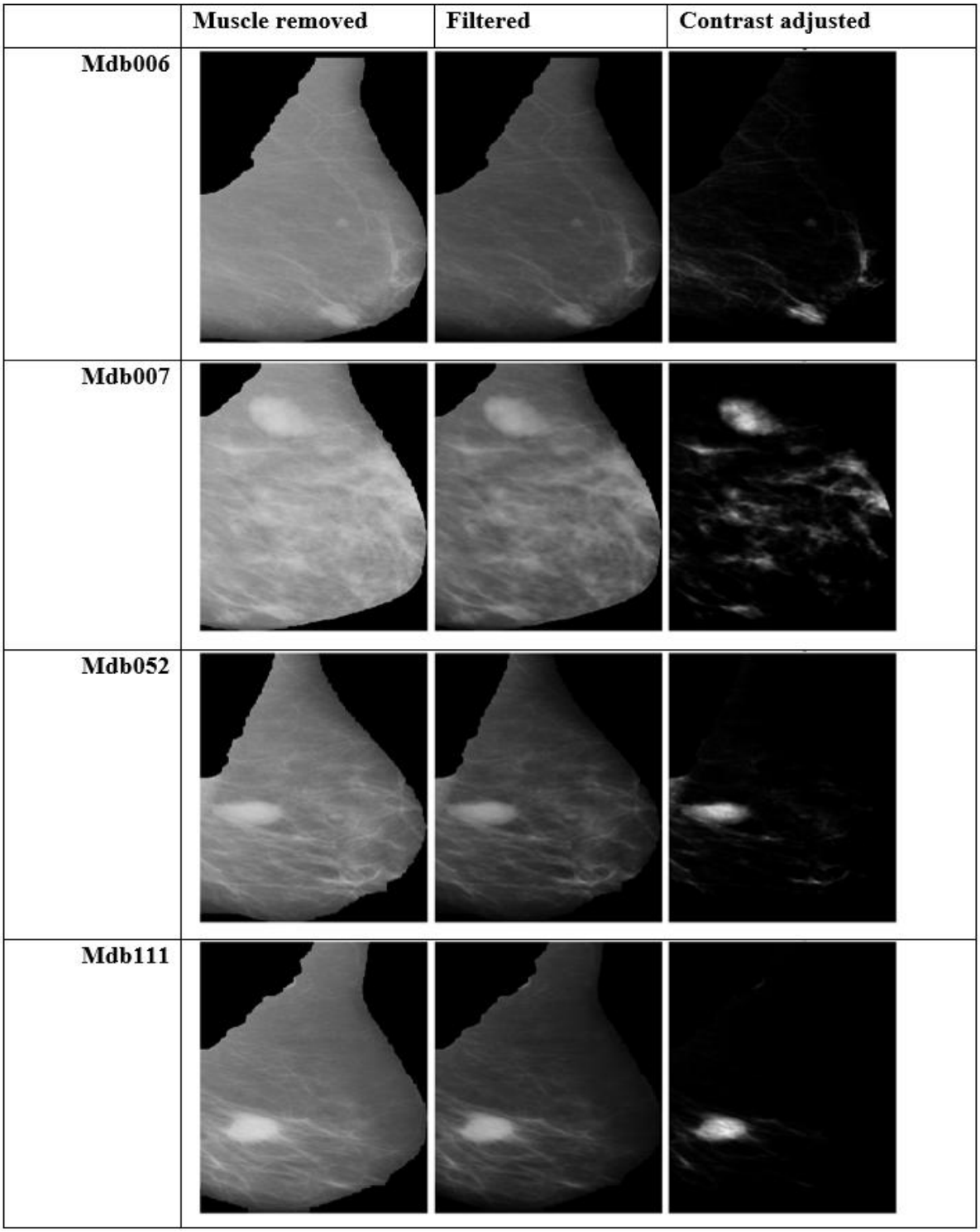



*Figure 0.3: Pectoral muscle removed image (left column) and contrast adjusted image (right column).*

## 4.3 Segmentation Results

Three methods explained in the methodology section on Chapter 3 are compared for their efficacy in accurately segmenting the breast abnormalities (tumors) in the mammogram images: Otsu, global thresholding and watershed. All of the methods are applied on the preprocessed images. The permanence of the segmentation is qualitatively assessed in comparison to the available ground truth, which consists of the contours provided by the physician utilized as the gold standard in this study. Figure 4.4 illustrates some results for demonstration. The qualitative evaluation indicated that the global thresholding and watershed techniques yielded significantly better segmentation results than Otsu.

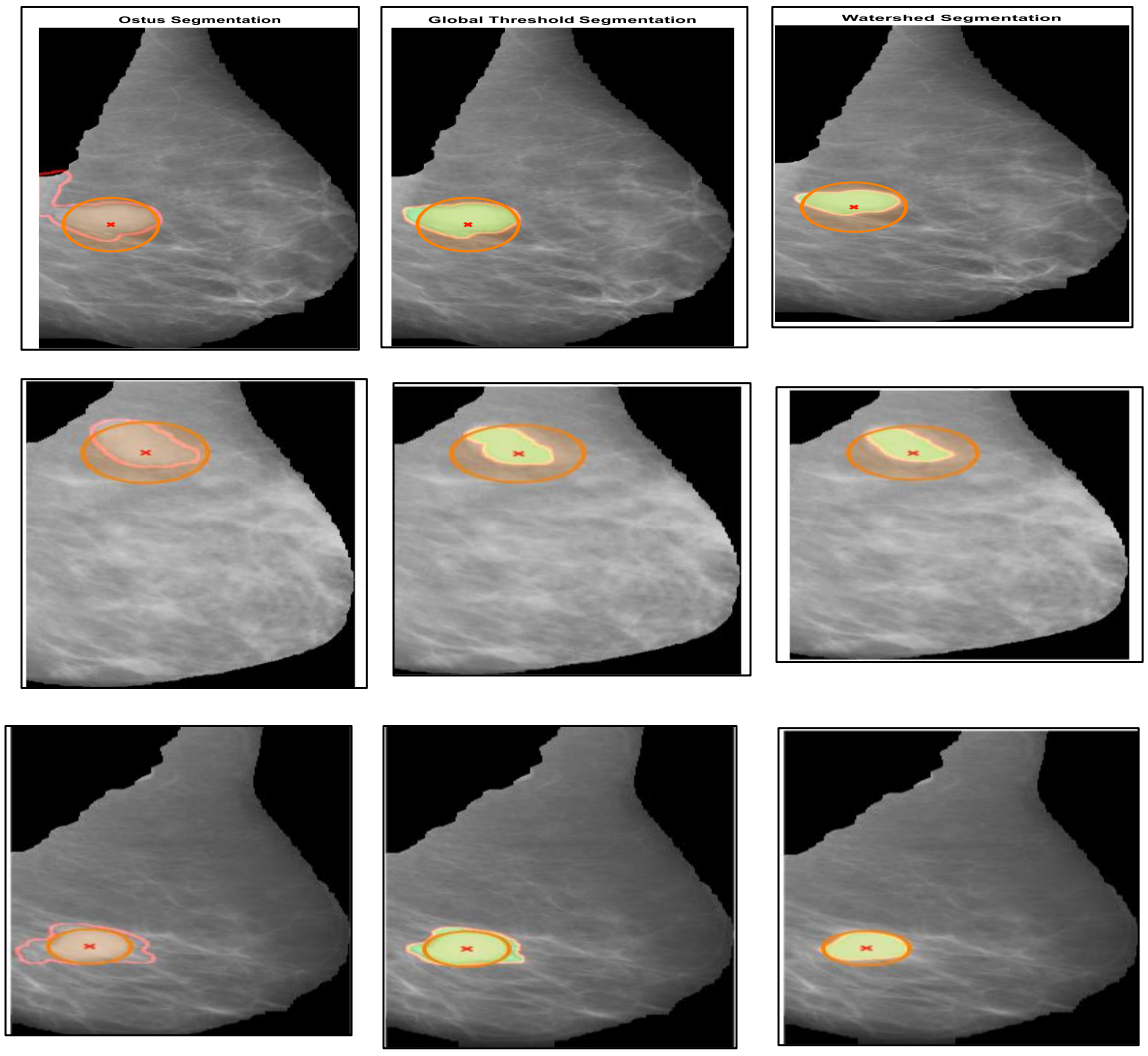



*Figure 0.4: Segmentation sample results: Otsu (left), global thresholding (middle) and watershed (right). Red contour indicates the ground truth information.*

## 4.4 Classification Results

Features were extracted from the finally segmented digital mammograms to automatically classify them into normal and abnormal (due to breast tumor). As indicated in the previous section, both global thresholding and watershed segmentation techniques offered good segmentation outputs. However, due to its simplicity and efficiency, the global thresholding method was chosen for feature extraction. A total of 274 mammograms were used for the classification work and the images come in two classes: 102 abnormal (malignant or benign) images and 172 normal images. 80% of the images from each class was used to train the SVM classification scheme and the rest 20% for testing.

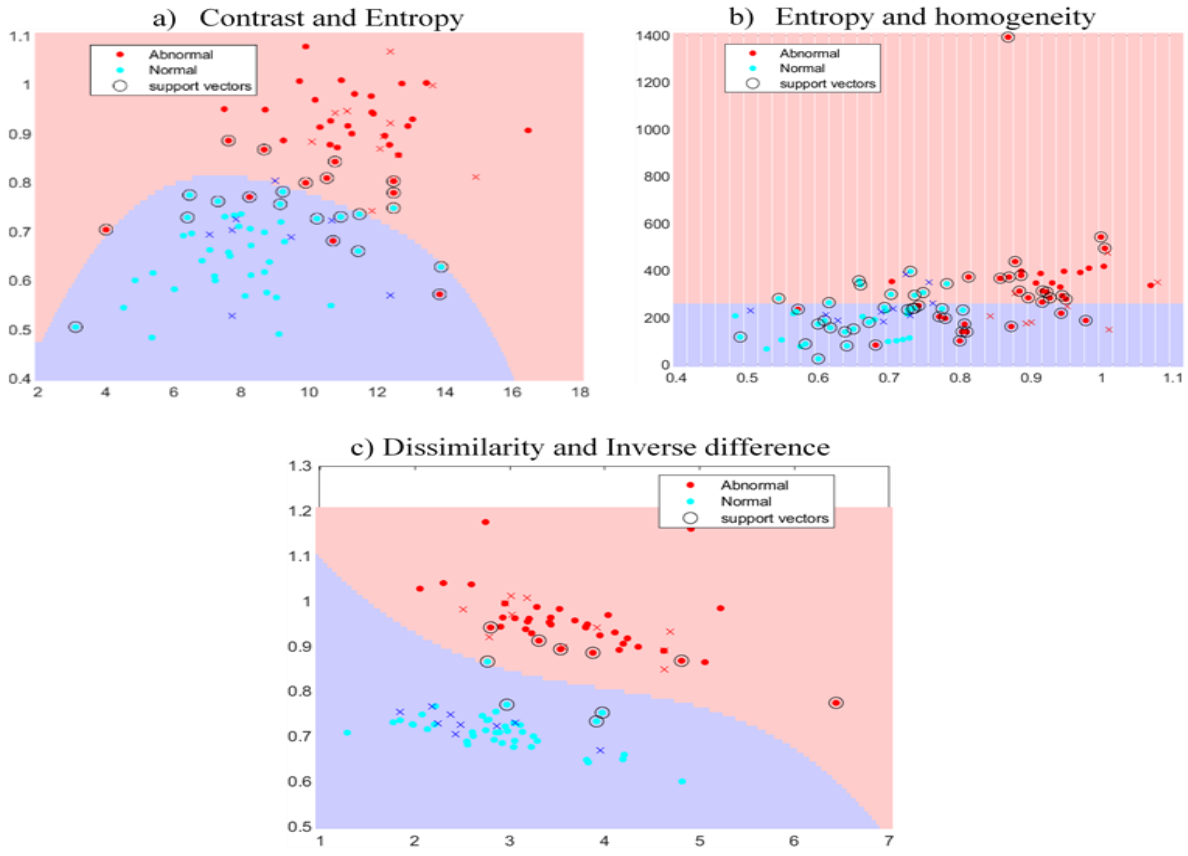



*Figure 0.5: SVM decision boundary for different GLCM feature combinations: (a) Contrast and entropy, (b) Entropy and homogeneity and (c) Dissimilarity and Inverse difference*

## 4.5 Feature Selection

Following preprocessing, segmentation and then obtaining the ROIs from the mammograms, features were extracted using GLCM statistical matrices, as already explained in Chapter 3. The

twentyfour GLCM features (computed in four directions) extracted from the segmented mammograms were compared for their efficacy in accurately classifying the samples into normal and abnormal. Then, after looking at the performance of the individual GLCM features in accurately classifying the mammograms, the best feature combination was sought to improve the classification performance. That was executed by generating the decision boundary generated by the SVM. Figure 4.5 presents the decision boundary created after the SVM classifier is applied. As can be seen on the figure, the combination of dissimilarity feature computed at 225° and the inverse difference feature computed at 90° offered the best overall classification accuracy.

## 4.6 System's Confusion Matrix

A confusion matrix is utilized to assess how well a classifier performs on a given feature dataset. The elements along the diagonal show the count of instances where the predicted and the actual label matches, whereas the off-diagonal elements highlight the instances that have been incorrectly labeled by the classifier. A higher value in the diagonal elements signifies better performance, as it reflects a greater number of correct predictions.

The confusion matrix generated for the proposed SVM based classification scheme is presented in Table 4.1. Out of the 102 abnormal and 172 normal mammogram samples, 95 abnormal and 155 normal cases have been correctly identified.

Table 0.1: SVM classifier system confusion matrix.

| **Ground truth** | **SVM Classification** | | **Total** |
|---|---|---|---|
| | **Abnormal** | **Normal** | |
| **Abnormal** | TP = 95 | FN = 7 | 102 |
| **Normal** | FP = 17 | TN = 155 | 172 |
| **Total** | 112 | 162 | **274** |

As can be seen from the confusion matrix shown in Figure Fig 4.6, out of the 102 abnormal images, 93.14% are classified as abnormal and 6.86% are misclassified as normal. Also, out of 172 normal images 90.12% are classified as normal images and 9.88% are misclassified as abnormal.

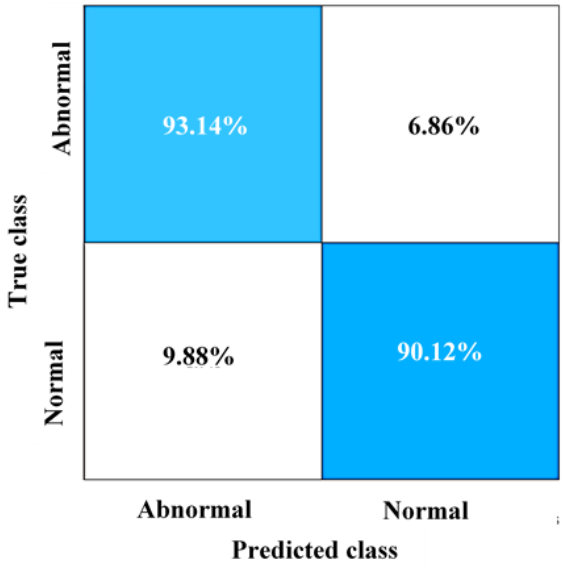



*Figure 0.6: SVM classifier performance: true class versus predicted class.*

**Sensitivity (SN) or Recall or true positive rate (TPR):**

$$SN = \frac{TP}{TP + FN} * 100\% = \frac{95}{95 + 7} * 100\% = 93.14\%$$

**Specificity (SP) or TNR:**

$$\text{SP} = \frac{TN}{TN + FP} * 100\% = \frac{155}{155 + 17} * 100\% = 90.12\%$$

**Positive predictive value (PPV) or precision:**

$$\text{SPPV} = \frac{TP}{TP + FP} * 100\% = \frac{95}{95 + 17} * 100\% = 84.82\%$$

**Negative predictive value (NPV):**

$$\text{NPV} = \frac{TN}{TN + FN} * 100\% = \frac{155}{155 + 7} * 100\% = 95.68\%$$

**Accuracy (ACC):**

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{95 + 155}{95 + 155 + 17 + 7} * 100\% = 91.24\%$$

**Error rate (ERR):**

$$\text{ERR} = \frac{FP + FN}{TP + TN + FP + FN} * 100\% = \frac{17 + 7}{95 + 155 + 17 + 7} * 100\% = 8.76\%$$

From the performance metrics, we can see that the probability that the system would detect breast abnormality is 93.14%, while the chance of it correctly classifying a patient's mammogram image as normal among those without breast abnormalities is 90.12%. Additionally, when the system classifies a mammogram as abnormal, there is an 84.82% probability that it is indeed abnormal, and the probability that a mammogram is truly normal when the system labels it as such is 95.68%.

## 4.7 ROC Curve

To more precisely measure the classification effectiveness of the suggested classification scheme, receiver operating characteristic curve (ROC) curves were created along with their corresponding area under the curve (AUC) values. Figure 4.7 shows the ROC curve generated with the corresponding AUC value of 0.92 showing the great promises the proposed method carries in accurately classifying breast mammograms into normal and abnormal.

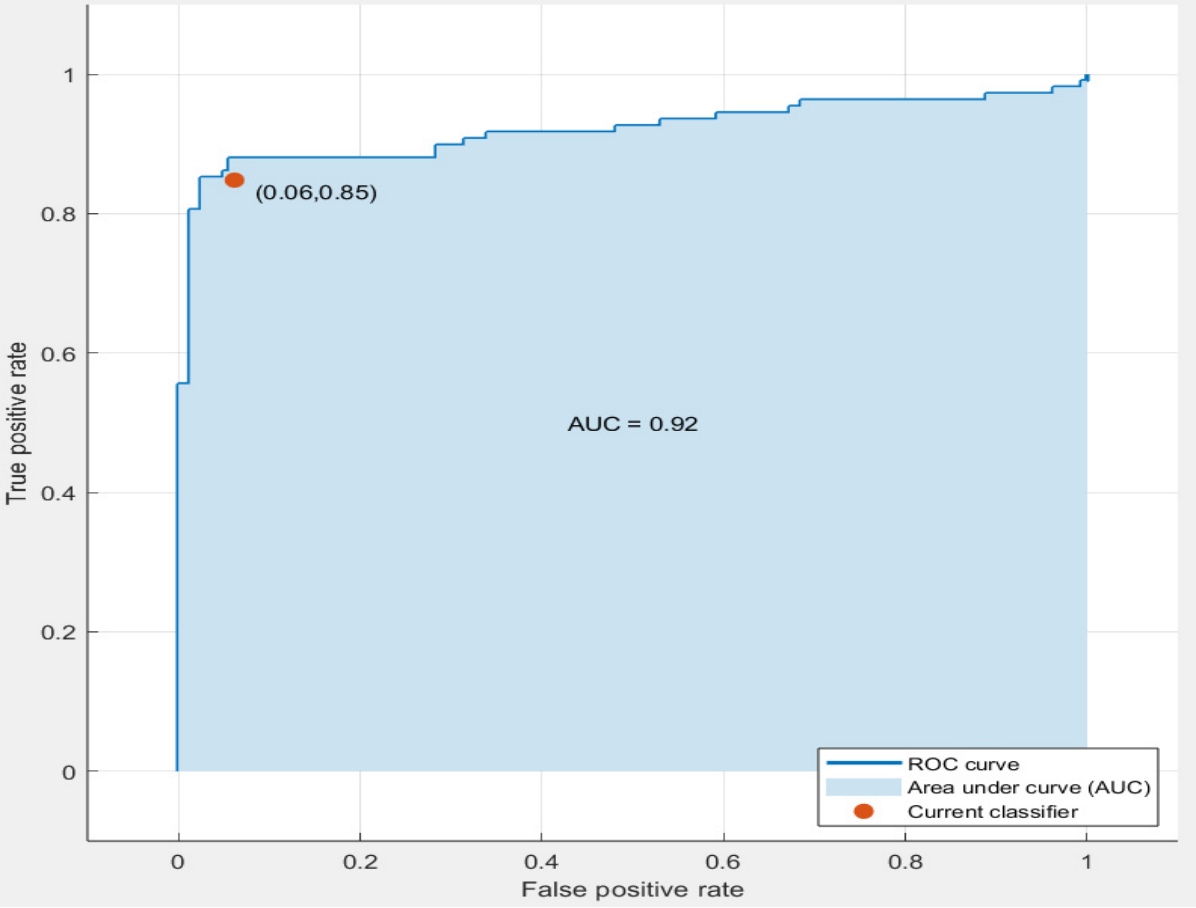



*Figure 0.7: ROC plot and the corresponding AUC value for the proposed SVM classifier.*

# Chapter 5

# Conclusion and Recommendation

## 5.1 Conclusion

The significant variability in interpretation when various radiologists analyze the same mammogram independently can result in diagnostic errors. Therefore, developing a more effective technique to support ROI identification and classification on mammograms is essential. This study has presented an automated scheme for segmentation of breast abnormalities as well as an SVM based classification scheme to differentiate between normal and abnormal breast mammograms exploiting GLCM features.

Before abnormalities are segmented and features are extracted, useful image pre-processing steps were utilized. The pre-processing steps include image filtering to remove noise, image enhancement for better feature extraction and pictorial muscle removal for better detection of region of interest. Twenty four GLCM features were extracted in four directions $(90°, 135°, 180°, 225°)$. Out of the twenty-four features, best classification was achieved by combining dissimilarity at 225° and inverse difference at 90°.

To measure the effectiveness of the proposed method, various performance evaluation matrices were utilized including sensitivity, specificity; positive predictive value (PPV), negative predictive value (NPV) and accuracy. From the performance metrics, the probability that the system would detect a breast abnormality is 0.93 (sensitivity = 93.14%). The probability that the system will identify a patient's mammogram image as benign among those who do not have cancer is 0.901 (specificity =90.12%). Furthermore, the likelihood that a mammogram is genuinely malignant when the system categorizes it as cancerous is 0.848 (PPV =84.82%) Additionally, the probability that a mammogram is benign when the system indicates it is not malignant is 0.957, meaning the NPV is 95.68. The performance metrics resulted in an AUC value nearing 0.92, indicating a very good classification performance.

Based on the performance evaluation results, the system has the potential to assist radiologists in decision-making during mammogram interpretation. This support could enhance their diagnostic abilities by enabling them to make prompt and more accurate decisions.

## 5.2 Recommendations

The research conducted in this thesis has highlighted several areas that warrant further exploration. For instance, the segmentation of the pectoral muscle and the breast region of interest was not consistently accurate, potentially due to limitations in pre-processing and segmentation techniques. Consequently, there is room to improve the classification accuracy of the proposed method. Moreover, the study relied on a relatively small data set for training and testing, and expanding the sample size could significantly enhance the classification performance.

This study focused on developing an algorithm specifically for the detection and classification of breast tumors. However, other breast abnormalities, such as masses and architectural distortions, which can serve as early indicators of breast cancer, were not addressed. Further research is needed to explore these areas and related topics, including effective methods to distinguish between benign and malignant breast tumors. Similar CAD systems could be proposed for the automatic detection of these abnormalities. Combining these systems could provide significant benefits to radiologists, aiding in the early diagnosis of breast abnormalities and potentially reducing the need for unnecessary biopsy procedures.

# Reference


[1] J. Uscher, “Screening and Testing for Breast Cancer.” .

[2] Society American Cancer, “Breast Cancer Statistics | How Common Is Breast Cancer? | American Cancer Society,” 2024. .

[3] R. Steen, André & Tiggelen, “Short history of mammography: a Belgian perspective,” vol. 90, 2007, [Online].Available:https://www.researchgate.net/publication/6143668 Short history of mammography a Belgian perspective.

[4] “WHO Report,” 2022. https://www.who.int/news-room/fact-sheets/detail/cancer.

[5] “Breast cancer statics,” [Online]. Available: https://www.cancer.net/cancer-types/breast-cancer/statistics#:~:text=In 2023%2C an estimated 297%2C790,half a percent each year.

[6] et al Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, “Cancer Today,” 2020. [Online]. Available: https://gco.iarc.fr/today.

[7] T. R. Vaishnavi Patil1, Shravani Burud2*, Goutami Pawar3 and S. B. Hebbale, “Breast Cancer Detection Using Matlab functions,” *Adv. Image Process. Pattern Recognit.*, vol. 3, no. 2, 2020.

[8] A. C. Society, “Cancer Facts and Figures 2023,” *Am. Cancer Soc.*, 2024, [Online]. Available:https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf.

[9] “Breastcancer.org,”*Breast Cancer Facts and Statistics*, 2024. https://www.breastcancer.org/facts-statistics.

[10] R. Menon, Radhika & Raha, Poulami & Kothari, Shweta & Chakraborty, Sumit & Chakrabarti, Indrajit & Karim, “Automated detection and classification of mass from breast ultrasound images,” *2015 Fifth Natl. Conf. Comput. Vision, Pattern Recognition, Image Process. Graph.*, 2015, doi: http://dx.doi.org/10.1109/NCVPRIPG.2015.7490070.

[11] M. R. Khalilabad, N. D., Hassanpour, H., & Abbaszadegan, “Fully automatic classification

of breast cancer microarray images," *ournal Electr. Syst. Inf. Technol.*, vol. 3, no. 2, pp. 348-359., 2016, doi: 10.1016/j.jesit.2016.06.001.

[12] P. S. K. Bandyopadhyay, "Pre-processing of Mammogram Images," *Int. J. Eng. Sci. Technol.*, 2010, [Online]. Available: https://www.researchgate.net/publication/50384221_Pre-processing_of_Mammogram_Images.

[13] S. Ponraj, D & Jenifer, M & Poongodi, P. & Manoharan, "A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 2, 2011, [Online].Available:https://www.researchgate.net/publication/266489422_A_Survey_on_the_Preprocessing Techniques of Mammogram for the detection of Breast Cancer.

[14] "Mammography Fact sheet," *Natl. Inst. Biomed. IMAGING Bioeng.*, 2019, [Online]. Available: https://www.nibib.nih.gov/sites/default/files/2020-03/Mammography Fact Sheet 2019.pdf.

[15] P. Bonnie Joe, MD, "Advances in Breast Imaging: Evolution & History of Mammography," *San Franscisco Med.*, 2015, [Online]. Available: https://issuu.com/sfmedsociety/docs/march_9c327b3b848572/1?e=3533752/11778000.

[16] Y. Guo, "Computer-Aided Detection of Breast Cancer Using Ultrasound Images," *https://digitalcommons.usu.edu/*, 2010, [Online]. Available: https://core.ac.uk/download/pdf/19682055.pdf.

[17] A. G. Sepideh Iranmakani1, Tohid Mortezazadeh2, Fakhrossadat Sajadian1, Mona Fazel Ghaziani1 and D. K. and A. E. Musa3, "A review of various modalities in breast imaging: technical aspects and clinical outcomes," *Egypt. J. Radiol. Nucl. Med.*, vol. 51, no. 1, 2020, doi: https://doi.org/10.1186/s43055-020-00175-5.

[18] "Breast Anatomy," *Cleveland Clinic*. https://my.clevelandclinic.org/health/articles/8330-breast-anatomy.

[19] A. L. Asma Javed, "Development Of The Human Breast," *Natl. Lib. Med.*, vol. 27, no. 1,

2013, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/issues/224872/.

[20] "Anatomy and Physiology of the Breast," *Johns Hopkins University Department of Pathology*. https://pathology.jhu.edu/breast/overview.

[21] H. Direct, "Breast Diseases," *Health Direct*. https://www.healthdirect.gov.au/breast-diseases.

[22] F. A. V. Samith Sandadi, David T. Rock, James W. Orr, "Breast diseases: Detection, Management, and Surveillance of Breast Disease," *Compr. Gynecol. (Eighth Ed. Elsevier*, pp. 289-322.e3, 2022, doi: https://doi.org/10.1016/B978-0-323-65399-2.00024-3.

[23] National Cancer Institute, "Does having dense breast tissue affect a mammogram?," *National Cancer Institute*, 2024. https://www.cancer.gov/types/breast/breast-changes/dense-breasts.

[24] Https://www.cancer.org/cancer/acs-medical-content and news-staff.html, "Limitations of Mammograms," *The American Cancer Society*. .

[25] R. F. dos S. Teixeira, "Automatic Analysis Of Mammography Images: Classification Of Breast Density," pp. 26–32, 2013, [Online]. Available: f:%5C302970018.pdf.

[26] "Breast Cancer," *Cleveland Clinic*, 2023. https://my.clevelandclinic.org/health/diseases/3986-breast-cancer.

[27] Allen + Clarke, "Breast Density," Australia, 2018. [Online]. Available: https://www.health.gov.au/sites/default/files/documents/2019/09/breast-density-literature-review_1.pdf.

[28] Https://www.cancer.org/cancer/acs-medical-content-and-news-staff.html, "Survival Rates for Breast Cancer," *American Cancer society*, 2024. https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html.

[29] G. C. Observatory, "No Title." .

[30] "Understanding Breast Changes and Conditions," *Natl. Cancer Inst.*, 2022, [Online].

Available: https://www.cancer.gov/types/breast/breast-changes/understanding-breast-changes.pdf.

[31] P. C. H. Berment, V. Becette, M. Mohallem, F. Ferreira, “Masses in mammography: What are the underlying anatomopathological lesions?,” *Diagn. Interv. Imaging*, vol. 95, no. 2, pp. 126–135, 2014, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211568413003872.

[32] V. H. A. Carlos B. Fiallos1, Maria G. Pérez2, Aura Conci3, “Automatic Detection of Injuries in Mammograms Using Image Analysis Techniques,” *IEEE*, 2015, doi: https://doi.org/10.1109/IWSSIP.2015.7314222.

[33] healthcare and editorial professionals The ACS, “What Does the Doctor Look for on a Mammogram?,” *American Cancer society*, 2022. https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/mammograms/what-does-the-doctor-look-for-on-a-mammogram.html.

[34] P. Mitra, “Computer Based Analysis Of Mammogram Images To Treat Breast Cancer,” 2013.

[35] “Breast Cancer Detection From Mammograms Using Image Processing Techniques Department Of Electronics And Communication.”

[36] “Types of Breast Cancer: What to Know About Common and Rare Forms,” *Breast Cancer Research Foundation*, 2023. https://www.bcrf.org/blog/types-of-breast-cancer/.

[37] A. S. G. K. Amrutha, P. Meghana, D. S. S. Dharma Teja, “Breast Cancer Detection From Mammograms Using Image Processing Techniques.”

[38] “Mammogram Procedure,” *Johns Hopkins Medicine*. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/mammogram-procedure.

[39] M. R. Bulevar Kralja Aleksandra, “Basic Principles Of Mammography.”

[40] “Breast Imaging: Mammography,” *RADIOLOGY KEY*. https://radiologykey.com/breast-

imaging-mammography/.

[41] P. J. D. MD, “The History of Breast Ultrasound,” *Journal Ultrasound Med.*, vol. 23, no. 7, pp. 887–894, [Online]. Available: https://onlinelibrary.wiley.com/doi/10.7863/jum.2004.23.7.887.

[42] D. Sabih, “Breast Ultrasonography,” *MEDSCAPE*, 2021. https://emedicine.medscape.com/article/1948269-overview.

[43] “Breast Ultrasound,” *RadiologyInfo.Org*, 2022. https://www.radiologyinfo.org/en/info/breastus.

[44] G. Katti, S. A. Ara, and A. Shireen, “Magnetic resonance imaging (MRI)--A review,” *Int. J. Dent. Clin.*, vol. 3, no. 1, pp. 65–70, 2011.

[45] M. Morrow, J. Waters, and E. Morris, “MRI for breast cancer screening, diagnosis, and treatment,” *Lancet*, vol. 378, no. 9805, pp. 1804–1811, 2011.

[46] E. A. Morris, “Breast cancer imaging with MRI,” *Radiol. Clin.*, vol. 40, no. 3, pp. 443–466, 2002.

[47] M. Salvatore and S. Del Vecchio, “Dynamic imaging: scintimammography,” *Eur. J. Radiol.*, vol. 27, pp. S259--S264, 1998.

[48] O. Schillaci, R. Danieli, P. Romano, R. Santoni, and G. Simonetti, “Scintimammography for the detection of breast cancer,” *Expert Rev. Med. Devices*, vol. 2, no. 2, pp. 191–196, 2005.

[49] B. M. Das BK, Biswal BM, Bhavaraju MDas BK, Biswal BM, “Role of Scintimammography in the Diagnosis of Breast Cancer,” *Malaysian J. Med. Sci.*, vol. 13, no. 1, pp. 52–57, 2006, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3347903/pdf/mjms-13-1-052.pdf.

[50] J. Fitzjohn, C. Zhou, and J. Geoffrey Chase, “Critical Assessment of Mammography Accuracy,” *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 5620–5625, 2023, doi: 10.1016/j.ifacol.2023.10.472.

[51] O. Schillaci and J. R. Buscombe, “Breast imaging with scintimammography,” *Breast Cancer Nucl. Med. Diagnosis Ther. Options*, pp. 57–70, 2008.

[52] B. Bagni *et al.*, “Scintimammography with 99mTc-MIBI and magnetic resonance imaging in the evaluation of breast cancer,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 30, pp. 1383–1388, 2003.

[53] J. Nahas-Neto *et al.*, “Evaluation of mammographic density and 99mTc-sestamibi scintimammographic uptake in postmenopausal women on hormone replacement therapy,” *Maturitas*, vol. 53, no. 1, pp. 97–106, 2006.

[54] P. A. Carney *et al.*, “Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography,” *Ann. Intern. Med.*, vol. 138, no. 3, pp. 168–175, 2003.

[55] T. Kurita *et al.*, “Roles of fine-needle aspiration and core needle biopsy in the diagnosis of breast cancer,” *Breast Cancer*, vol. 19, pp. 23–29, 2012.

[56] Y.-H. Yu, W. Wei, and J.-L. Liu, “Diagnostic value of fine-needle aspiration biopsy for breast mass: a systematic review and meta-analysis,” *BMC Cancer*, vol. 12, pp. 1–14, 2012.

[57] S. M. Willems, C. H. M. Van Deurzen, and P. J. Van Diest, “Diagnosis of breast lesions: fine-needle aspiration cytology or core needle biopsy? A review,” *J. Clin. Pathol.*, vol. 65, no. 4, pp. 287–292, 2012.

[58] M. J. G. Kshema and D. A. S. Dhas, “Preprocessing filters for mammogram images: a review,” in *IEEE Conference on Emerging Devices and Smart Systems*, 2017, pp. 3–4.

[59] R. Srisha and A. M. Khan, “Morphological operations for image processing: understanding and its applications,” *NCVSComs-13*, vol. 13, no. 17–19, p. 19, 2013.

[60] M. L. Comer and E. J. Delp III, “Morphological operations for color image processing,” *J. Electron. Imaging*, vol. 8, no. 3, pp. 279–289, 1999.

[61] K. Parvati, B. S. Prakasa Rao, and M. Mariya Das, “Image Segmentation Using Gray-Scale Morphology and Marker-ControlledWatershed Transformation,” *Discret. Dyn. Nat. Soc.*,

vol. 2008, no. 1, p. 384346, 2008.

[62] D. Raba, A. Oliver, J. Mart\'\i, M. Peracaula, and J. Espunya, "Breast segmentation with pectoral muscle suppression on digital mammograms," in *Pattern Recognition and Image Analysis: Second Iberian Conference, IbPRIA 2005, Estoril, Portugal, June 7-9, 2005, Proceedings, Part II 2*, 2005, pp. 471–478.

[63] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, R. A. Borges, and A. F. Frere, "Automatic identification of the pectoral muscle in mammograms," *IEEE Trans. Med. Imaging*, vol. 23, no. 2, pp. 232–245, 2004.

[64] S. Bandyopadhyay, I. K. Maitra, S. Nag, and S. K. Bandyopadhyay, "Automated Digital Mammogram Segmentation for Detection of Abnormal Masses using Binary Homogeneity Enhancement Algorithm." 2011.

[65] S. Wu, S. Yu, Y. Yang, and Y. Xie, "Feature and contrast enhancement of mammographic image based on multiscale analysis and morphology," *Comput. Math. Methods Med.*, vol. 2013, 2013, doi: 10.1155/2013/716948.

[66] J. K. Bates, "Multidimensional Systems: Signal Processing and Modeling Techniques," in *Control and Dynamic Systems*, Second., vol. 69, ScienceDirect, 1995, pp. 299–435.

[67] O. Magud, E. V. A. Tuba, and N. Bacanin, "Medical ultrasound image speckle noise reduction by adaptive median filter," *Wseas Trans. Biol. Biomed*, vol. 14, pp. 38–46, 2017.

[68] S. Hemali, A. Smita, P. Oza, S. Tanwar, and A. Alkhayyat, "Mammogram Pre-processing Using filtering methods for Breast Cancer Diagnosis," *Int. J. Image, Graph. Signal Process.*, vol. 15, no. 4, pp. 44–58, 2023.

[69] P. S. Hiremath, P. T., and S. Badiger, "Speckle Noise Reduction in Medical Ultrasound Images," *Adv. Break. Ultrasound Imaging*, 2013, doi: 10.5772/56519.

[70] I. U. I. Wani, M. C. Hanumantharaju, M. T. Gopalakrishna, and others, "Review of mammogram enhancement techniques for detecting breast cancer," *IJCA Proc. ICICT*, 2014.

[71] M. I. K. Abir, “Contrast enhancement of digital mammography based on multi-scale analysis,” 2011.

[72] R. Suresh, A. N. Rao, and B. E. Reddy, “Detection and classification of normal and abnormal patterns in mammograms using deep neural network,” in *Concurrency and Computation: Practice and Experience*, Jul. 2019, vol. 31, no. 14, doi: 10.1002/cpe.5293.

[73] R. E. W. Gonzalez, Rafael C. (University of Tennessee), *Digital Image Processing*, Fourth Edi. 330 Hudson Street, New York, NY 10013: Pearson Education, 2018.

[74] N. Pradeep, H. Girisha, B. Sreepathi, and K. Karibasappa, “Feature extraction of mammograms,” *Int. J. Bioinforma. Res. ISSN*, vol. 9753087, pp. 136–138, 2012.

[75] D. C. R. Novitasari, A. Lubab, A. Sawiji, and A. H. Asyhar, “Application of feature extraction for breast cancer using one order statistic, GLCM, GLRLM, and GLDM,” *Adv. Sci. Technol. Eng. Syst. J.*, vol. 4, no. 4, pp. 115–120, 2019.

[76] V. P. Singh, A. Srivastava, D. Kulshreshtha, A. Chaudhary, and R. Srivastava, “Mammogram classification using selected GLCM features and random forest classifier,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 6, pp. 82–87, 2016.

[77] A. Qayyum and A. Basit, “Automatic breast segmentation and cancer detection via SVM in mammograms,” in *2016 International conference on emerging technologies (ICET)*, 2016, pp. 1–6.

[78] V. J. Gaikwad, “Detection of Breast Cancer in Mammogram using Support Vector Machine.” .

[79] S. Kavitha and K. K. Thyagharajan, “Features based mammogram image classification using weighted feature support vector machine,” in *International Conference on Computing and Communication Systems*, 2011, pp. 320–329.

[80] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, “Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform,” *Meas. J. Int. Meas. Confed.*, vol. 146, pp. 800–805, Nov. 2019, doi: 10.1016/j.measurement.2019.05.083.

[81] A. A. Kayode, N. O. Akande, A. A. Adegun, and M. O. Adebiyi, “An automated mammogram classification system using modified support vector machine,” *Med. Devices Evid. Res.*, vol. 12, pp. 275–284, 2019, doi: 10.2147/MDER.S206973.