DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA

MINISTRY OF HIGHEST EDUCATION AND SCIENTIFIC RESEARCHES

IBN KHALDOUN UNIVERSITY – TIARET

THESIS

Posted to :

MATHEMATICS AND COMPUTER SCIENCE FACULTY

For The graduation of :

MASTER

Specialty : [Software Engineering]

Presented by:

[BOUDERBALA Amina]

s[KHALFA Chifaa Teriba]

On the subject :

Mammographic Image segmentation by deformable models

Publicly supported on **/07/2023 in Tiaret in front of the jury composed of:

Mr. DAOUD Bachir.	MAA	Ibn-Khaldoun Tiaret University	President
Mr. CHENINE Abdelkader.	MAA	Ibn-Khaldoun Tiaret University	Supervisor
Mr. GAFOUR Yacine.	MCB	Ibn-Khaldoun Tiaret University	Examiner

Acknowledgements

First of all, we would like to thank Almighty God and merciful, who gave us the strength and the patience to accomplish this modest work.

We would like to express our genuine gratitude to our supervisor Mr. CHENINE Abdelkader for his precious advice and guidance.

We express our heartfelt appreciation to all our qualified teachers for their scholarly generosity for five years of unwavering sincerity. We are deeply pleased and honored.

We should also thank the honorable jury for its dedication to correct this work.

Finally, we would like to thank all the people who participated directly or indirectly in the development of this modest work. Thank you to all of you ...!

Dedications

From the bottom of my heart, I dedicate this work to those who are dear to me

To my very dear father,

To the only man who is precious to me, the best offer Allah offered me, to whom I owe my success and all my respect, my dear father my source of affection who has always been there, may God keep you for us!

To my dearest mother,

No dedication can express the love, esteem, devotion and respect that I always had for you, and nothing in the world is worth the efforts put in day and night for my education and my well-being, you have always been present by my side and comforting me when you had to keep supporting me, without your help, your advice and your encouragement this work would not have seen the light of day

To my dear sisters and brothers,

My dears, words are not enough to express the attachment, love and the affection I have for you, my guardian angels and my faithful companions in the most delicate moments of this life, I am lucky to have you have by my side. I dedicate this work to you with all my wishes for happiness, health and of success

To my best friends and my binome,

In testimony to the friendship that unites us and memories of all the times we spent together, I dedicate this work and I wish you a life full of health and happiness.

I appreciate you all...!

KHALFA chifaa Teriba

Dedication

I dedicate this work

To the soul of my father who wished to see me graduated this year but his candle melted too soon, feelings are not enough for describe what I would have felt if he was present today.

To the most important person in my life. My beloved mother, who taught me to trust Allah and believe in myself.

To my sister and brothers who were always here for help, for all their support and care.

To my friend Fadhila thank you for making my life colorful.

To the best partner I ever could get, Chifaa Thank you for your efforts and the hard work you put on this project, I'm grateful to have you as a partner.

To everyone who has supported me and helped me get to where I am now,

Thank you all.

Amina.

Abstract

Breast cancer is a significant global health concern, necessitating early and accurate diagnosis for effective treatment and improved patient outcomes. Medical imaging techniques, particularly mammography, play a crucial role in breast cancer detection. However, segmenting cancerous regions from mammographic images is challenging due to the complexity and variability of breast tissue, as well as the presence of noise and overlapping structures. This study focuses on two deformable methods, the Chan-Vese and Snakes (Kass) methods, for breast cancer image segmentation. By conducting a comparative study using the publicly available MIAS database, we assess the effectiveness of these methods in accurately segmenting breast cancer regions and compare their performance in terms of accuracy, robustness, and computational efficiency. The results of this research aim to contribute insights into the strengths and limitations of deformable methods for segmentation, with the potential to enhance the development of more robust and accurate algorithms for breast cancer diagnosis and treatment planning.

Résumé

Le cancer du sein est une préoccupation majeure pour la santé mondiale, nécessitant un diagnostic précoce et précis pour un traitement efficace et des résultats améliorés pour les patients. Les techniques d'imagerie médicale, en particulier la mammographie, jouent un rôle crucial dans la détection du cancer du sein. Cependant, la segmentation des régions cancéreuses à partir d'images mammographiques est difficile en raison de la complexité et de la variabilité des tissus mammaires, ainsi que de la présence de bruit et de structures superposées. Cette étude se concentre sur deux méthodes déformables, la méthode de Chan-Vese et la méthode des Serpents (Kass), pour la segmentation d'images de cancer du sein. En réalisant une étude comparative à l'aide de la base de données publique MIAS, nous évaluons l'efficacité de ces méthodes pour segmenter avec précision les régions de cancer du sein et comparons leurs performances en termes de précision, de robustesse et d'efficacité informatique. Les résultats de cette recherche visent à apporter des informations sur les forces et les limites des méthodes déformables pour la segmentation, avec le potentiel d'améliorer le développement d'algorithmes plus robustes et précis pour le diagnostic et la planification du traitement du cancer du sein.

ملخص

يعتبر سرطان الثدي من الشواغل الصحية العالمية الهامة ، مما يستلزم التشخيص المبكر والدقيق للعلاج الفعال وتحسين نتائج المرضى. تلعب تقنيات التصوير الطبي ، وخاصة التصوير الشعاعي للثدي ، دورًا مهمًا في اكتشاف سرطان الثدي. ومع ذلك ، فإن تجزئة المناطق السرطانية من الصور الشعاعية للثدي يمثل تحديًا بسبب تعقيد وتنوع أنسجة الثدي ، فضلاً عن وجود ضوضاء وهياكل متداخلة. تركز هذه الدراسة على طريقتين قابلتين للتشوه ، وهما طريقة Chan-Vese و Snakes (Kass) لتجزئة صورة سرطان الثدي. من خلال إجراء دراسة مقارنة باستخدام قاعدة بيانات MIAS المتاحة للجمهور ، نقوم بتقييم فعالية هذه الأساليب في تقسيم مناطق سرطان الثدي بدقة ومقارنة أدائها من حيث الدقة والمتانة والكفاءة الحسابية. تهدف نتائج هذا البحث إلى المساهمة في رؤى نقاط القوة والقيود في الأساليب القابلة للتشوه للتجزئة ، مع إمكانية تعزيز تطوير خوارزميات أكثر قوة ودقة لتشخيص سرطان الثدي وتخطيط العلاج

Table of Contents

Table of figures

Introduction

Breast cancer is a formidable health concern affecting millions of individuals worldwide. With its potentially devastating consequences, early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. Medical imaging techniques play a pivotal role in breast cancer detection and diagnosis, with mammography being one of the primary imaging modalities employed in clinical practice. Mammograms provide detailed radiographic images of the breast tissue, enabling the identification of abnormalities and potential malignancies.

Segmentation, the process of delineating and separating regions of interest within an image, is a fundamental step in analyzing mammographic images for breast cancer diagnosis. Accurate segmentation of cancerous regions is essential for subsequent stages of analysis, such as feature extraction, classification, and treatment planning. However, due to the inherent complexity and variability of breast tissue, as well as the presence of noise, artifacts, and overlapping structures, segmentation of breast cancer from mammograms poses a significant challenge.

To address this challenge, various segmentation techniques have been developed, with deformable models gaining prominence in recent years. Deformable models are a class of algorithms that iteratively evolve a contour or surface to fit the desired boundaries within an image. These models offer the advantage of adaptability and flexibility, allowing them to capture the irregular shape and contour variations associated with breast tumors.

In this study, we focus on two deformable methods for breast cancer image segmentation: the Chan-Vese method and the Snakes (Kass method). The Chan-Vese method is a region-based deformable model that aims to find a globally optimal segmentation by iteratively minimizing an energy functional based on region properties. On the other hand, the Snakes method is an edge-based deformable model that seeks to identify object boundaries by minimizing an energy functional related to edge information.

To evaluate the performance and suitability of these deformable methods, we conducted a comparative study using the publicly available Mammographic Image Analysis Society (MIAS) database. This database contains a diverse collection of mammographic images, encompassing different breast tissues, lesion types, and imaging conditions. By applying the Chan-Vese and Snakes methods to the MIAS database, we aimed to assess their effectiveness in accurately segmenting breast cancer regions and compare their performance in terms of accuracy, robustness, and computational efficiency.

Through this research, we aim to contribute to the field of breast cancer image analysis by providing insights into the strengths and limitations of deformable methods for segmentation. The findings of this study may help improve the development of more robust and accurate segmentation algorithms, ultimately enhancing the efficacy of breast cancer diagnosis and treatment planning.

Chapter 1 BREAST CANCER

1.1 Introduction

With 3,500 deaths per year, breast cancer is the leading cause of death among women in Algeria. It accounts for 40.45% of all female cancers. It is the fifth leading cause of cancer-related deaths worldwide, accounting for 6.9% of all cancer deaths. Women are more susceptible to breast cancer's hold than males are since it primarily affects them. While men can also develop breast cancer, women are much more likely to do so. This huge gender gap emphasizes the need for awareness campaigns and preventative actions that are especially geared toward women. Moreover, preventing the terrible effects of breast cancer is greatly helped by early diagnosis. Early disease detection improves the possibility of effective therapy, lowers the danger of metastasis, and eventually raises the chance of survival. Early detection is a potent tool in the fight against breast cancer.

1.2 Understanding the breast

In humans and other mammals, the breast is a glandular organ located on the front of the chest. It is made up of specialized tissue that produces milk to feed offspring. The nipple, areola, ducts, lobules, each lobe contains numerous smaller structures known as lobules. These result in dozens of tiny bulbs capable of producing milk.

The lobes, lobules, and bulbs are all linked together by thin tubes known as ducts. These ducts lead to the nipple, which is located in the center of a dark area of skin known as the areola.

Front view of the female breast anatomy illustration   A side view of the anatomy of the female breast.The spaces between lobules and ducts are filled with fat. The breast has no muscles, but muscles lie beneath each breast and cover the ribs. Each breast also has blood vessels and lymphatic vessels. Lymph vessels connect to small bean-shaped organs known as lymph nodes. These lymph nodes are found in groups under the skin.

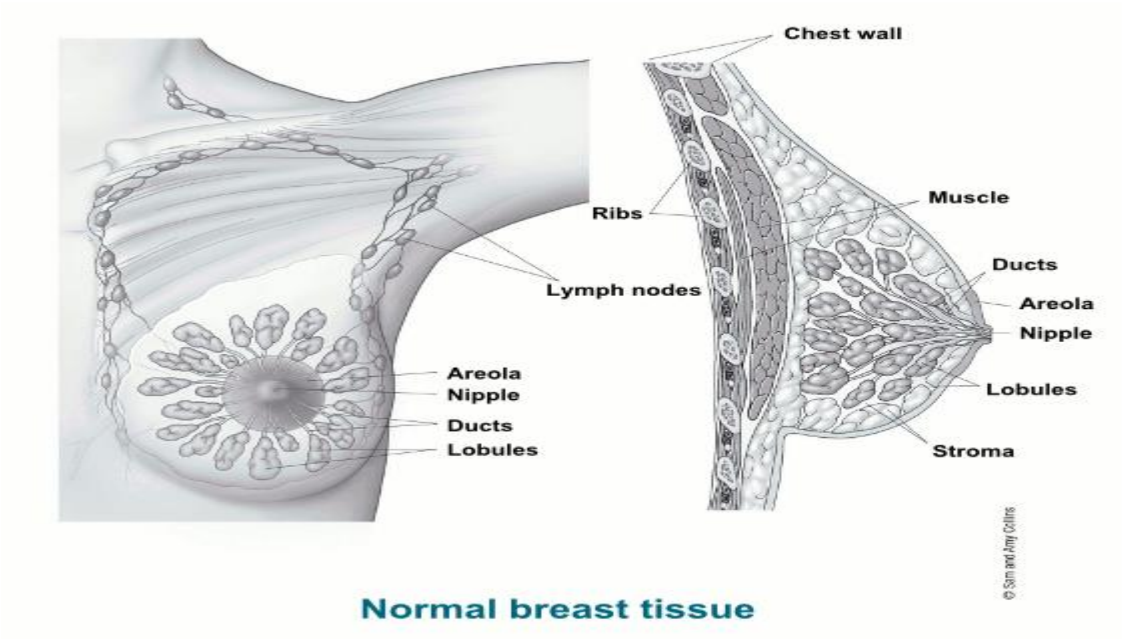


Figure 1-1 Structure of the breast [2]

1.3 Breast cancer

Breast cancer is a cancer that begins in the breast. It can begin in either one or both breasts.

Cancer develops when cells begin to proliferate uncontrollably.

Breast cancer affects almost entirely women, but men can develop it as well [2].

Breast cancer happens when cells in your breast grow and divide in an uncontrolled way, creating a mass of tissue called a tumor. Signs of breast cancer can include feeling a lump in your breast, experiencing a change in the size of your breast and seeing changes to the skin on your breasts. Mammograms can help with early detection[3].

1.4 Causes and risk of breast cancer

In the case of rare breast cancer, the causes and risk factors are the same. Causes and risk factors are related but not identical in the context of breast cancer. The term "causes" refers to the specific factors or events that contribute to the development of breast cancer. In some cases, mutations in certain genes, such as BRCA1 and BRCA2, have been identified as a cause of breast cancer. Risk factors, on the other hand, are factors that increase the developing breast cancer but do not necessarily cause it. Age, gender, family history, certain medical conditions, and lifestyle factors such as smoking, alcohol consumption, and physical activity level can all be risk factors.

1.4.1 Causes

1) A previous history of breast cancer

A woman who has had breast cancer back in her life may get cancer in the other breast again. [4].

2) Family history

If several members of the patient's family have had certain types of cancer, the patient may be at a higher risk of developing breast cancer**[**4].

3) Hormonal causes

A change in hormonal levels may lead to breast cancer. It could be accompanied by the onset and cessation of periods (Menstrual Cycle), early pregnancy, hormone replacement therapy, the use of oral pills, and so on [4].

4) Lifestyle and dietary causes

A sedentary lifestyle, a high fat diet, and obesity, particularly in postmenopausal women, can all lead to breast cancer. Another cause of breast cancer is the use of alcohol. The risk rises as the amount of alcohol consumed rises. Women who drink two to five alcoholic beverages per day have a risk of developing breast cancer that is roughly one and a half times that of nondrinkers [4].

7) Environmental causes

There is a slight increase in risk in women who work with low doses of radiation for an extended period, such as X-ray technicians[4].

1.4.2 Risk factors for breast cancer

1) Age

Aside from sex, one of the most important risk factors for breast cancer is aging, because the incidence of breast cancer rises with age." Women over the age of 40 accounted for 99.3% of all breast cancer-related deaths in the United States in 2016 3. As a result, women over the age of 40 should schedule a mammogram screening in advance [5].

2) Sex

Women are far more likely than men to develop breast cancer[3].

3) **Alcohol consumption**

According to research, drinking alcohol may increase your risk of developing certain types of breast cancer.[3]

4) **Dense breast tissue**

Women with dense breast tissue may be at a higher risk of developing breast cancer because there are more cancerous cells.

Dense breast tissue can also make a mammogram (breast scan) difficult to read because lumps or areas of abnormal tissue are more difficult to see.

Breasts are denser in younger women. As you age, the amount of glandular tissue in your breasts decreases and is replaced by fat, resulting in less dense breasts[6].

1.4.3 Symptoms

Knowing how your breasts look and feel is an important part of maintaining your breast health. Although regular breast cancer screening tests are important, mammograms do not detect all breast cancers. This means that you should be aware of any changes in your breasts by knowing what they normally look and feel like.[14]

A new lump or mass is the most common symptom of breast cancer (though most breast lumps are not cancer). Breast cancer is more likely to be a painless, hard mass with irregular edges, but it can also be soft, round, tender, or even painful.

Other possible breast cancer symptoms include:

. Swelling of the entire or a portion of a breast (even if no lump is felt)

. Skin dimpling (which can resemble an orange peel)

. Breast or nipple discomfort

. Nipple retraction (inward turn)

. Red, dry, flaking, or thickened nipple or breast skin

. Nipple discharge (aside from breast milk)

. Swollen lymph nodes under the arm or near the collarbone (This can sometimes indicate breast cancer spread before the original tumor in the breast is large enough to be felt.)

Many of these symptoms can be attributed to benign (non-cancerous) breast conditions. Still, any new breast mass, lump, or other change should be checked by an experienced health care professional so that the cause can be identified and treated, if necessary[14].

1.5 *Diagnosis of breast cancer*

Some benign breast changes may cause symptoms or signs (for example, breast lumps, pain, or nipple discharge), while others may be discovered during a mammogram. In either case, they can be difficult to distinguish from breast cancer, so additional exams or tests may be required to be certain [14].

1) **Mammogram**

Mammogram — A mammogram is a low-dose X-ray examination of the breast. The breast tissue is compressed for the X-ray, which reduces tissue thickness and holds the breast in place, allowing the radiologist to find abnormalities more accurately. To ensure that all tissue is examined, each breast is compressed between two panels and x-rayed from two directions (top-down and side-to-side). Mammograms are currently the most effective screening method for breast cancer. Some mammograms capture images digitally, which provides better clarity, the ability to adjust the image, and a lower likelihood of having to return on a different day for repeat images [8].

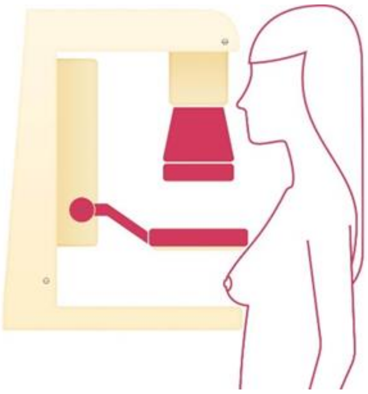


Figure 1-2: Mammogram diagnosis [10]

2) Biopsy

A biopsy is a test that takes a sample of cells from your breast and tests them to see if they are cancerous.

A scan and a needle test on lymph nodes in your armpit (axilla) may also be required to determine whether these are also affected[7].

3) Magnetic resonance imaging (MRI)

Breast MRI — Magnetic resonance imaging (MRI) creates a detailed image of a part of the body using a powerful magnet. It does not employ X-rays or radiation, but it does necessitate the injection of a contrast agent (a material that appears on imaging) into a vein. A blood test is performed prior to administering the contrast to ensure that you are able to receive the contrast[8].

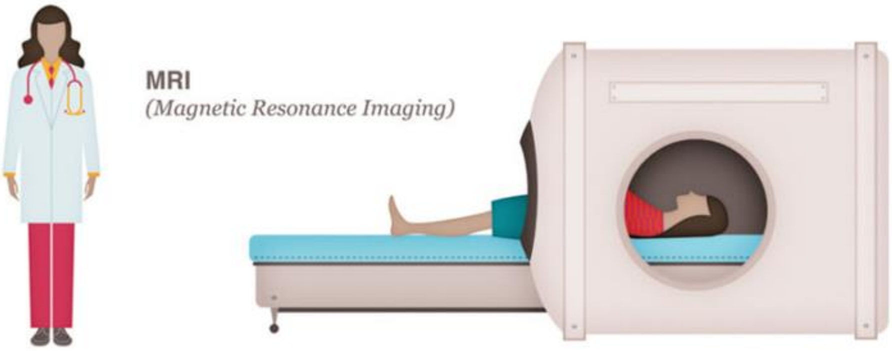


Figure 1-3: Magnetic resonance imaging [10]

4)Ultrasonographic screening

When a suspicious area is detected in your breast through a breast self-exam or on a screening mammogram, your doctor may request an ultrasound of the breast tissue. A breast ultrasound is a scan that uses penetrating sound waves that do not affect or damage the tissue and cannot be heard by humans. The breast tissue deflects these waves causing echoes, which a computer uses to paint a picture of what’s going on inside the breast tissue (no radiation is involved). A mass filled with liquid (which is a benign cyst) shows up differently than a solid mass [10].

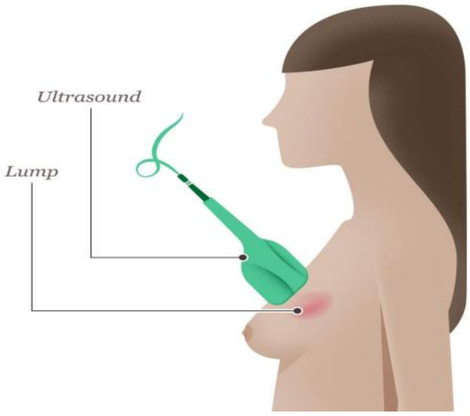


Figure 1-4: Breast ultrasound [10]

1.6 Stages of Breast Cancer

When a person is diagnosed with a malignant tumor or breast cancer, tests are performed to determine whether cancer cells have spread within the breast or to other parts of the body *[10]*.

There are five stages of breast cancer, starting at zero and going up to four:

1) **Stage 0**

The disease is not contagious. This means it hasn't escaped from your breast ducts [3].

2) Stage 1

Cancer cells have spread to neighboring breast tissue [3].

3)Stage 2

The tumor is either less than 2 centimeters in diameter and has spread to the underarm lymph nodes, or it is more than 5 centimeters in diameter but has not spread to the underarm lymph nodes. Tumors at this stage can range in size from 2 to 5 centimeters in diameter and may or may not affect nearby lymph nodes [3].

4)Stage 3

At this point, the cancer has spread beyond its original site. Although it has infiltrated nearby tissue and lymph nodes, it has not spread to distant organs. Stage III breast cancer is also known as locally advanced breast cancer [3].

5)Stage 4

The cancer has spread to other parts of your body, including your bones, liver, lungs, and brain. Breast cancer in stage IV is also known as metastatic breast cancer [3].

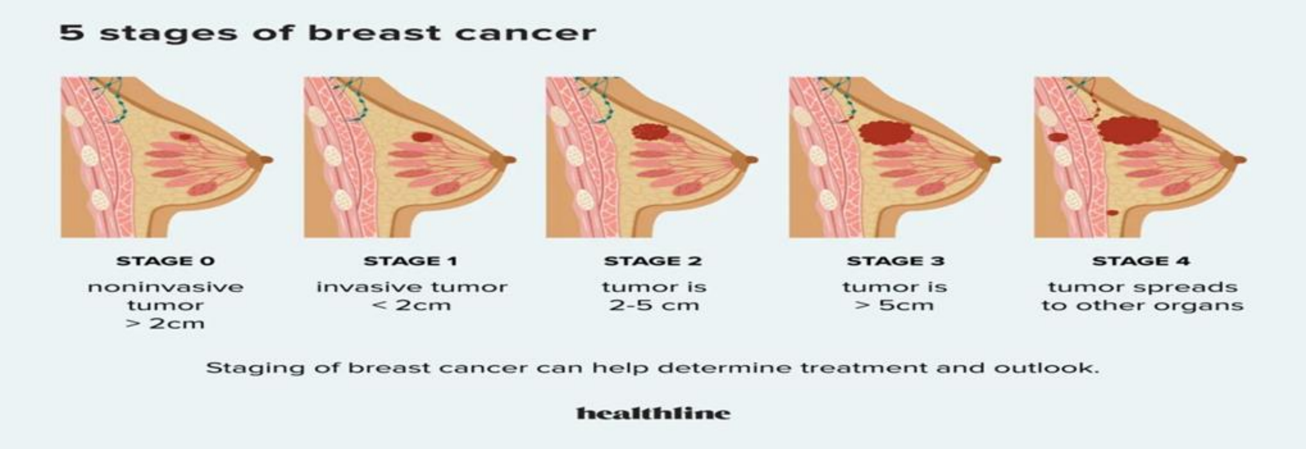


Figure 1-5: The five stages of breast cancer [9]

1.7 Types of breast cancer

There are numerous types of breast cancer and numerous ways to describe them. The specific cells in the breast that become cancer determine the type of breast cancer. The most common types of breast cancer are:

1.7.1 Non-invasive breast cancers

Breast cancer cells that are confined to the ducts and do not invade the surrounding fatty and connective tissues.

1) Ductal carcinoma in situ (DCIS)

DCIS, the most common type of non-invasive breast cancer, is limited to the breast ducts. Consider ductal comedocarcinoma [12].

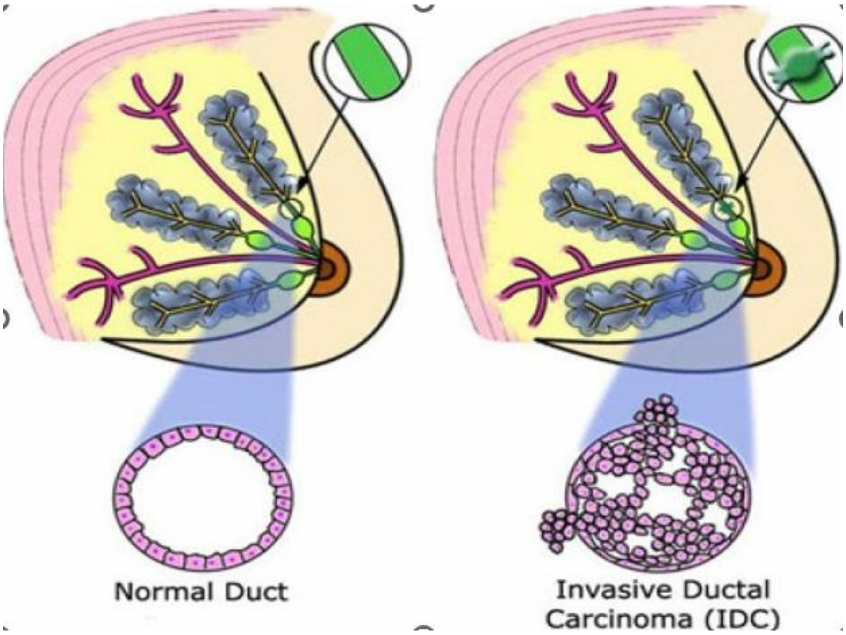


Figure 1-6 : Typical Structure associated with ductal carcinoma [12].

2) **Lobular carcinoma in situ (LCIS)**

LCIS is a type of non-invasive breast cancer that develops in the lobules (the milk-producing glands at the ends of the breast ducts). It isn't invasive because it hasn't spread to the surrounding breast tissue. Although LCIS is not fatal, it does increase the risk of developing invasive breast cancer later in life [11].

1.7.2 Invasive breast cancers

Invasive breast cancers have spread outside of the breast's ducts or lobules into surrounding breast tissue. The term 'early breast cancer' refers to cancer that is contained within the breast and may have spread to nearby lymph nodes in the breast or armpit (axilla) but not elsewhere in the body [11].

- The most common types of breast cancers are:

1) **Invasive ductal carcinoma**

IDC (invasive ductal carcinoma) is the most common type of breast cancer." Invasive ductal carcinomas account for approximately 80% of all breast cancers. Invasive ductal carcinoma occurs when a cancer that began in the milk ducts of the breast has spread through the duct lining and into the surrounding breast tissue. Invasive ductal breast cancer can spread to the lymph nodes and potentially other parts of the body over time [11].

2) Invasive lobular carcinoma

After invasive ductal carcinoma, invasive lobular carcinoma (ILC) is the most common type of breast cancer. Invasive lobular carcinoma indicates that the cancer began in the milk-producing lobules of the breast and has spread to surrounding breast tissue. Invasive lobular breast cancer can spread to the lymph nodes and potentially other parts of the body over time [11].

- Less common types of invasive breast cancer are:

1) Locally advanced breast cancer.

2) Metastatic breast cancer.

3) Inflammatory breast cancer.

1.8 Treatment options

Breast cancer treatment options are determined by a variety of factors, including the type and stage of the cancer, the patient's overall health, and their personal preferences. Surgery, radiation therapy, chemotherapy, hormone therapy, targeted therapy, or a combination of these treatments are the most common primary treatments for breast cancer.

1.8.1 Surgery

Surgery for breast cancer consists of two main options.

In breast-conserving surgery, only the tumor and an area of normal tissue surrounding it are removed. Breast conserving surgery includes the following:

Lumpectomy: A small amount of surrounding normal tissue is removed.

Wide excision: Also called as partial mastectomy in which somewhat larger amount of the surrounding normal tissue is removed

Quadrantectomy: About one fourth of the breast is removed [12].

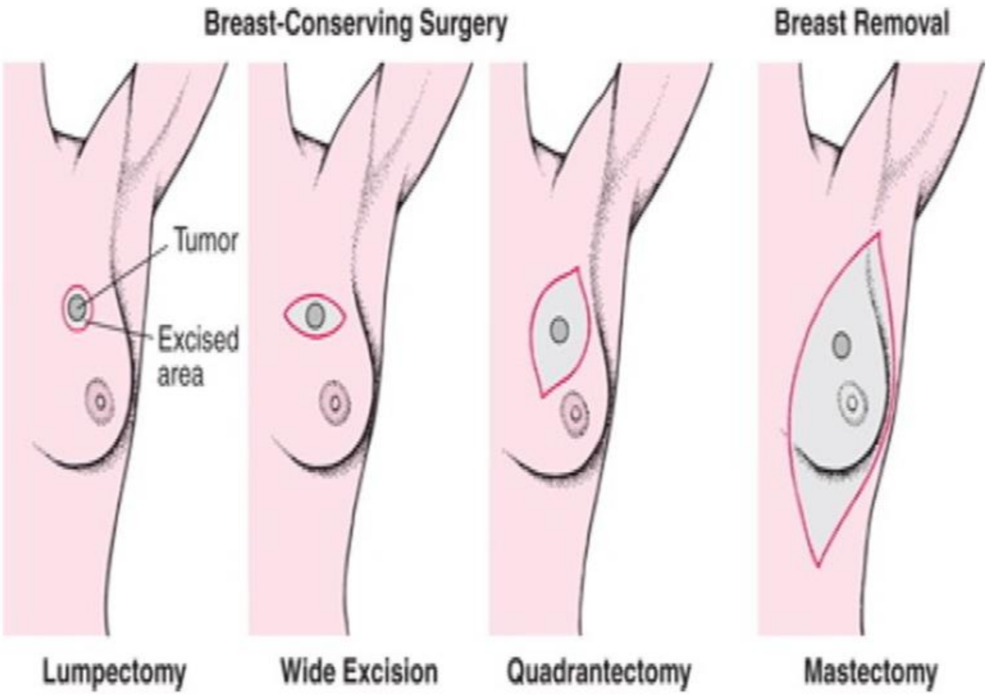


Figure 1-7:various types of surgery applied for breast cancer [12].

1.8.2 Radiation therapy

Radiation therapy is the application of high-energy X-rays or gamma rays to a tumor or post-surgery tumor site. These radiations are extremely effective at killing cancer cells that may remain after surgery or reappear in areas where the tumor was removed. Implanted radioactive catheters (brachytherapy), similar to those used in prostate cancer treatment, can be used in addition to this treatment. However, electron beam radiotherapy to the breast scar has surpassed this treatment option. Radiation therapy for breast cancer is typically administered following surgery and is an essential component of breast-conserving therapy. The radiation dose must be high enough to ensure the elimination of cancer cells. Treatments are typically administered over a five- to seven-week period, five days per week. Each treatment takes about 15 minutes [12].

Chemotherapy

Chemotherapy is the treatment of cancerous cells with anti-cancer drugs. Treatment for breast cancer will be determined by factors such as overall health, medical history, age (whether menstruation is present or not), type and stage of the cancer, tolerance for specific medications and procedures, and so on. Chemotherapy treatments are frequently administered in cycles, with one treatment followed by a recovery period and then another. Chemotherapy can be used prior to surgery to shrink the tumor and, in some cases, allow for breast conserving surgery rather than a mastectomy. It is frequently administered following surgery and may be administered every three weeks or every two weeks in a "dose dense" fashion. Agents commonly used in breast cancer chemotherapy include [12].

1.8.3 Hormone therapy

If your breast cancer has estrogen, progesterone, or both hormone receptors, your doctor will most likely recommend hormonal therapy, also known as anti-estrogen therapy or endocrine therapy. Hormonal therapy medications work by lowering estrogen levels in the body or by blocking estrogen's action on breast cancer cells [13].

1.9 Conclusion

In conclusion, breast cancer is a significant health concern affecting the breast tissue, and early detection is crucial for effective management. We discussed the definition, causes, diagnosis, and symptoms of breast cancer, emphasizing the importance of regular screening and awareness. Treatment options, including surgery, radiation therapy, chemotherapy, hormonal therapy, and targeted therapy; depend on the stage and type of cancer. Looking ahead, our next chapter will delve into the field of image processing and its role in improving diagnostic accuracy and treatment outcomes. By exploring advanced techniques in image analysis, we can enhance the detection and characterization of breast cancer, enabling more personalized and precise interventions.

Chapter 2 Image processing

2.1 Introduction

Image processing is to manipulate and analyze digital images using various algorithms and techniques. It involves techniques for enhancing, modifying, and extracting information from images to improve their visual quality, extract useful information, or enable automated analysis. Image processing is essential for image segmentation because it enables us to extract relevant features, and develop and optimize segmentation algorithms. Without image processing, it would be much more difficult to accurately and efficiently segment images, which is an important task in many applications.

2.2 Image

In the context of digital image processing, an image can be defined as a two-dimensional array of pixels, where each pixel represents a small area of the image and is assigned a numerical value that represents the brightness or color of that area.

An image can be defined as a two-dimensional function, f (x, y), where x and y are spatial coordinates and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point [15].

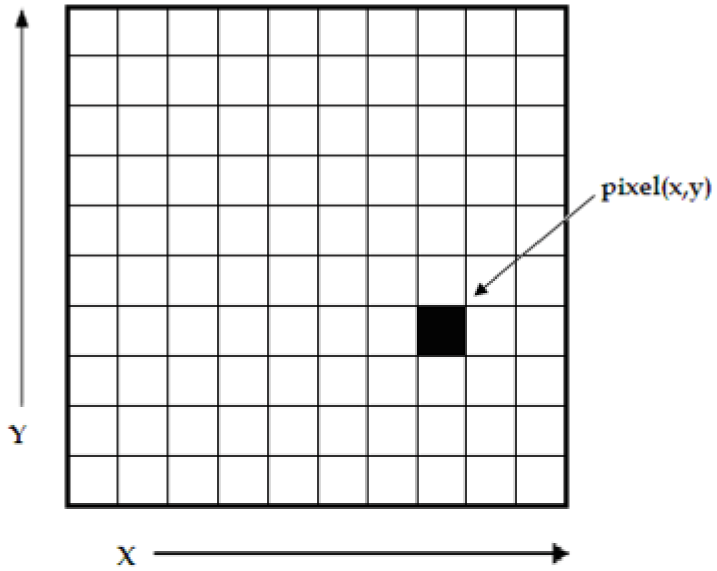


Figure 2-1: Image pixels[16]

2.3 Types of images

Images can be represented in various formats, including binary, grayscale, and color

2.3.1 Binary images

A binary image is a kind of digital image where each pixel can only have one of two possible values, typically black or white, or 0 or 1. A bi-level or black-and-white image is another name for it. The term "binary" refers to the fact that each pixel can have only one of two possible states. In a binary image, a single bit of information, represents each pixel where a value of zero usually corresponds to black or background, and a value of 1 corresponds to white or foreground. This representation allows for a simple and efficient way to store and process images, especially in applications that require binary operations like shape recognition, object detection, or image segmentation[15].

2.3.2 Grayscale images

A grayscale image is an image in which each pixel has a value between 0 and 255 that represents its brightness. In a grayscale image, the color of each pixel is represented by a single channel, and the intensity of that channel determines the pixel's brightness, each pixel is represented by a single channel, typically referred to as the gray channel or the intensity channel. The value of each pixel in the gray channel represents the brightness of that pixel, with 0 representing black and 255 representing white. For example, a pixel with a value of 128 would be a medium-gray color[15].

2.3.3 Color images

A color image is an image that contains color information, in addition to brightness and contrast information, for each pixel. A color image is typically composed of three channels, each representing a primary color, such as red, green, or blue. The combination of these primary colors results in a wide range of colors and shades. In color images each pixel is represented by three channels, typically referred to as the red, green, and blue channels (RGB). The value of each pixel in each channel represents the intensity of that color, with 0 representing no color and 255 representing maximum intensity. The combination of these three channels creates a wide range of colors and shades. For example, a pixel with a value of (255,0,0) would be pure red, while a pixel with a value of (128,128,128) would be a medium-gray color.[15]

Figure 2-2:a) Color image (b) Grayscale image (c) Binary image [16]

2.4 Image characteristics

2.4.1 Noise

Noise is a random variation of image Intensity and visible as grains in the image. It may arise in the image as effects of basic physics-like photon nature of light or thermal energy of heat inside the image sensors. It may produce at the time of capturing or image transmission. Noise means, the pixels in the image show different intensity values instead of true pixel values. Noise can degrade the image at the time of capturing or transmission of the image. Before applying image processing tools to an image, noise removal is very necessary. [17]

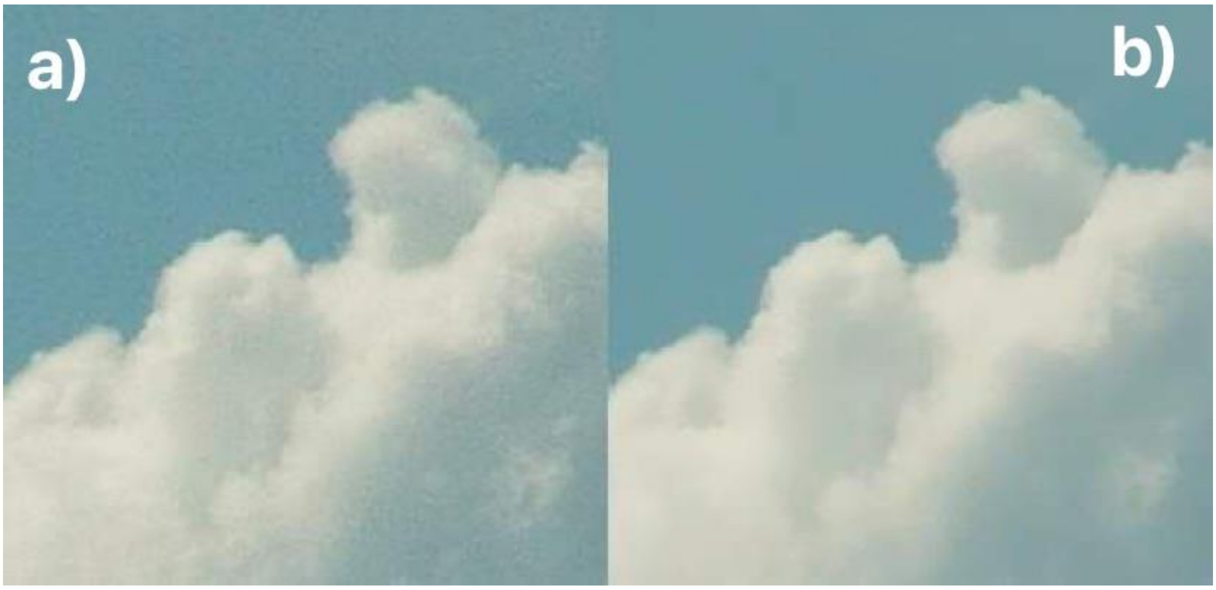


Figure 2-3: *a) Image with noise b) Image without noise [18]*

2.4.2 Histogram

The histogram in the context of image processing is the operation by which the occurrences of each intensity value in the image is shown. Normally, the histogram is a graph showing the number of pixels in an image at each different intensity value found in that image. For an 8-bit grayscale image there are 256 different possible intensities, and so the histogram will graphically display 256 numbers showing the distribution of pixels amongst those grayscale values. Histogram modification is a classical method for image enhancement [19]

2.4.3 Homogeneity

Homogeneity is local information that corresponds to the uniform character of a region. A region in an image is said to be homogeneous if it groups together a set of pixels that have similar or uniform characteristics. These characteristics can be for example the variance of the level of gray, the color, the texture [20].

2.5 Image acquisition

Image acquisition techniques refer to the methods used to capture or acquire digital images from a physical object or scene. There are various techniques for image acquisition, each with its own advantages and limitations. Some of the commun image acquisition techniques are :

1. Digital cameras: Digital cameras are one of the most widely used techniques for image acquisition. They use a sensor to capture light and convert it into digital data, which is then stored in the camera's memory or on a storage device. Digital cameras are available in different types, such as point-and-shoot, DSLR, mirrorless, and smartphone cameras, each with varying levels of image quality and features.

2. Scanners: Scanners are used to digitize physical documents or images, such as photographs, paintings, or drawings. They work by using a light source and a sensor to capture the reflected light from the document and convert it into digital data. Scanners are commonly used in offices, libraries, and archives for document digitization and preservation.

3. Microscopes: Microscopes are used to capture high-resolution images of small objects, such as cells, bacteria, or microstructures. They use a combination of lenses and sensors to magnify and capture the image of the object, which can then be viewed and analyzed on a computer or display device.

4. X-ray imaging: X-ray imaging is a technique used to capture images of the internal structures of objects, such as bones, organs, or machines. It works by passing X-rays through the object and detecting the radiation that is transmitted or scattered by the object. X-ray imaging is commonly used in medical imaging, industrial inspection, and security screening.

5. Magnetic resonance imaging (MRI): MRI is a medical imaging technique used to capture images of the internal structures of the body. It works by using a strong magnetic field and radio waves to generate images of the body's tissues and organs. MRI is commonly used in diagnosis and treatment planning for various medical conditions.

Image acquisition techniques are an important part of digital image processing, as they determine the quality and properties of the acquired image. The choice of image acquisition technique depends on the application, the type of object or scene being captured, and the desired level of image quality and resolution [15].

2.6 Image enhancement

Image enhancement refers to the process of improving the visual quality of a digital image by applying various techniques and algorithms. The goal of image enhancement is to make the image more visually appealing, easier to analyze, or more suitable for a particular application.

There are various techniques and algorithms that can be used for image enhancement, such as:

1. **Contrast enhancement:** This technique is used to improve the contrast between different regions of an image, making it easier to distinguish between them. Examples of contrast enhancement techniques include histogram equalization and contrast stretching.

2. **Color adjustment:** This technique is used to adjust the color balance and intensity of an image. Color adjustment techniques can be used to correct colorcast, white balance, or saturation issues in an image.

3. **Noise reduction:** This technique is used to reduce the amount of noise in an image, which can be caused by low light conditions, a high ISO setting, or a poor quality sensor. Examples of noise reduction techniques include median filtering and wavelet denoising.

4. **Sharpening:** This technique is used to enhance the details and edges in an image, making it appear sharper and more defined. Examples of sharpening techniques include unsharp masking and Laplacian filtering.

Image enhancement is a fundamental step in digital image processing and is widely used in various fields, such as photography, medical imaging, and computer vision. By improving the visual quality of an image, image enhancement can make it easier to analyze, interpret, and extract useful information from the image [15].

2.7 Image restoration

Image restoration is the process of recovering a degraded or corrupted digital image to improve its visual quality and restore its original appearance. Image degradation can be caused by various factors, such as noise, blur, and compression artifacts, which can affect the quality of the image and make it more difficult to analyze or interpret.

The goal of image restoration is to remove or reduce the effects of degradation and recover the original image as much as possible. This can be achieved through various techniques and algorithms, such as:

1. **Filtering:** This technique is used to remove noise and blur from an image by applying various filters, such as median filtering, Wiener filtering, or Gaussian filtering. Filtering can help to smooth out the image and make it more visually appealing, but it may also remove some of the fine details in the image.
2. **Deconvolution:** This technique is used to remove blur from an image by estimating the point spread function of the blur and applying inverse filtering or regularized deconvolution. Deconvolution can help to recover the fine details in an image and improve its sharpness, but it may also introduce noise or artifacts if the estimation of the point spread function is inaccurate.
3. **Super-resolution:** This technique is used to increase the resolution and detail of an image by combining multiple low-resolution images or using deep learning algorithms to generate high-resolution images. Super-resolution can be useful for applications such as satellite imaging or medical imaging, where high-resolution images are required for accurate analysis.

Image restoration is an important step in digital image processing, as it can help to improve the quality and usefulness of a degraded image. However, the success of image restoration depends on various factors, such as the degree of degradation, the properties of the image, and the accuracy of the restoration technique [15].

2.8 Conclusion

Image processing is of utmost importance in image segmentation as it provides the necessary tools, techniques, and algorithms for preprocessing, feature extraction, and implementing segmentation algorithms. Through preprocessing, image quality is enhanced, noise is reduced, artifacts are corrected, and image properties are normalized, enabling more accurate and reliable segmentation results. Image processing techniques also facilitate feature extraction, allowing for the extraction of relevant features such as color, texture, intensity, shape, and spatial relationships. These features serve as important cues for differentiating and classifying regions or objects during the segmentation process. Moreover, image processing plays a vital role in designing and implementing segmentation algorithms by leveraging operations like filtering, edge detection, morphological operations, clustering, and region growing. This leads to improved accuracy and reliability in segmenting images, as image processing operations enhance the visibility of object boundaries, reduce noise interference, handle illumination variations, and mitigate the impact of artifacts. Overall, image processing is integral to the success of image segmentation, enhancing its accuracy, reliability, and efficiency in extracting meaningful information from image.

Chapter 3 Segmentation

3.1 Introduction

Breast cancer is one of the most common cancers affecting women worldwide. Early detection of breast cancer is crucial for successful treatment, and mammography is an effective imaging technique for breast cancer screening. However, mammograms can be difficult to interpret due to the complex anatomy of the breast, and computer-aided diagnosis

(CAD) systems have been developed to assist radiologists in detecting and diagnosing breast cancer. Mammogram segmentation is an important step in CAD systems, as it involves identifying and extracting the breast tissue from the background and other structures, such as the pectoral muscle and adipose tissue. Accurate segmentation of mammograms can improve the performance of CAD systems and reduce the workload of radiologists. In recent years, various segmentation techniques, including active contour models, deep learning-based methods, and hybrid approaches, have been proposed for mammogram segmentation. This chapter will provide an overview of mammogram segmentation using active contour and its application in breast cancer diagnosis.

3.2 Mammogram

A mammography is a type of medical imaging that uses low-dose X-rays to create images of the breast tissue. It is a standard screening tool for the detection of breast cancer in its early stages, before any symptoms are present. During a mammogram, the breast is compressed between two plates to spread out the tissue and create a clearer image. The images are then examined by a radiologist who can identify any abnormalities or signs of cancer [21].

3.3 Types of mammograms

3.3.1 Screening mammogram

A screening mammogram is a type of mammogram that is used for breast cancer screening in women who have no symptoms of breast cancer. It involves taking two X-ray images of each breast from different angles to detect any changes or abnormalities in the breast tissue. The American Cancer Society recommends that women aged 40 and older should have annual screening mammograms [22].

3.3.2 Diagnostic mammogram

Diagnostic mammogram is a type of mammogram that is used for further evaluation of the breast when a change or abnormality has been detected, such as a lump, nipple discharge, or breast pain. It involves taking additional X-ray images from different angles to get a closer look at the area of concern. Diagnostic mammograms may also include additional imaging tests, such as ultrasound or magnetic resonance imaging (MRI), depending on the specific situation [1].

In addition, there are also some specialized types of mammograms that may be used in certain cases:

3.3.3 3D mammography (digital breast tomosynthesis):

This is an advanced type of mammography that takes multiple X-ray images of the breast from different angles to create a 3D image of the breast tissue. This may be used in combination with a standard mammogram for women with dense breast tissue or for further evaluation of an abnormality found on a standard mammogram [23].

3.3.4 Contrast-enhanced mammography:

This type of mammogram involves injecting a contrast dye into the bloodstream to enhance the visibility of any abnormal areas in the breast tissue. It may be used for women with a high risk of breast cancer or for further evaluation of a suspicious area found on a standard mammogram [22].

3.4 The main differences between screening mammograms and diagnostic mammograms are:

1. **Purpose:** The purpose of a screening mammogram is to detect breast cancer in women who do not have any symptoms or signs of the disease. A diagnostic mammogram is done to evaluate a breast problem that has been found by the woman or her doctor, such as a lump, nipple discharge, or a suspicious finding on a screening mammogram.

2. **Number of images**: A screening mammogram typically involves two X-ray images of each breast, while a diagnostic mammogram may involve additional images that focus on a specific area of the breast where a problem has been detected.

3. **Compression:** During a screening mammogram, the breast is compressed between two plates to obtain a clear X-ray image. During a diagnostic mammogram, the breast may be compressed more firmly or from different angles to obtain additional images of the area of concern.
4. **Follow-up:** If a screening mammogram detects a suspicious finding, a woman may be called back for additional imaging or testing, which may include a diagnostic mammogram. If a diagnostic mammogram detects a problem, additional tests, such as a breast ultrasound or biopsy, may be needed to make a diagnosis.

3.5 Craniocaudal view and Mediolateral oblique view:

CC view and MLO view are two common projections used in mammography to examine the breast tissue from different angles.

1.3.1 The CC view, or craniocaudal view, is an X-ray image of the breast taken from above and angled downward. This view allows for the visualization of the breast tissue from the nipple to the chest wall.

1.3.2 The MLO view, or mediolateral oblique view, is an X-ray image of the breast taken from the side at an oblique angle. This view allows for the visualization of the breast tissue from the upper to the lower regions of the breast.

Both views are important in mammography, as they can help detect abnormalities in different areas of the breast tissue. According to the American College of Radiology, the combination of CC and MLO views is the standard protocol for screening mammography [22].

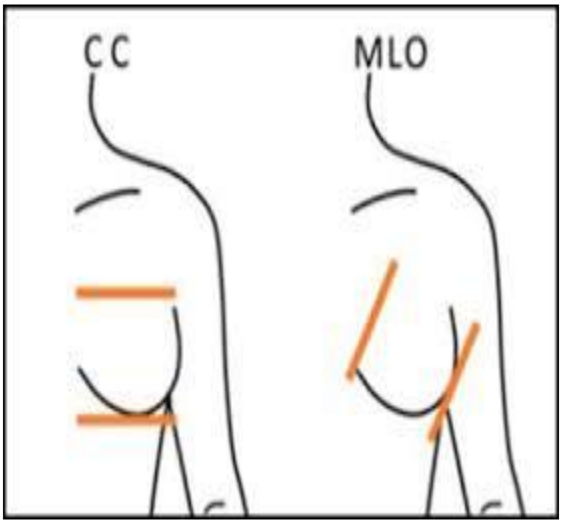
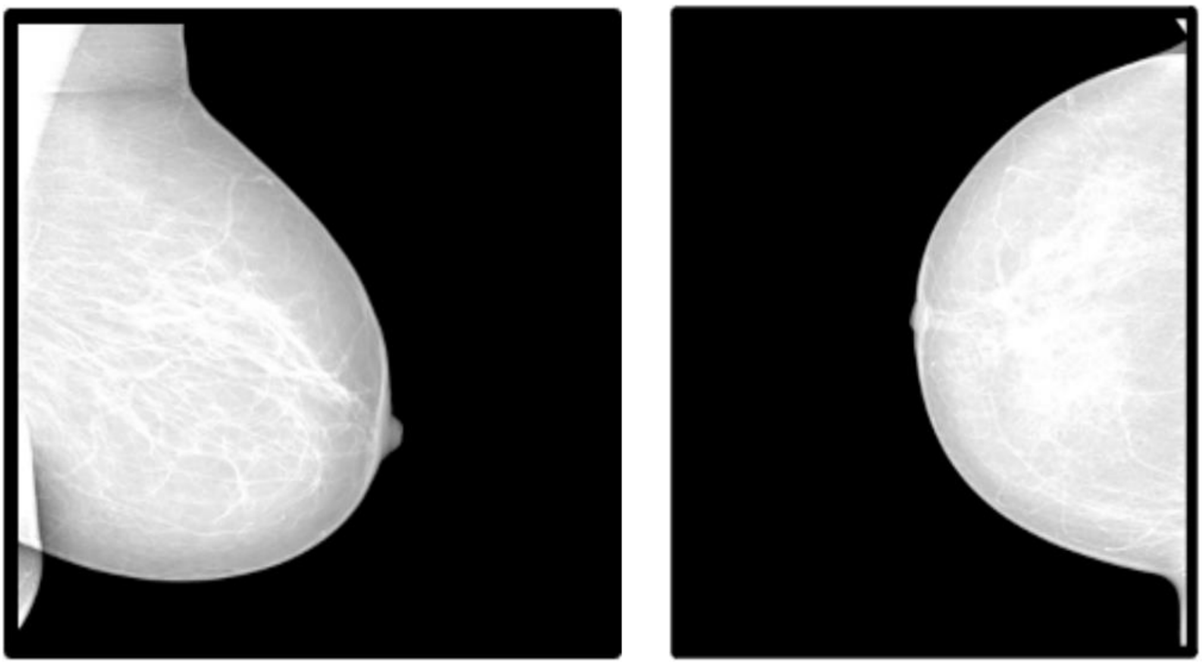


Figure 3-1: The difference between cc and Mlo views [24]



[(b) MLO view] [(a) CC view]

Figure 3-2: Cc view mammogram and view mammogram [24]

3.6 Mammogram segmentation

Mammogram segmentation refers to the process of identifying and extracting the regions of interest (ROIs) from a mammogram image, which can be used for further analysis.

Mammogram segmentation can be classified into three categories:

- Manual segmentation
- Semi-automatic segmentation
- Fully automatic segmentation.

Manual segmentation involves human experts manually marking the ROIs on the mammogram image. Semi-automatic segmentation involves a combination of manual and automated techniques, where the initial segmentation is performed automatically, and then corrected by human experts if necessary. Fully automatic segmentation uses computer algorithms to automatically segment the ROIs without human intervention [25].

3.7 Image segmentation:

Image segmentation refers to the process of partitioning a digital image into multiple segments, each representing a distinct region or object within the image. The goal of image segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

Image segmentation is the process of partitioning an image into multiple segments, each of which corresponds to a different region or object within the image. The goal of image segmentation is to simplify the representation of an image into meaningful and useful parts, with each part being easier to analyze and manipulate [15].

3.8 Segmentation techniques

3.8.1 Thresholding Segmentation

This method separates objects based on their intensity values, which can be set to a fixed threshold or adaptively determined [26].

Advantages:

- Simple and computationally efficient: Thresholding segmentation is a simple and computationally efficient method that can be implemented quickly on a wide range of hardware and software platforms.
- Easy to understand and use: It is easy to understand and use because it involves only a single parameter, which is the threshold value.
- Effective in separating foreground and background: Thresholding segmentation is effective in separating the foreground and background of an

image, making it useful in applications such as image binarization, object detection, and image enhancement.

Disadvantages:

- Sensitive to noise and illumination variations: Thresholding segmentation is sensitive to noise and variations in illumination, which can affect the accuracy of the segmentation result.
- Limited ability to handle complex images: Thresholding segmentation is not suitable for images that have complex backgrounds or foregrounds, as it may not accurately separate the regions of interest.
- May require preprocessing: Thresholding segmentation may require preprocessing to enhance the contrast of the image and remove noise to improve the accuracy of the segmentation result.

3.8.2 Edge-based Segmentation

This method detects edges in the image and segments regions based on the discontinuities in the intensity values [26].

Advantages:

- Effective in segmenting objects with distinct boundaries: Edge-based segmentation is particularly effective in segmenting objects with clear boundaries, such as objects with sharp edges or corners.
- Can be combined with other techniques: Edge-based segmentation can be combined with other segmentation techniques, such as region-based segmentation or clustering-based segmentation, to improve the accuracy of the segmentation result.
- Robust to noise: Edge-based segmentation can be more robust to noise compared to other segmentation techniques, as it focuses on detecting abrupt changes in intensity.

Disadvantages:

- Sensitive to the quality of edge detection: Edge-based segmentation is highly dependent on the quality of edge detection, which can be affected by noise, illumination, or contrast variations in the image.
- May not accurately segment objects with ambiguous boundaries: Edge-based segmentation may not accurately segment objects with ambiguous or blurry boundaries, as detecting edges can be difficult in such cases.
- May result in over-segmentation or under-segmentation: Edge-based segmentation may result in over-segmentation (splitting one object into multiple segments) or under-segmentation (merging multiple objects into a single segment) of objects, depending on the quality of edge detection.
- Computationally expensive: Edge-based segmentation can be computationally expensive, as it requires detecting edges by performing several convolutions and edge detection algorithms, which can slow down the segmentation process.

3.8.3 Region-based Segmentation

This method partitions an image into regions with similar properties, such as color, texture, or intensity [26].

Advantages:

- Can accurately segment objects with non-uniform properties: Region-based segmentation can accurately segment objects with non-uniform properties, such as images with non-uniform lighting, shadows, or reflections.
- Robust to noise and edge variations: Region-based segmentation is robust to noise and edge variations, as it focuses on identifying regions with similar properties rather than detecting edges or thresholds.
- Can be used for texture segmentation: Region-based segmentation can be used for texture segmentation, where the goal is to segment regions with similar textures, such as in medical imaging or material science.

Disadvantages:

- Computationally expensive: Region-based segmentation can be computationally expensive, as it requires performing iterative optimization or clustering algorithms to segment the image into regions.
- Requires careful parameter selection: Region-based segmentation requires careful selection of parameters, such as the number of regions, the similarity measure, and the stopping criteria, which can be challenging and time-consuming.
- May over-segment or under-segment objects: Region-based segmentation may over-segment objects by splitting them into smaller regions, or under-segment objects by grouping them with similar but distinct regions, depending on the quality of the similarity measure.

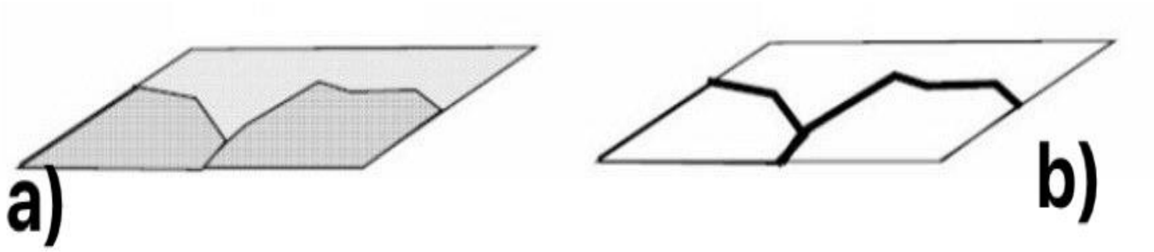


Figure 3-3*:* a) Region based segmentation b) Edge based segmentation

3.8.4 Clustering-based Segmentation

This method groups pixels into clusters based on their similarity in features such as color, texture, or intensity [26]

3.8.5 Watershed Segmentation

This method considers the image as a topographical surface and simulates flooding of the surface to create basins, which correspond to image regions [26].

3.8.6 Semantic Segmentation

This method assigns each pixel in the image a label that corresponds to a specific object or region in the scene [26].

3.8.7 Instance Segmentation

This method identifies and segments individual objects in the image, where each object is assigned a unique label [26].

3.8.8 Hybrid Segmentation

This method combines two or more of the above segmentation techniques to achieve better segmentation results [26].

3.9 Active contour model

The active contour method is a group of image segmentation techniques that employ flexible curves or surfaces to automatically identify and outline object boundaries in an image. This approach involves iteratively refining the contour to minimize an energy function, which typically consists of two components: an internal energy term ensuring smoothness of the contour and a data term capturing the desired properties of the segmented objects.

Snakes, also referred to as active contours, are a specific formulation of the active contour method. They are parametric curves or surfaces that deform in order to align with object boundaries by minimizing an energy function. Snakes are commonly utilized for tasks such as object localization and boundary extraction.

Chan-Vese is a variational active contour model that incorporates information about regions into the segmentation process. It utilizes the Mumford-Shah functional to minimize the energy function, which includes terms related to data fitting, regularization, and a constant term representing the segmentation contour.

Level set methods are numerical techniques employed to implicitly evolve the contour. Rather than explicitly representing the contour as a parametric curve or surface, it is defined as the zero level set of a higher-dimensional function. By solving a partial differential equation (PDE) that describes the evolution of the level set function, the contour can be continuously refined to capture object boundaries [26].

3.9.1 Snakes (Edge-based active contour model)

Michael Kass's active contour model, commonly known as the snake model, is a popular technique for image segmentation. It was introduced in the paper titled "Snakes: Active Contour Models" published in 1987 by Michael Kass, Andrew Witkin, and Demetri Terzopoulos. The snake model is based on the concept of deformable contours that can adapt and fit object boundaries in an image. The snake begins with an initial contour, which is typically manually placed close to the object boundaries of interest, the contour evolution is driven by an energy functional that combines internal and external energies while the internal energy term, also known as the elasticity term, encourages smoothness and controls the contour's deformation. It penalizes high curvatures, long contour lengths, or other deformations undesirable for segmentation. The external energy term, derived from image gradients or other image features, attracts the contour towards object boundaries. It encourages the contour to align with edges or regions of interest. The energy functional is minimized iteratively to deform the contour and converge towards the desired object boundaries by using gradient descent or other optimization techniques. The snake model has been extended and modified by researchers to enhance its performance and address specific challenges [27].

In the following, I denotes an image on a domain$\Omega \subset \mathbb{R}^2$. The active contour model is defined as a functional applied to a curve in the image domainΩ. Formally, we aim to compute a path v

$$\begin{aligned} v\colon [0,1] &\to \Omega \subset \mathbb{R}^2 \\ s &\mapsto v(s) = (x(s), y(s)), \end{aligned} \tag{1}$$

Such that $v(0) = v(1)$ and $v \in \mathcal{C}^2$. At the position s, the active contour is represented by $v(s) = (x(s), y(s))$ where x and y are snake point coordinate vectors, defined on Ω.

Let us consider the following energy

$$E(s) = \int_0^1 E_{\text{int}}(s) + E_{\text{ext}}(s)\mathrm{d}s \tag{2}$$

Where E_{int} denotes the internal energy and E_{ext} the external energy. The internal energy penalizes non-smooth curves and its minimization favors the regularity of v. This is the a priori part of the functional. The external energy depends on the image and enables the curve to fit the area of interest. This last term can be seen as the data fidelity term of the functional.

Internal energy. To regularize the shape of the snake, the internal energy can be written as

$$E_{\text{int}} = \frac{1}{2}\left(\alpha(s)\|v'(s)\|^2 + \beta(s)\|v''(s)\|^2\right) \tag{3}$$

Where $\alpha(s)$ and $\beta(s)$ are weights, which control the snake regularity. In the following, α and β are considered constant to simplify the problem for a numerical application. The minimization of the term E_{int} enforces the regularity of the curve.

External energy. The external energy represents the data-fidelity term. It is defined as a potential P on the image domain Ω

$$\begin{aligned} P\colon \Omega &\to \mathbb{R} \\ (x,y) &\mapsto P(x,y), \end{aligned} \tag{4}$$

Such that the minimum of P is the contour of interest in the image. In case of detection of a black contour in a white background, the potential can be the image itself. For smooth images, we consider that the external energy is only represented by the image gradient. Indeed, the active contour must be brought to high gradient areas (which represent image edges) to make the snake close to the contours. For this purpose, the image gradient has to be as high as possible. The external energy can therefore be represented by $E_{ext} = -P(v(s))$ where

$$P(x,y) = \left\| \left(\frac{\partial I}{\partial x}(x,y), \frac{\partial I}{\partial y}(x,y) \right) \right\|^2. \tag{5}$$

This potential has an analogous definition with the Canny edges detector.

Global energy. Combining Equations (2), (3) and (6), our problem can be written as

$$\min_{v} \int_0^1 G(s, v, v', v'')\mathrm{d}s \tag{6}$$

With

$$G(s, v, v', v'') = \frac{1}{2}\left(\alpha \|v'(s)\|^2 + \beta \|v''(s)\|^2\right) - P(v(s)) \tag{7}$$

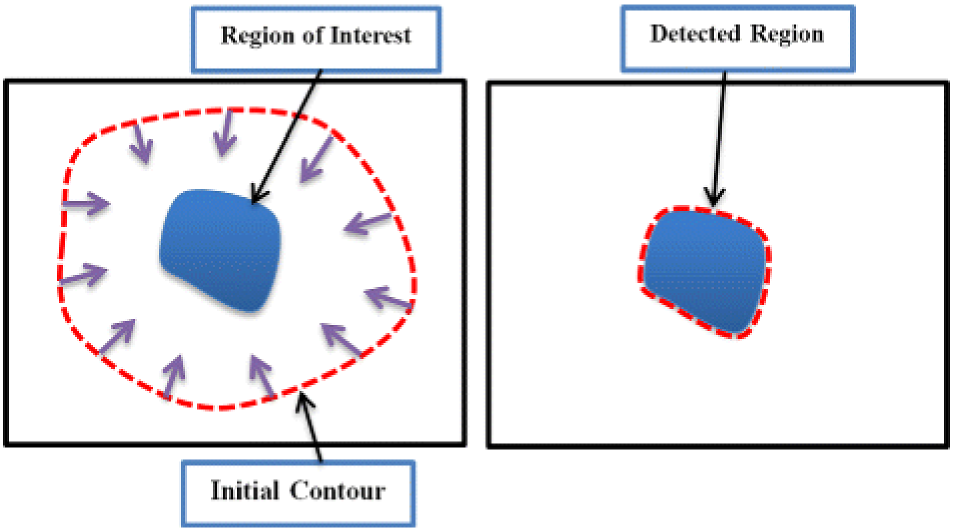


Figure 3-4: Level set models concept [16]

3.9.2 *The region-based active contour model*

Also known as Chan-Vese model is a popular variation of the active contour model that is commonly used for image segmentation, it was introduced in the year 2001 and proposed by Tony Chan and Luminita Vese in their paper titled "Active Contours Without Edges," published in the journal IEEE Transactions on Image Processing. Unlike traditional active contour models that use image gradients to detect object boundaries, the Chan-Vese model uses region-based information to segment the image. Chan-Vese's active contour model defines an energy functional that consists of region-based terms. It uses statistical properties of the object and background regions to drive the segmentation. The energy functional is minimized to find the optimal segmentation that balances the homogeneity within regions and the smoothness of the contour by capturing the difference in pixel intensities between the object and background regions. The Chan-Vese model does not require an initial contour. Instead, it initializes the segmentation by dividing the image into two regions, object and background. The Chan-Vese model is known for its robustness to noise and its ability to handle images with weak or indistinct edges [28].

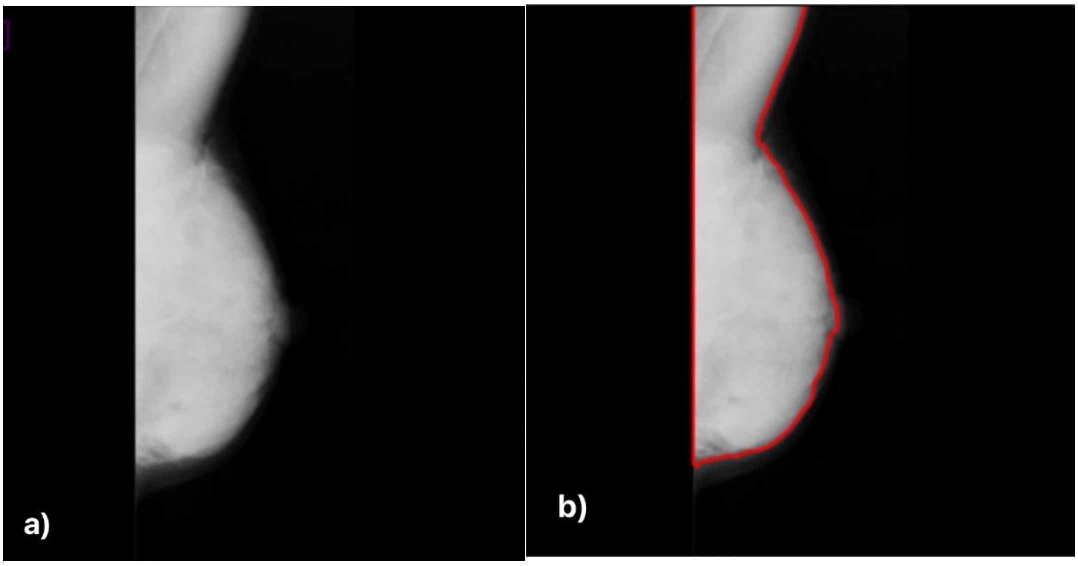


Figure 3-5:a) Mammogram without segmentation b) Mammogram with Region based segmentation (Chan-Vese)

3.10 Conclusion

Active contour is a powerful technique for image segmentation that can be used to extract the boundaries of objects from an image. The method uses a contour that is iteratively adjusted to fit the edges of the object by minimizing an energy function that takes into account the desired properties of the contour and the intensity gradients of the image. There are several different types of active contour methods that can be used for image segmentation, each with its own strengths and weaknesses.

Chapter 4: Tools and results

Chapter 4 Tools and results

4.1 Introduction

In this chapter, our focus is on image segmentation of the MIAS (Mammographic Image Analysis Society) dataset. We use the active contour method, specifically the region-based active contour model proposed by Chan and Vese, to accomplish this. We implement this algorithm using various libraries and tools available in the Python ecosystem, leveraging the power of Python. To make our development process easier, we use Google Colab, a web-based integrated development environment (IDE) that allows us to write and execute Python code in the cloud. We hope to accurately segment the images and extract meaningful information for further analysis and research by combining Python, the region-based active contour algorithm, and the MIAS dataset.

4.2 Tools used

We used Python as the programming language of choice due to its versatility and rich ecosystem. We used Google Colab, a web-based integrated development environment (IDE) that allowed us to write and execute Python code in the cloud, to speed up our word. We incorporated specific libraries tailored to our project's needs, such as NumPy, Pandas, and Matplotlib, leveraging the power of Python's extensive library support and allowing us to efficiently manipulate and visualize data. We used the MIAS (Mammographic Image Analysis Society) database, which is a popular resource in breast cancer research. We were able to tackle our project effectively and derive meaningful insights from it by leveraging Python, Google Colab, specific libraries, and the MIAS database.

4.3 Environment of execution

4.3.1 Google colab

Google Colaboratory, or "Colab" as it is colloquially known, is a cloud-based Jupyter notebook environment. It runs in your browser and allows anyone with internet access to experiment with machine learning and artificial intelligence coding. You can write and run Python code, as well as share and edit it with other team members [27].

Figure 4-1*:* Google colab logo

4.3.2 Google colab advantages

Free computing resources

Colab enables users to take advantage of powerful hardware resources without the need for costly local setups. This is especially useful for tasks that require a significant amount of computational power, such as training deep learning models.

Integration with Jupyter Notebook

Colab is built on Jupyter Notebook, which provides an interactive coding environment. It allows you to experiment, iterate, and document your work by allowing you to write and execute code in cells.

Colab supports real-time collaboration

allowing multiple users to work on the same notebook at the same time. Notebooks can be shared with others by sending them a link, making it easy for collaboration and code sharing.

Pre-installed libraries and packages

Colab includes a number of libraries and packages that are commonly used in data science and machine learning, such as TensorFlow, PyTorch, NumPy, and pandas. Because you don't have to manually install and configure these libraries, you save time and effort.

Version control and history

Colab automatically saves the version history of your notebooks, allowing you to roll back to previous versions if necessary. This feature adds an extra layer of security and allows you to track changes over time.

Educational resources and tutorials

Colab have a repository of example notebooks and tutorials contributed by the community. This is a valuable resource for learning new concepts, experimenting with different techniques, and gaining insights from experts in various fields.

Integration with Google services

Because Colab was created by Google, it seamlessly integrates with other Google services. You can, for example, load and process data directly in your Colab notebook using Google Sheets or BigQuery.

Extensibility and customization

Colab allow you to install additional packages and libraries, allowing you to tailor your environment to your specific requirements. Additionally, you can run shell commands and install system-level dependencies.

Portability and accessibility

Colab is a web-based platform, which means you can access your notebooks from anywhere with an internet connection. It eliminates the need for local installations and configurations, making it easy to switch between devices or work remotely.

4.3.3 Python

Python is an object-oriented, high-level programming language with dynamic semantics that is interpreted. It is high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very appealing for use as a scripting or glue language to connect existing components together. Python's simple, easy-to-learn syntax emphasizes readability, lowering program maintenance costs. Python provides support for modules and packages, which promotes program modularity and code reuse. The Python interpreter and extensive standard library are free to use and distribute in source or binary form for all major platforms[29].

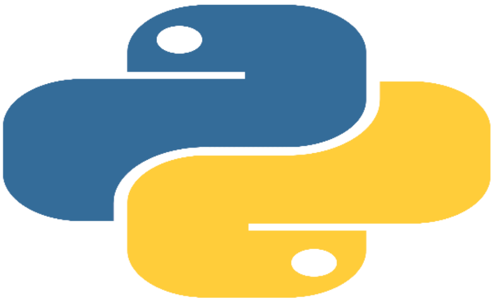


Figure 4-2: Python logo [29]

4.3.4 Python libraries

Pandas

Pandas is a powerful open-source library in Python that provides data structures and data analysis tools for efficient handling and manipulation of structured data. It simplifies tasks such as data cleaning, exploration, and analysis, making it an essential tool for data scientists and analysts.

Figure 4-3: Pandas library logo.

NumPy

NumPy is a fundamental library in Python for numerical computing. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently. NumPy is widely used in scientific computing and forms the foundation for many other Python libraries.

Figure 4-4: NumPy library logo.

Matplotlib

Matplotlib is a comprehensive plotting library in Python that enables the creation of high-quality 2D and 3D visualizations. It offers a wide range of plotting options and customization features, making it suitable for various data visualization needs. Matplotlib is extensively used for generating charts, graphs, histograms, and other visual representations of data.

Figure 4-5: Matlplotlib library logo.

Pylab

Pylab is a module in the Matplotlib library that combines the functionality of Matplotlib with NumPy in an interactive environment. It provides a convenient interface for creating plots and visualizations by importing both Matplotlib and NumPy together. Pylab allows users to generate plots and explore data interactively, making it useful for data analysis and exploration tasks.

4.3.5 Jupyter notebook

Jupyter Notebook is an open-source web application that allows you to create and share documents with code, equations, visualizations, and explanatory text.

It enables us to clean and transform data, perform numerical simulation, statistical modeling, and machine learning, among other things.

4.3.6 Database

We utilized the MIAS (Mammographic Image Analysis Society) dataset that is a collection of mammographic images that were obtained through various sources and protocols. These images were collected over a period of time and made available for research and development purposes within the field of mammographic image analysis. The MIAS (Mammographic Image Analysis Society) dataset contains 322 images, each 1024 by 1024 pixels in size. This dataset has seen widespread use in breast cancer research and analysis. The MIAS dataset, with its large collection of high-resolution mammographic images, has played a critical role in the advancement of algorithms and techniques for breast cancer detection and diagnosis. This dataset is used by researchers and practitioners to train machine-learning models, evaluate image segmentation methods, and conduct a variety of studies aimed at improving the accuracy and efficiency of breast cancer screening and assessment. The MIAS dataset's large number of images and high resolution make it a valuable resource for both academic and clinical applications in the field of mammography.

4.3.7 How to read MIAS images

1st column: MIAS database reference number.

2nd column: Character of background tissue:

F - Fatty

G - Fatty-glandular

D - Dense-glandular

3rd column: Class of abnormality present:

CALC - Calcification

CIRC - Well-defined/circumscribed masses

SPIC - Spiculated masses

MISC - Other, ill-defined masses

ARCH - Architectural distortion

ASYM - Asymmetry

NORM - Normal

4th column: Severity of abnormality;

B - Benign

M - Malignant

5th, 6th columns: x, y image-coordinates of centre of abnormality.

7th column: Approximate radius (in pixels) of a circle enclosing the abnormality.

Example: how to read this image from MIAS database

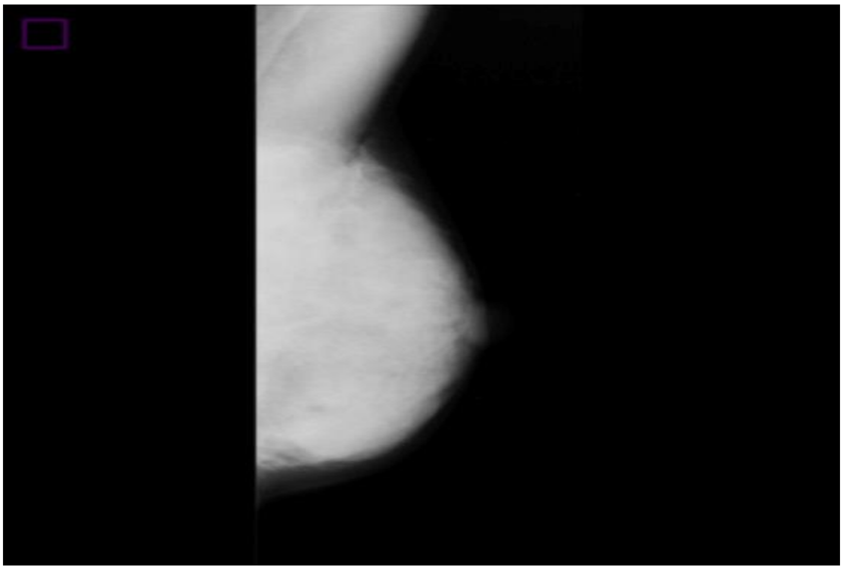


Figure 4-6: Mammogram image from MIAS database

This image name is “mdb058 D MISC M 318 359 27”

mdb058 is the number of the image in the database ‘58’

D means that the character of background tissue is Dense-glandular

MISC means that the class of abnormality present is ill-defined masses or other

M means that the severity of abnormality is malignant

318 359 are the image-coordinates of centre of abnormality.

27 is the approximate radius (in pixels) of a circle enclosing the abnormality.

4.4 Segmentation with snakes on MIAS database

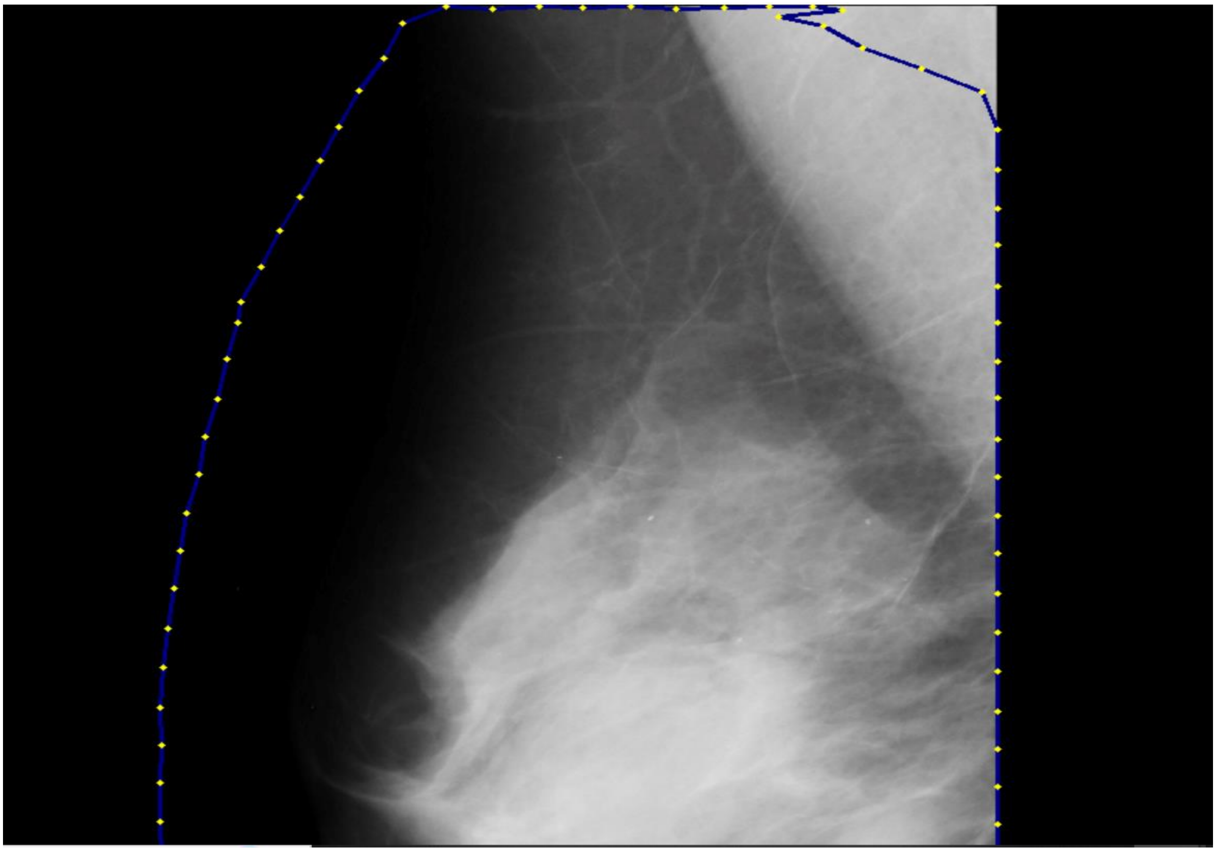


Figure 4-7: Segmented mammogram image from MIAS database with snakes.

As the picture shows, the snake is not really contouring the object(breast) due to this main cause:

The Snake (Kass method) lies in its reliance on well-defined edges in the image for accurate segmentation. While the snake is an edge-based segmentation technique, it heavily depends on clear and distinguishable boundaries between objects. In scenarios where the edges are not well defined, such as in images with low contrast, noise, or complex textures as it is in our case with MIAS database, the snake may struggle to accurately locate and track the desired object boundaries. This limitation can lead to suboptimal results, as the snake may deform incorrectly or fail to converge to the desired object contours. Therefore, the effectiveness of the Snake method is contingent upon the presence of clearly discernible edges in the image, posing a challenge when dealing with images that lack distinct boundaries.

Explanation:

The MIAS database consists of scanned mammographic images that are not in a digital format. This characteristic poses a challenge as the edges within these images may not be well defined.

Scanned images can suffer from various artifacts and imperfections, such as noise, blurriness, and inconsistencies in contrast. These factors can hinder the accurate detection and localization of object boundaries, making the segmentation task more challenging.

4.5 Segmentation with region based (chan-vese) on MIAS database

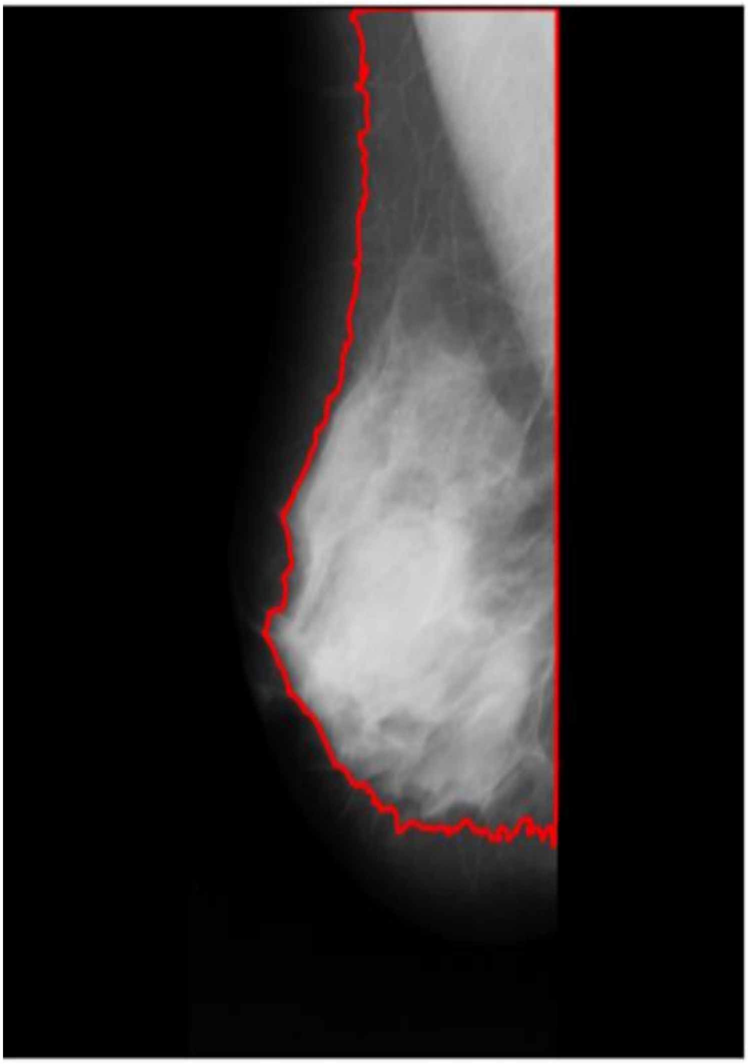


Figure 4-8: Segmented mammogram image (22) from MIAS database with chan-vese.

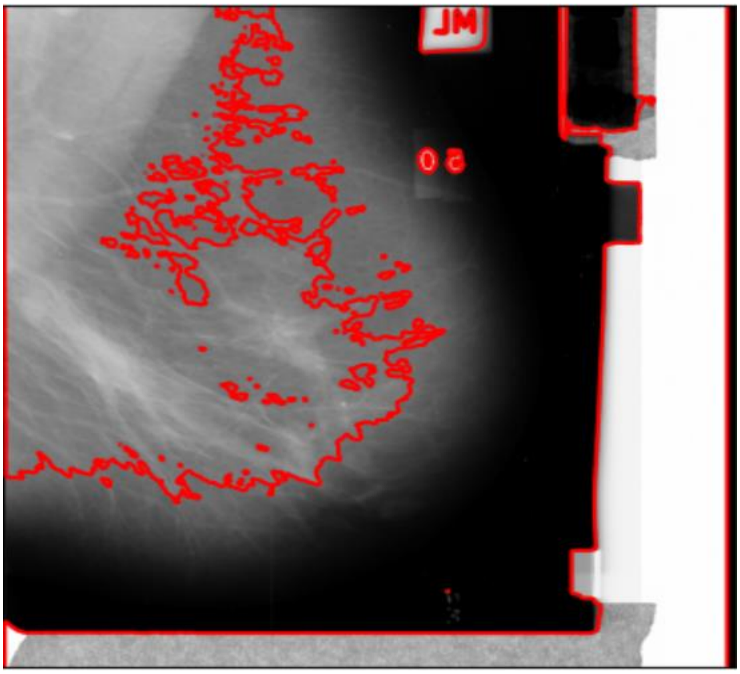


Figure 4-9: Segmented mammogram image (274) from MIAS database with chan-vese.

Explanation:

Unlike edge-based methods like the Snake, the Chan-Vese method performs region-based segmentation. It focuses on identifying coherent regions within an image rather than relying solely on edge information. This makes it particularly effective for segmenting objects with indistinct or weak boundaries. The Chan-Vese method incorporates a regularization term that helps to mitigate the influence of noise on the segmentation process. By considering the overall region properties, it can effectively handle images with high levels of noise without compromising the segmentation accuracy.

4.6 Conclusion

We conducted an experiment where we applied the Snake (Kass method) and the Chan-Vese model to the MIAS database images. Our objective was to compare the performance of these two segmentation techniques. Upon analysis, we found that the Chan-Vese method, being a region-based segmentation approach, yielded superior results compared to the Snake method. The Chan-Vese model showcased better adaptability to varying object appearances and handled images with indistinct or weak boundaries more effectively. However, we also examined the functionality of the Kass method (Snakes) and recognized its strengths in edge-based segmentation. Further exploration of the Kass method provided insights into its advantages and limitations, offering valuable perspectives for future studies and applications.

General conclusion

In this study, we look into the effectiveness of deformable methods, specifically the Chan-Vese and Snakes (Kass method), for breast cancer image segmentation using the MIAS database. The segmentation of breast cancer regions is a crucial step in diagnosing and treating this life-threatening disease. By comparing the performance of these two deformable models, we aimed to provide insights into their applicability and efficacy in accurately delineating cancerous regions within mammographic images.

Our experimental findings demonstrate that the Chan-Vese method, a region-based deformable model, outperforms the Snakes method in terms of accuracy, robustness, and computational efficiency. The Chan-Vese method utilizes the global optimization framework, considering both the region properties and contour smoothness, to achieve more consistent and reliable segmentation results. It effectively handles images with indistinct boundaries and varying object appearances, making it suitable for breast cancer segmentation tasks.

While the Snakes method, an edge-based deformable model, has its advantages in certain scenarios, such as well-defined edges, it proved to be less effective in accurately segmenting breast cancer regions within the MIAS database. The sensitivity of the Snakes method to noise and its dependence on well-defined edges limited its performance and hindered its adaptability to the diverse range of breast tissue characteristics.

Our research contributes to the field of breast cancer image analysis by highlighting the significance of deformable methods, particularly the Chan-Vese method, in achieving accurate and reliable segmentation of breast cancer regions. The results suggest that region-based approaches offer superior performance in handling the complexities and variabilities inherent in mammographic images.

Moving forward, further advancements in deformable models and segmentation techniques can build upon the findings of this study. Exploring hybrid approaches that combine the strengths of region-based and edge-based methods may yield even better results. Additionally, incorporating machine learning techniques, such as deep learning, could enhance the segmentation accuracy and further improve the diagnosis and treatment planning of breast cancer.

In conclusion, our research demonstrates that deformable methods, specifically the Chan-Vese model, hold significant potential for accurate breast cancer image segmentation. The ability to precisely delineate cancerous regions within mammographic images aids in early diagnosis, treatment planning, and monitoring of the disease. By improving the segmentation process, we can contribute to more effective and personalized care for breast cancer patients, ultimately leading to improved outcomes and enhanced quality of life.

It still important to refine the results obtained from firstly segmented objects in particular eliminating the artifacts that the active model detected.

Bibliography

[1] Johns Hopkins Medicine

https://www.hopkinsmedicine.org/ visited on 03/05/2023

[2] American cancer society

https://www.cancer.org/ visited on 06/05/2023

[3] Cleveland Clinic

https://my.clevelandclinic.org/ visited on 20/05/2023

[4] National library of medicine

https://www.ncbi.nlm.nih.gov/pmc/ visited on 27/05/2023

[5] Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., ... & Zhu, H. P. (2017). Risk factors and preventions of breast cancer. International journal of biological sciences, 13(11), 1387.

[6] NHS

https://www.nhs.uk/ visited on 01/06/2023

[7] NHS

https://www.nhs.uk/conditions/breast-cancer/diagnosis/ visited on 04/06/2023

[8]UpToDate

https://www.uptodate.com/contents/search visited on 07/06/2023

[9] Healthline

https://www.healthline.com/ visited on 10/06/2023

[10] National Breast cancer foundation inc

https://www.nationalbreastcancer.org/ visited on 16/06/2023

[11] Breast cancer network Australia

https://www.bcna.org.au/ visited on 19/06/2023

[12] Sharma, G. N., Dave, R., Sanadya, J., Sharma, P., & Sharma, K. (2010). Various types and management of breast cancer: an overview. Journal of advanced pharmaceutical technology & research, 1(2), 109.

[13] Breast cancer.org

https://www.breastcancer.org/ 20/06/2023

[14] American cancer society

https://www.cancer.org/ 22/06/2023

[15] "Digital Image Processing" book, by Gonzalez and Woods

[16] Research gate

https://www.researchgate.net/ visited on 24/06/2023

[17] Verma, R., & Ali, J. (2013). A comparative study of various types of image noise and efficient noise removal techniques. International Journal of advanced research in computer science and software engineering, 3(10).

[18] photo toolbox

https://www.photo-toolbox.com/index.html visited on 24/06/2023

[19] Bagade, S. S., & Shandilya, V. K. (2011). Use of histogram equalization in image processing for image enhancement. International Journal of Software Engineering Research & Practices, 1(2), 6-10.

[20] Cheng, H. D., & Sun, Y. (2000). A hierarchical approach to color image segmentation using homogeneity. IEEE Transactions on image processing, 9(12), 2071-2082.

[21] National cancer insitute

https://www.cancer.gov/types/breast/mammograms-fact-sheet visited on 25/06/2023

[22] American college of Radiology

https://www.acr.org/ visited on 25/06/2023

[23] https://www.cancer.org/

[24] Justaniah, E., Alhothali, A., & Aldabbagh, G. (2021). Mammogram Segmentation Techniques: A Review. International Journal of Advanced Computer Science and Applications, 12(5).

[25] American cancer society

https://www.ncbi.nlm.nih.gov/ visited on 25/06/2023

[26] Padhi, S., Rup, S., Saxena, S., & Mohanty, F. (2019, September). Mammogram segmentation methods: a brief review. In 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) (pp. 218-223). IEEE.

[27] Kass, Michael, Andrew Witkin, and Demetri Terzopoulos. "Snakes: Active contour models." International journal of computer vision 1, no. 4 (1988): 321-331.

[28] Chan, T. F., & Vese, L. A. (2001). Active contours without edges. IEEE Transactions on Image Processing, 10(2), 266-277.

[29] python

https://www.python.org/ visited on 25/06/2023