FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

A Deep Segmentation Approach for the Retrieval of Breast Cancer Treatment Outcomes

Carlos Miguel de Sá Veloso

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Master in Data Science and Engineering

Supervisor: Hélder Filipe Pinto de Oliveira, PhD

Co-supervisor: Mohammad Hossein Zolfagharnasab, MSc

Co-supervisor: Nuno Pinheiro da Silva Mota Freitas, MSc

Co-supervisor: Tiago Filipe Sousa Gonçalves, MSc

July 26, 2024

A Deep Segmentation Approach for the Retrieval of Breast Cancer Treatment Outcomes

Carlos Miguel de Sá Veloso

Master in Data Science and Engineering

July 26, 2024

Abstract

Breast cancer is one of the most common types of cancer among women globally, with significant impacts on health and quality of life. Developments in the treatment of this disease have led to better survival rates, however, the heterogeneity of clinical manifestations and therapeutic responses requires continuous advances in the personalization of treatment. Currently, breast cancer patients can maintain a normal quality of life due to advancements in treatments and increased awareness for population screening. This thesis investigates the potential of using Deep Learning (DL) techniques to evaluate post-treatment aesthetic results through medical image feedback.

The article starts with a discussion of breast cancer pathology, covering epidemiology, underlying biological mechanisms, and current treatment modalities. Next, it analyzes imaging feedback strategies as tools for improving therapeutic choice and monitoring disease progression. Convolutional Neural Networks (CNN) have revolutionized medical image retrieval, allowing for the identification and retrieval of breast cancer cases with similar visual and clinical characteristics from large databases.

The methodology initially focuses on the exploitation of image segmentation, with the aim of detecting the region of interest to patients (the breast region) and creating masks with a certain level of precision. This enables advanced and pre-trained deep learning models to extract the characteristics of the image exclusively from that area. In a second phase, two approaches were taken. The first employed tabular data (biometrics) from the patients, while the second used data extracted from the segmented images. The objective was to ascertain which of the two approaches yielded the most optimal outcomes and whether the utilisation of images cropped to the region of interest could facilitate more efficacious image retrieval outcomes than those observed with the original images. The Euclidean distance was employed to assess the degree of similarity and dissimilarity between patients and the model, with image segmentation demonstrating the most favourable outcomes. Finally, to demonstrate the efficacy of our experiment, image segmentation was incorporated into the pipeline of a framework, currently under review, which employs models trained and evaluated using triplet loss. Once more, the segmented images demonstrated superior performance to those without the segmentation step.

Acknowledgments

I would like to express my sincerest gratitude to all those who have contributed, directly or indirectly, to the completion of this thesis. Firstly, I would like to thank Professor Hélder Oliveira, my dissertation supervisor, for setting essential challenges for the project and for helping to solve other challenges that arose.

To Tiago Gonçalves for his invaluable guidance, patience, and support throughout this project. His advice, encouragement, and brainstorming sessions were instrumental in the completion of this work.

I would also like to thank the rest of the supervision team, Nuno Freitas and Mohammad Zolfagharnasab, for their availability, valuable suggestions, and constructive criticism, which greatly contributed to the improvement of this project.

To the FEUP Force team study group, not only for their support during the dissertation, but also throughout the master's degree. Their contributions were invaluable in bringing this master's degree to a successful conclusion.

I would also like to thank all my colleagues on the master's program in Data Science and Engineering, with whom I had the pleasure of sharing the classroom.

I would also like to thank my parents, family and girlfriend for their love, patience and unconditional encouragement. Their emotional and moral support was fundamental during the challenging moments of this journey.

Finally, I would like to express my gratitude to Victor Piglialarme, an exemplary manager, who not only facilitated the integration of work and academic schedules, but also provided invaluable encouragement during pivotal moments.

Carlos Miguel de Sá Veloso

"The real problem is not whether machines think but whether men do"

B. F. Skinnerl

Contents

List of Figures

List of Tables

Abbreviations

ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
AnDCG	Adjusted Normalized Discounted Cumulative Gain
BCCT	Breast Cancer Conservative Treatment
BCE	Breast Compliance Evaluation
BRA	Breast Retraction Assessment
BoF	Bag of Features
CBIR	Content-Based Image Retrieval
CBMIR	Content-Based Medical Image Retrieval
CNN	Convolutional Neural Network
CT	Computed Tomographic
DBMS	Database Management System
DC	Dice coefficient
DCG	Discounted Cumulative Gain
DCIS	Ductal Carcinoma in situ
DL	Deep Learning
DNN	Deep Neural Network
FNA	Fine Needle Aspiration
IDC	Infiltrating Ductal Carcinoma
IDCG	Ideal Discounted Cumulative Gain
ILC	Infiltrating Lobular Carcinoma
LBC	Lower Breast Contour
LCIS	Lobular Carcinoma in situ
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NDCG	Normalized Discounted Cumulative Gain
PACS	Picture Archiving and Communication Systems
ReLU	Rectified Linear Units
ROI	Region of Interest
SVM	Support Vector Machine
UNR	Upward Nipple Retraction
ViT	Vision Transformer

Chapter 1

Introduction

1.1 Context

1.1.1 Breast Cancer Overview

The technological revolution in medical imaging is particularly crucial in the context of cancer diagnosis and treatment. According to World Health Organization[1], cancer is a large group of diseases that can start in almost any part of the body (organs and tissues) when abnormal cells spawn in an uncontrolled process and invade other parts of the body, spreading to other organs. One of the major causes of death is metastization, the last process of cancer[2].

There are several stages to breast cancer, each of which has different forms of treatment (they can be more or less invasive). The stages range from O to IV, and the higher the stage, the more advanced the disease and the more invasive the treatment will have to be (in this case, for example, it is common for chemotherapy to be used). According to the Portuguese National Health Service[3], the probability of a woman getting breast cancer in her lifetime is 1 in 8.

Science has evolved not only in the different techniques for detecting cancer, helping to enable its detection at an increasingly early stage, but also in intervention and treatment techniques, allowing for increased survival rates and breast conservation - survival rates have increased significantly, according to the Portuguese Cancer League (Liga Portuguesa Contra o Cancro) [3], the survival rate is over 74%. With the rise of all these indicators also comes a greater interest in a better quality of life after treatment, where importance is given to the conservation and aesthetics of the breast. Among the many approaches used, breast cancer conservative treatment (BCCT) can be highlighted, which, as the name suggests, aims to preserve the breast [6]. The BCCT is shown in the Figure 1.1 below [4].

[1] Accessed on 10/01/2024 - https://www.who.int/health-topics/cancer
[2] According with Breast Cancer Foundation - https://www.bcrf.org/blog/types-of-breast-cancer/
[3] Accessed in 29/11/2023 https://www.ligacontracancro.pt/cancro-da-mama/
[4] Image from https://www.stlukes-stl.com/health-content/health-ency-multimedia/1/000913.htm

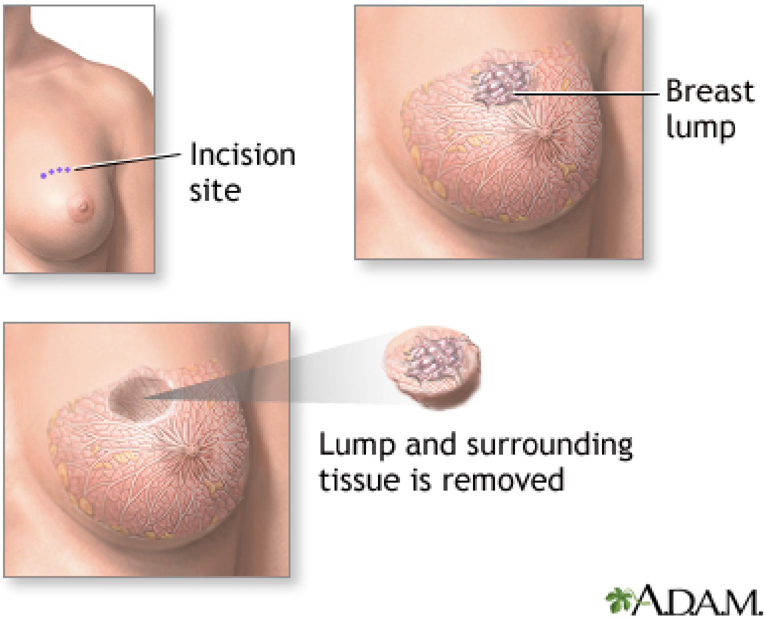


Figure 1.1: Female breast cancer conservative treatment.

1.1.2 Medical Image Retrieval

In more recent decades, the symbiotic relationship between science and technology has been evident and indispensable, with a union that has rocketed us towards unprecedented advances and many innovations. A clear example of this relationship can be seen in the health sector, particularly in the evolution of medical imaging and information technologies. The rapid increase in the use of picture archiving and communication systems (PACS) in hospitals and the medical community has led to a sharp rise in the number of medical images stored. Such an increase in information available has a strong impact on the efficiency and effectiveness of the retrieval process, as it allows for more material to be compared for each type of patient or disease [7]. Retrieving these images can be done through text-based or content-based methods [8]. However, the former faces more challenges than the latter, particularly in terms of data scalability [8]. Several studies have demonstrated the applicability of image retrieval in medicine for both diagnostic and treatment purposes [1, 9, 10]. These approaches have kept pace with the evolution of computational techniques. AI concepts have already been introduced to support image retrieval [10].

1.2 Motivation

The establishment of a standard for results, using artificial intelligence (AI) and advanced deep learning models, could lead to significant improvements in the management of patient expectations before surgery, whereby patients with very similar characteristics can be compared with others. The similarity of the patient's characteristics is crucial to the success of the retrieval: a patient with a specific breast size, skin color and age, for instance, should be compared with other patients with the matching or equivalent characteristics.

This disappointment with the results is due, among other reasons, to poor management of expectations with the patient, combined with the fact that there is no objective standard. Currently, the most common method of evaluating aesthetic results is through one or more observers (i.e., doctors) using photographs or directly with the patient. In addition to the potential reduction in the percentage of dissatisfied patients after surgery, this method increases the fluidity of the process, allowing the doctors or professionals involved to spend less time on these tasks [1]. The studies conducted by [11] and [3] were significant contributions to the field of breast cancer research. Nevertheless, it is important to highlight two softwares, the Breast Analysing Tool - BAT© and the Breast Cancer Conservative Treatment cosmetic results (BCCT.core), the first tool measures the difference between the size of the left and right breasts using images, and the second tool, known as BCCT.core, evaluates key components such as breast symmetry, color differences, and scar visibility. BCCT.core introduces a semi-automated and standardized analysis based on digital images, providing an objective assessment of cosmetic results in breast cancer treatment; softwares such as this is capable of establishing the objective standards that have been sought after. However, these methods still do not provide the desired results and have some difficulties in distinguishing some classes, for example, in the case of another software, the Breast Symmetry Index (BSI), the classes good and fair are very well detected, but in the case of fair and poor we are no longer able to obtain satisfactory results [12], thus showing the need for progress in this area. As technology becomes increasingly prevalent in all sectors, it would make sense for it to become a presence in this sector as well, making this process more autonomous, faster and more accurate. It has already been proven that technology, namely AI in the field of image retrieval, can improve medical tasks [12], thus leading to a better relationship between professionals and patients. The objective of this work is to develop an image retrieval system that can identify and present patients with images of similar clinical cases. This will allow patients to view results from cases analogous to their own, which could help illustrate possible disease pathways and treatment options. Identification with similar cases can be a powerful tool for patients, providing them with a more tangible and individualized perspective on what they can expect during their diagnosis and treatment. This image retrieval strategy is designed to strengthen the doctor-patient relationship by providing visual communication that can be more impactful than abstract clinical descriptions.

1.3 Goals

This dissertation aims to explore the capabilities of deep image segmentation to enhance the retrieval of medical images, providing more accurate and objective results. The goal is to train a segmentation model to accurately detect the breast region, utilizing these detected regions in a retrieval pipeline to improve the retrieval (visually) of the most similar and dissimilar patient cases.

1.4 Sustainable Development Goals

The work presented in this dissertation offers healthcare professionals tools to assist them in making informed prognoses and communicating aesthetic and clinical expectations to patients. This project aligns with SDG 3, which aims to enhance access to advanced medical technologies. By making medical imaging techniques more accessible to health facilities with varying levels of resources, it democratises technology and improves the quality of health in less favoured regions. Furthermore, it can help to reduce health inequalities and can assist in reducing disparities in the analyses and results of medical images.

1.5 Outline

Besides Introduction, this thesis contains three more Chapters.

Chapter 2 does an extensive review of the breast cancer and the types of the breast cancer and its current possible treatments.

Chapter 3 reviews the literature and the state of the art for Image retrieval and aesthetic evaluation of BCCT.

Chapter 4 explores the the image segmentation, using an U-Net architecture to generate masks of the patients region of interest.

Chapter 5 focus on the image retrieval experimental work comparing two different approaches, the first using images and deep learning algorithms and the second using tabular data.

Chapter 6 reviews the studies conducted to date and discusses future work.

Chapter 2

Background

2.1 Introduction

2.1.1 Breast Cancer

Breast cancer is a pathology in which specific cells in the breast become abnormal and multiply uncontrollably, forming a tumor. Most prevalent in women, this disease can also affect men, with ductal cancer - beginning in the cells of the milk ducts - being the most common in both sexes. In men, lobular cancer is rare, due to the scarce presence of lobular tissue [13]. This disease has great biological and molecular heterogeneity, with types varying according to their degree of invasiveness. Genetic mutations, especially in the BRCA1 or BRCA2 genes, substantially elevate the risk of developing this pathology. There are similarities between normal breast development and cancer progression, which suggests that breast cancer may originate from mammary cancer stem cells. Moreover, heterogeneity is influenced by epigenetic regulation and non-coding RNAs [14].

In the United States, breast cancer is the most commonly diagnosed cancer in women, excluding skin cancers, and the second leading cause of cancer death after lung cancer. Since 2012, there has been a significant increase in incidence and mortality, especially in less economically developed regions. In these regions, the main causes of high mortality are mainly lack of screening and late diagnosis of cancer [15]. Within Europe, the scenario was similar to the US, with projections showing around 380,000 new cases, of which 99% affect women. This accounts for approximately 13.8% of all cancer diagnoses[1].

Treatment strategies for breast cancer can vary depending on the molecular subtype and can include surgery, radiotherapy, chemotherapy, endocrine therapy for hormone receptor-positive disease, anti-HER2 therapy for HER2-positive disease and, more recently, immunotherapy. There is also an impact on the type of treatment depending on the stage of the disease, which is usually more invasive in the more advanced stages [16].

[1] https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/cancer-cases-and-deaths-rise-eu-2023-10-02_en

2.1.2 Types of Breast Cancer

It is in fact a common misconception that breast cancer is often mistaken for a monolithic disease rather than a complex constellation of conditions. Each type of breast cancer is distinguished by its molecular characteristics, prognosis and therapeutic responses, as well as different origins in the breast tissue [17] . The multifaceted nature of breast cancer challenges the medical community in the development of new treatment techniques and in the need to understand each of the various biological underpinnings [18]. In 2001, Dr. Charles Perou [2], classified for the first time, the breast cancer into subtypes based on gemonic patterns-discoveries that transformed our understanding of the disease and how it's treated.

There are several types of breast cancer [19, 20] and it is possible to divide them into two main groups according to cell type: **non-invasive breast cancer cells** and **invasive breast cancer cells**. Invasive breast cancer cells are characterized by the fact that these cells are not confined to the ducts and do not invade the surrounding adipose and connective tissues of the breast, whereas invasive cells tend to break through the duct and lobular wall and invade these same surrounding adipose and connective tissues of the breast. Breast cancer can be categorized into two main groups: Frequently occurring breast cancer and less commonly occurring breast cancer. Within these subgroups, there are several specific types of cancer, each with unique characteristics in terms of development, progression and response to treatment [19, 20].

2.1.2.1 Frequently occurring breast cancer

- **Ductal carcinoma in situ (DCIS)** - starts in the cells lining the milk ducts of the breast;
- **Lobular carcinoma in situ (LCIS)** - a much rarer non-invasive breast cancer, starts in the cells lining the breast lobules, the glands that make milk. Unlike DCIS, LCIS is not thought to generally progress to invasive breast cancer, but instead they are believed to increase a person's risk of breast cancer in the future. Some refer to LCIS as lobular neoplasia instead, as neoplasia refers generally to an abnormal growth of cells that can be benign or malignant;
- **Infiltrating lobular carcinoma (ILC)** - also known as invasive lobular carcinoma. ILC begins in the milk glands (lobules) of the breast, but often spreads (metastatizes) to other regions of the body. ILC accounts for 10% to 15% of breast cancers. This type of cancer is more common among postmenopausal women and have an increased risk of bilateral incidence. They are almost always estrogen and progesterone receptor positive;
- **Infiltrating ductal carcinoma (IDC)**: IDC begins in the milk ducts of the breast and penetrates the wall of the duct, invading the fatty tissue of the breast and possibly other regions of the body. IDC is the most common type of breast cancer, accounting for 80% of breast cancer diagnoses.

[2]Accessed on 10/01/2024 https://www.bcrf.org/blog/types-of-breast-cancer/

2.1.2.2 Less commonly occurring breast cancer

- **Medullary carcinoma** - It represents a form of invasive breast cancer characterized by a clear demarcation between the tumor tissue and the healthy tissue. Accounting for about 5% of all breast cancer cases, this type of carcinoma tends to be more prevalent among younger women, particularly those with a mutation in the BRCA1 gene;

- **Mutinous carcinoma** - Also known as colloid carcinoma, is an uncommon variant of breast cancer composed of cells that produce mucous. This type of cancer typically carries a more favorable prognosis compared to the more widespread forms of invasive breast carcinoma;

- **Tubular carcinoma** - Tubular carcinomas are a special type of infiltrating (invasive) breast carcinoma. Women with tubular carcinoma generally have a better prognosis than women with more common types of invasive carcinoma. Tubular carcinomas account for around 2% of breast cancer diagnoses;

- **Inflammatory breast cancer** - The manifestation of inflamed mammary tissue, characterized by redness and warmth, accompanied by dimpling or pronounced ridges, can be attributed to the obstruction of lymphatic vessels or channels within the skin overlaying the breast by cancerous cells. Despite its rarity, comprising merely 1% of all breast cancer cases, inflammatory breast cancer is known for its aggressive progression;

- **Paget's disease of the nipple** - A rare form of breast cancer that begins in the milk ducts and spreads to the skin of the nipple and areola, Paget's disease of the nipple only accounts for about 1% of breast cancers;

- **Phylloides tumors** - (also spelled "phyllodes") are can be either benign (non-cancerous) or malignant (cancerous). Phylloides tumors develop in the connective tissues of the breast and may be treated by surgical removal. Phylloides tumors are very rare; less than 10 women die of this type of breast cancer each year in the United States.

2.1.3 Breast Cancer Diagnosis

As mentioned earlier, the importance of early diagnosis has a major impact on survival rates in breast cancer patients [21]. Literature shows that when there is an early diagnosis, mortality rates are significantly reduced in the long term [22]. Breast cancer diagnosis is a multifaceted process that involves a combination of techniques to determine the presence and extent of the disease. This process can include, but is not limited to, the following methods: [19, 22]

- Breast Self-Examination and Clinical Examination;

- Mammography;

- Breast Ultrasound;

- Magnetic Resonance Imaging (MRI);

- Biopsy.

2.1.3.1 Breast Self-Examination and Clinical Examination

Doctors and various organizations involved in the study of breast cancer or patient support are increasingly urging women to have their breasts checked regularly, and this awareness has been growing in recent years. One of the most well-known symptoms of this disease is the presence of a lump, often identifiable in the breast tissue or in the axillary region. A monthly breast self-examination is one of the medical community's recommendations that allows women to familiarize themselves with the usual texture of their breasts, the cyclical changes, size and condition of the skin, thus increasing the likelihood of early detection of any abnormalities. There are a number of general warning signs of breast cancer which can be manifested by swelling or lumps in the breast, swelling in the axillary area, nipple discharge (clear or blood-tinged), pain or hypersensitivity in the nipple, or even a change in the appearance of the nipple itself. These signals do not appear in all cases, yet when they do, they are warning signs that require immediate medical assessment for a more appropriate diagnosis. Medical assessment by touch is also common practice, where the doctor will look for the same signs as described above, but mainly for the presence of a lump. In the metastatic stage, the most advanced stage of the disease, other symptoms may appear due to the possible involvement of axiliary lymph nodes. Generally, the patient begins to feel some bone pain, shortness of breath, weight loss (hepatocyte metastases) and other neurological symptoms such as headaches and weakness, which may indicate brain or nervous system metastases [19, 23, 24].

2.1.3.2 Mammography

Mammography is another essential breast cancer screening procedure in which a specialized x-ray machine is used to visualize breast tissue. The scanner emits electromagnetic radiation (at a much lower dose than standard x-rays to ensure patient safety) under the breast, which is placed on a flat, compressed surface to disperse the breast tissue and obtain a clearer image. The images are taken at different angles so that there is a better view of the whole breast [24]. Digital mammography is a very similar process, in which radiologists use images of the breast, but with the advantage that, as it is stored digitally, the brightness, contrast and other values that can help detect any abnormalities can be altered [24]. The studies presented by [25] regarding the mammography with contrast for the diagnosis of breast cancer highlighted promising results over traditional digital or conventional methods particularly beneficial for evaluating patients who present with ambiguous results from conventional mammography.

2.1.3.3 Breast Ultrasound

Breast ultrasound, an effective tool for differentiating between solid and cystic breast masses, is particularly useful in dense tissue and in situations where mammography is inconclusive. Primarily indicated for patients at high risk of breast cancer or for those who cannot be submitted to mammography, ultrasound enhances the sensitivity of screening but can decrease specificity and increase the rate of unnecessary biopsies. Its use as complementary screening is valuable, despite possible risks such as false positives, and can be optimized when targeted at women who are more likely to fail mammography screening [22, 26].

2.1.3.4 Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging (MRI) is outstanding for creating conditions to identify small lesions that may remain hidden on mammography. Nonetheless, this approach has drawbacks, including a high cost and low specificity, which can lead to over-diagnosis. The MRI technology generates images in distinct slices using a strong magnetic field and radio-frequency signals, with the option of adding a contrast agent to improve image resolution. Although it is recommended for individuals at a high risk of breast cancer, it is not generally recommended for the general population due to its proclivity to produce false-positive results, as well as the associated costs, time consumption, the need for an adequate number of units and experienced radiologists, and the lack of broad clinical utility [22, 24].

2.1.3.5 Biopsy

Three major breast biopsy procedures have been adjusted to various diagnostic scenarios. Fine needle aspiration (FNA) biopsy collects cytological samples from solid masses using a 20-gauge needle, which is sometimes guided by ultrasound or stereotactic technology for non-palpable tumors. Core biopsy involves removing cylinders of tissue with a 14-gauge needle, either with or without imaging guidance, and requires a minor skin incision and local anesthetic. Excisional biopsy is recommended when needle biopsies are unreliable yet clinical suspicion of malignancy persists, and it may be the first choice if the likelihood of malignancy is significant. In non-palpable masses, wire localization is employed before biopsy. These treatments are typically performed with local anesthesia and intravenous sedation, allowing the patient to go home the same day.

2.1.3.6 Breast Cancer Treatment

Women diagnosed with breast cancer have a diverse number of treatment options available to them, including surgery, chemotherapy, radiotherapy, hormone therapy and specific treatments such as targeted therapies. The choice of therapeutic approach is often influenced by the stage of the disease. However, the final word rests with the patient and this process involves a collaborative evaluation between doctor and patient[3]. Women diagnosed with breast cancer have a diverse number of treatment options available to them, including surgery, chemotherapy, radiotherapy, hormone therapy and specific treatments such as targeted therapies. The final verdict on the therapeutic approach is often affected according to the stage of the disease. However, the final word rests with the patient and this process involves a collaborative evaluation between doctor and patient. An increasing trend has been observed among young women diagnosed with early-stage

[3] http://www.ligacontracancro.pt/cancro-da-mama/

and unilateral breast cancer to opt for surgery, whereas in the past surgery was something to be avoided, this is also due to new surgical techniques and the successes they have achieved. Depending on the type of treatment, in less advanced cases such as ductal carcinoma in situ, which can evolve into invasive cancer, treatment is usually through conservative surgery together with radiotherapy, without the need for systemic therapy. In stages I and II, conservative surgery followed by radiotherapy is also common, which reduces mortality and recurrence. In this case, it may be followed by a sentinel lymph node biopsy, depending on the type of cancer. Cancers with positive lymph nodes are treated systemically with chemotherapy, endocrine therapy and trastuzumab. At a progressed stage such as stage III, chemotherapy is advised to reduce the size of the tumor before conservative surgery. Nevertheless, if this cancer is of the inflammatory type, which is considered more aggressive, masectomy is recommended instead of conservative surgery. Stage IV is the most advanced stage. Once here, the main concern is no longer a cure but trying to slow down the progression of the disease, hormone therapy is used, if there are positive hormone receptors, chemotherapy or both. Immunotherapy may also be used [27, 28, 29].

2.2 Conclusions

Breast cancer patients now have the perspective of a longer and better quality life, as a result of advances in the medical field [21]. The importance of prevention has been emphasized and disseminated widely, partly due to advances in the diagnostic equipment and techniques available [19, 22]. Prevention can be started in the comfort of the home, through self-examination, a recommended practice for early detection. The prognosis and treatment of breast cancer vary according to the stage at which the disease is discovered; earlier diagnosis tends to result in better outcomes and often allows for less invasive therapeutic approaches [16, 21].

Chapter 3

Literature Review

3.1 Introduction

The increase in breast cancer survival rates, which has risen in recent years[1], is largely due to a series of awareness-raising programs by the many entities linked to the study and prevention of the disease (including governments and hospitals), which have led to a greater concern for prevention and regular check-ups, resulting in more and more cases being detected in the early stages of the disease, rather than in the more terminal and complicated stages. Combined with the development of surgical treatments and the promising advances in radiotherapy [30], this has resulted in an increase in survival rates [31]. Patients have therefore become more concerned about their quality of life post-treatment, with aesthetic results being one of the main areas of focus [12]. In breast cancer, the consequences of treatment can be quite aggressive, not only in psychological terms, because of all the risk involved with the disease and the fact that, although the average life expectancy for these patients has increased, it remains a disease with a high mortality rate [32], but also due to the physical consequences, in cases where the cancer is more advanced, there may be recourse to partial or total mastectomy of the breast [29], which means that the patient may lose part or all of the breast. Aesthetic outcomes may not always be the most pleasing, as noted by [33]. Additionally, these outcomes can have a significant impact on patients' lives, potentially leading to a lack of trust and difficulties in intimate relationships, as reported by [12]. These issues are not limited to cases of total mastectomy, but also occur in cases of conservative cancer treatment. Nevertheless, with all the technology available, it is possible to detect breast cancer at the earliest stages, which means that the patient would not have to go through such an invasive process, but would have the option of a more conservative treatment. Here, the breast is partially or completely preserved, unlike the treatment described above [34].

[1] https://www.breastcancer.org/facts-statistics

3.2 Medical Image Retrieval

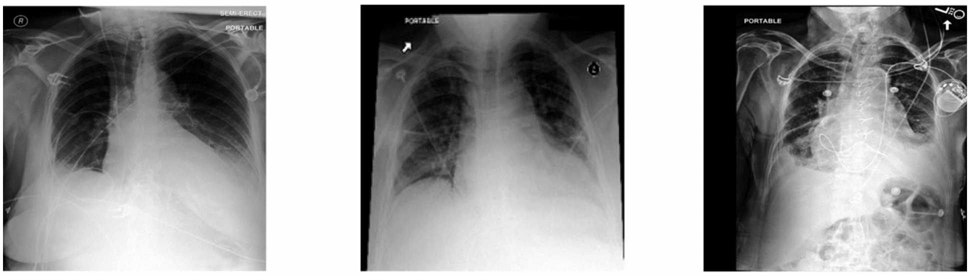
Imaging has been an important element of modern medicine, becoming a powerful tool for detecting and supporting the treatment of certain diseases [35]. With the increasing accumulation of medical data and images and the growing use of picture archiving and communication systems (PACS), it is becoming possible to take advantage of this collection of data that has been stored in recent years to support research [36]. Currently, these images can be retrieved in two ways: text-based methods and content-based methods [8]. Traditionally, the ability to search for and retrieve images was facilitated by text-based labeling systems. Each image in a database was manually annotated with textual descriptors such as keywords, tags or categories that described its content. These descriptors were then indexed by a Database Management System (DBMS), which could consult these annotations to find images matching the user's search terms. The manual annotation approach encounters some challenges such as [8, 37]:

- High labor intensity: The process requires a considerable investment of time and effort;
- Annotation inaccuracy: The subjectivity of annotators can lead to inconsistencies and inaccuracies in annotations;
- Semantic gap: There is a lack of alignment between the textual descriptors and the visual complexity of the images;
- Scalability issues: The exponential growth in the number of digital images challenges the updating and maintenance of annotations.

To overcome these problems, Content-Based Image Retrieval (CBIR) systems use visual characteristics of images [7, 38] to index and retrieve images from a database. These systems are designed to allow the search and retrieve of images similar to a query image provided by the user, based on visual characteristics rather than textual metadata or manual descriptions. The visual characteristics are among others the following [7, 38]:

- Color: Color is one of the most commonly used visual characteristics in CBIR. Image retrieval systems usually use color to represent and compare images. However, given that in medicine many images are used in the gray scale, it is not so decisive;
- Texture: Texture refers to the appearance or feel of a surface;
- Shape: The shape of objects in images can be a strong distinguishing feature.

Despite these advances in relation to the limitations of manual annotation, CBIR also encounters challenges to its success and feasibility, the main one being image similarity [39]. The difficulty with this similarity lies in extracting features that are similar for a given specific task, for example, the extraction of color is unlikely to be interesting for a task in which the images are gray-scale ultrasound images. Feature extractions should be targeted at each type of problem and when we have global features, as in the case of those described above, it becomes complicated to generalize the result for each individual task and to try to overcome the problem, the use of local features (at the level of the image pixels) can be a solution [39]. In medicine, when the objective is the detection or diagnosis of a disease, the choice of the most similar image differs from many other tasks. Here, the main focus of the concept of similarity is not the selection of the most similar image, in the general sense, but rather, in the objective sense of the disease [1] (see Figure 3.1).



(a) Test Image **(b)** Least Similar Image **(c)** Most Similar Image

Figure 3.1: Pleural Effusion test image and the least and most similar images of the catalogue [1].

Studies reported using the ASSERT CBIR system have already shown good diagnostic results with lung images [40], as has another study in which [9] used computed tomographic (CT) images for retrieval. The aim was to develop and preliminary test a system designed to facilitate the identification and retrieval of radiological images characterized by lesions of similar appearance.

3.2.1 Artificial Intelligence Approaches

Even more recently, Deep Learning (DL) models have shown good performance in CBIR applications, [10], based on the [41] model to propose a classification model using an enhanced version of Resnet to improve the accuracy and efficiency of the Content-Based Medical Image Retrieval (CBMIR) system. In this model, the main change proposed was to replace the final 7x7 average clustering layer with a 7x7x2048 convolutional layer. The reason for this replacement is that the classification of medical images presents the challenge of high similarity between different classes, which requires the extraction of more useful features by the CNN for effective classification. With this model, they showed results that surpassed those of the original ResNet50 model.

Silva et al[1], conducted an experiment in which they use four different methods for the Chest X-ray retrieval task focusing specifically on the application of structural similarity indices, convolutional neural networks, interpretability-guided networks and attention mechanisms to improve the accuracy and effectiveness of image retrieval. For this study, they use a CNN with the DenseNet121 architecture to automatically identify relevant features, but in this case, unlike previous approaches, the initial weights are derived from pre-trained models on specific medical conditions, the pleural effusion CheXpert CNN model by [42] instead of ImageNet. Similarity is calculated by the euclidean distance in feature space. The third method uses the same architecture as the CNN but uses saliency maps generated by the Deep Taylor method as input, focusing on features related to the disease. The network is trained with specific models for medical conditions, in this case the CheXpert IG model from [42], and similarity is assessed in the same way as the CNN. Finally, it introduces a method based on attention mechanisms to learn relevant features at different levels and scales. It uses a CNN with an attention mechanism (self-attention and multi-level attention) to classify medical images, focusing on more focused and relevant features. Similarity is computed in a manner similar to previous approaches. This work demonstrates the effectiveness of an interpretability-driven medical image retrieval method that uses attention mechanisms to improve medical decision support in radiology. This method outperformed other approaches tested and proved capable of capturing diagnosis-relevant information and effectively representing similar clinical cases, which is particularly valuable in difficult diagnostic scenarios. In addition, it was found to match the performance of board-certified radiologists in both conditions studied, with remarkable performance even in conditions where there was disagreement in diagnosis, such as pneumonia.

3.3 Retrieval of Breast Cancer Treatment Outcomes

3.3.1 Aesthetic outcome evaluation

Aesthetic outcome evaluation is a key element in the patient's recovery and one that has a major influence on their quality of life after treatment. At the moment, there is no standard for evaluating these results [12], which often leads to the patient's expectations not truly being managed in the most effective terms. As there is no standard for comparison, the patient's expectations may be higher than the outcomes, or even lower than the anticipated result, leading them to go through a less stimulating period, as described above. The characterization of the results will always depend on the complexity of the cancer, where more advanced stages require more in-depth interventions, compared to earlier stages where the intervention may be more conservative (BCCT). Naturally, the aesthetic results of two patients with different stages should not be compared, as it is possible that the more advanced case will have poorer results than the earlier one [3]. Currently, in the literature, we can find two types of evaluation: subjective and objective [12].

3.3.1.1 Subjective evaluation

The subjective method was first described in [43]. Harris introduced a system of subjective cosmetic assessment that has become a widely accepted standard (see Figure 3.2): excellent (the treated breast looks almost identical to the untreated breast), good (the treated breast is slightly different from the untreated breast), fair (the treated breast is clearly different from the untreated breast, but not severely distorted) and poor (the treated breast is severely distorted) [43]. Presently, subjective evaluation done by one or more observers is the most frequently employed assessment of the aesthetic result. This evaluation can be performed through direct observation of the patient or through photographic representations. However, as the professionals involved in the treatment are often part of the group of observers, impartiality may not be guaranteed. In fact, direct observation of the individual patient can prove to be quite uncomfortable. It should also be highlighted that, even among professionals, it is fairly common for there to be discrepancies regarding the appraisal result, calling into question the reliability of the method [12, 44]. This subjective approach, despite being broadly adopted, faces significant challenges when it comes to the accuracy and consistency of the results. The subjective nature of this assessment can lead to marked variations in perceptions of the aesthetic result, influenced by the experience, training and personal expectations of the observers. This not only compromises the neutrality of the process, but also questions its validity as a reliable measure of treatment success. The participation of professionals directly involved in the treatment as evaluators increases the risk of bias, as they may have specific expectations or preferences regarding the results, affecting their assessment. This lack of neutrality can distort the results, undermining the objectivity required for an accurate assessment of aesthetic results. In tandem, direct observation of the patient can add a layer of discomfort for both the patient and the observer, and can influence the perception of the results. Such a situation can be problematic in contexts where privacy and sensitivity are essential. The divergence of opinions among professionals about the evaluation of results highlights an urgent need for more objective and standardized criteria for aesthetic evaluation.Therefore, although subjective methods play a crucial role in the evaluation of aesthetic results, it is imperative to recognize and address their limitations. This motivated the development and implementation of more objective assessment tools, such as standardized scales and image analysis, to complement and potentially improve the accuracy and reliability of aesthetic assessments in clinical practice [12, 44].

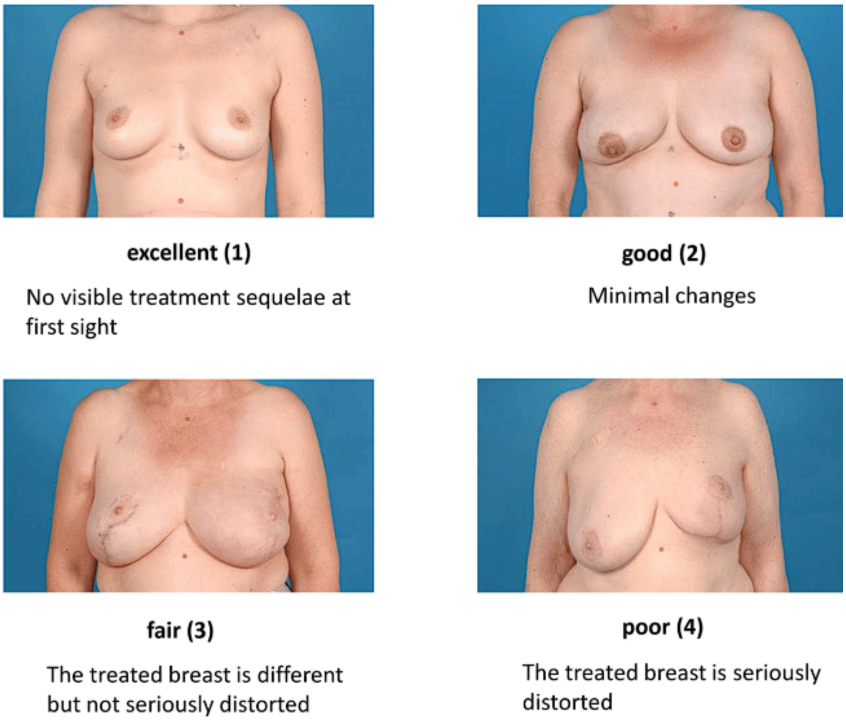


Figure 3.2: Harris subjective method, from [2].

3.3.1.2 Objective evaluation

New methods had to be introduced to address the limitations of subjective methods in evaluating breast cancer conservative treatment (BCCT). Pezner et al. [45] pioneered the objective method, which correlated the aesthetic result of surgery, the technique used, and the effects of surgery. They adopted the first objective measure to assess asymmetry, the breast retraction assessment (BRA), which directly correlates the aesthetic result of surgery with breast asymmetry, one of the main components of aesthetic assessment. BRA can be utilized to quantify physical changes in the treated breast by comparing the distances between fixed anatomical points on the skin of the breast and a reference point on the chest, before and after treatment. This method provides an accurate and reproducible assessment of cosmetic results, overcoming the limitations of subjective assessments that are prone to significant variations between different observers [45]. Following [45], several other authors have proposed their own models, some of them along the same lines, such as [46], who proposes a quantitative scoring system involving evaluation criteria such as breast symmetry, skin texture, appearance of the scar and some other presence in the breast that can be considered a change [46] or [47] who have proposed the breast compliance evaluation (BCE) method in which they quantitatively measured the compliance of the two breasts (the normal one and the one subjected to radiation). The method measures the difference between the anterior breast surface length from the infra-mammary fold to the nipple between the patient in erect and supine positions.Noguchi et al. [6] employed a Moiré topography camera to objectively assess breast asymmetry, providing a quantitative measure of this specific feature. In addition, more subjective elements such as breast tissue atrophy, skin modifications and the appearance of surgical scars were assessed by human observers. By combining these objective and subjective assessments, [6] developed a comprehensive approach to determining the overall aesthetic result, thereby establishing the foundations for a more integrated and objective assessment of post-surgical breast aesthetics. These initial generations of models brought objective quantification to evaluate individual components of the overall aesthetic result, moving beyond total dependence on expert judgment. Despite the fact that these innovations have not yet significantly differentiated the methodology of operative evaluation. The persistence in subjective evaluations for individual aspects keeps the final decision as subjective and irreproducible.

To address the challenges of the first models, innovative methods have been introduced for evaluating aesthetic results in breast-conserving surgeries. These techniques use software tools capable of automatically and objectively extracting individual characteristics from patient photographs in order to predict the overall aesthetic result. Such attempts have sought to compensate for the lack of effective tools by offering a computational solution that promises efficiency, simplicity, speed, reliability and reproducibility in the assessment of post-surgical care in breast cancer patients. To make this feasible, Fitzal et al. [11] set about creating a database of patient images, subjectively classified by experts, establishing a ground truth for the development and comparison of new methods. Fitzal et al. [11] therefore introduced the Breast Symmetry Index (BSI), using the Breast Analysing Tool (BAT©) software, to assess the size differences between the left and right breasts in patients' digital images. The sum of these differences makes up the BSI score, which quantifies variations in breast size. The BSI has been shown to be accurate, reproducible and can be used by experts as the learning curve is short and the results are consistent. The BSI score, obtained from frontal images, showed a high correlation with experts' subjective assessments and a significant distinction between good and reasonable cosmetic results. However, the BSI showed no correlation with the patients' self-assessment and was unable to distinguish between excellent and good results or between reasonable and poor results. These data indicate that the BSI, calculated using the BAT©, can be partially used in the aesthetic assessment after BCCT.

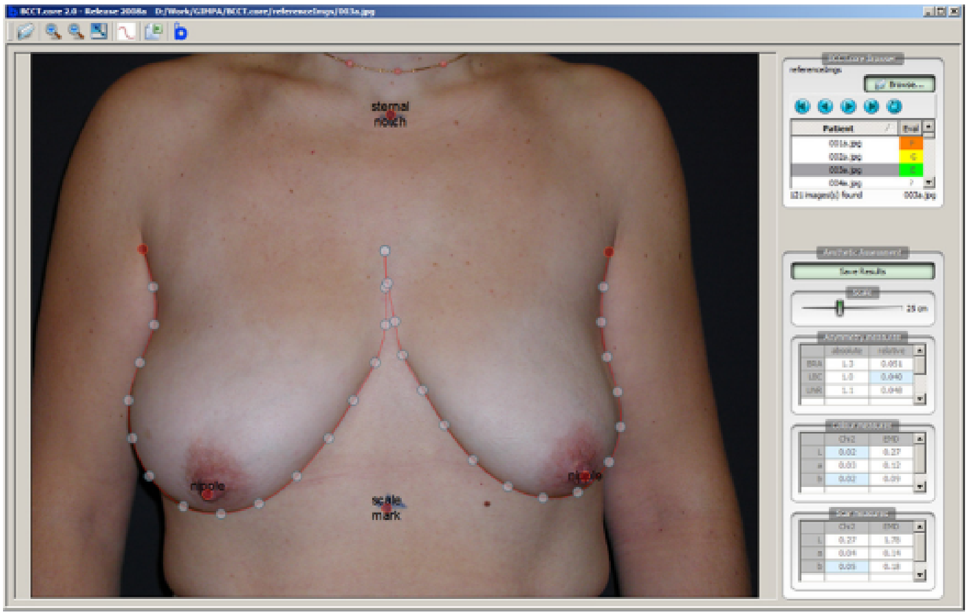


Figure 3.3: User interface of the BCCT.core software, from [3].

A very prominent method, introduced by [3], represents a significant advance in the objective assessment of cosmetic results in conservative breast cancer treatments. Known as Breast Cancer Conservative Treatment cosmetic results (BCCT.core), it introduces a semi-automated and standardized analysis based on digital images. BCCT.core (see Figure 3.3) uses a series of components that are considered key to assessing cosmetic results:

- Breast Symmetry: The software evaluates the concordance between the operated breast and the contralateral breast, where discrepancies can impact the patient's aesthetic perception and self-image;

- Color Differences: BCCT.core quantifies color variations between the breasts, which can signal changes in the skin resulting from treatment, such as the effects of radiotherapy;

- Scar Visibility: The tool analyzes how visible surgical scars are, which is a relevant factor in post-treatment aesthetic evaluation.

The accuracy of the cosmetic classification achieved by BCCT.core was around 70% compared to the subjective assessment of a panel of experts. BCCT.core, which includes additional parameters such as color differences and scar appearance, showed good performance with high quality photographs, but moderate performance with lower quality images.

3.4 Intelligent systems for BCCT aesthetic evaluation

In the field of aesthetic evaluation of breast cancer treatments, the method proposed by Cardoso et al. [3], which uses a support vector machine (SVM) for classification, represents the current benchmark. However, this method requires the semi-automatic annotation of key points, such as the contour of the breast and the position of the nipples, as well as the calculation of asymmetry features, thus requiring human intervention and preventing its incorporation into an end-to-end image generation model to predict aesthetic results. This method has inspired a number of studies and several authors who have used Cardoso et al. [3] work as a basis, such as Silva et al. [12].

3.4.1 Key point detection

3.4.1.1 Breast contour detection

Cardoso et al.[48] were among the pioneers in presenting an algorithm capable of automatically identifying breast contours in digital photographs. To achieve their goal, they analyze the image gradient and transform it into a weighted graph, taking into account the pixel gradient, value and position. On the assumption that the end points of the breast contour are known in advance, the central challenge of the algorithm is to find the shortest path that connects these two end points, passing through the breast contour. After developing the breast contour detection work, Cardoso et al [49] proposed a methodology to automatically identify the end points of the contour. It is based on the assumption that the photograph only captures the patient's torso, which allows us to assume that the external end point of the breast contour is located at the point where the contour of the arm meets the contour of the torso. However, in most of the photos, the patients have their arms down, making the contour of the arm almost indistinguishable from the contour of the torso. This leads to the definition that the external end point of the breast is the highest point of its contour.

3.4.1.2 Nipple location

In 2007, Udpa et al. [50] proposed an automatic method for nipple detection based on normalized cross-correlation. The method combines a bank of Gaussian and Laplace Gaussian filters with a probability map for likely nipple locations, thus reducing the number of false positives. The nipples are identified in the regions of the images where the pixels are most intense. In 2015, Cardoso et al [51]. proposed an alternative approach for identifying possible nipple locations and detecting aureole contours using a corner descriptor and a closed contour method. The best match between nipple candidates and closed contours is selected based on four high-level features, and a SVM classifier is trained with these features, resulting in comparable results. These advances in automatically detecting breast features in digital photographs are essential for objectively and accurately assessing aesthetic results in patients treated for breast cancer. This represents a significant contribution to the field of breast reconstruction and post-treatment aesthetic assessment.

3.4.2 Deep Learning approaches

Recently, Silva et al. [12] presented a more innovative approach, inspired by the ideas described by Cardoso and Cardoso [49], which uses a highly regularized DNN to automatically evaluate aesthetic results. High regularization was essential due to the small size of the data set and the subjective nature of the aesthetic results. After experiments with conventional CNN networks, such as DenseNet-121 and ResNet50, the need to significantly reduce the number of parameters to combat overfitting was identified. This resulted in the development of a simpler DNN, following the "conv-conv-pooling" scheme. Luini et al. [30] in order to improve accuracy, introduces an intermediate supervision in the detection of important key points and integrates predefined functions to translate these points into specific asymmetry measures (LBC (difference between the lower breast contours), BCE (difference between inframammary fold distances), UNR (difference between nipple levels), and BRA (breast retraction assessment)) using ”lambda" layers. Training takes place in two main phases: the first phase focuses on the detection of key points, using the mean square error of the coordinates of the key points. The second phase consists of multitask training that simultaneous optimizes keypoint detection and classification performance, with different weights for keypoint detection and classification losses. This method was a breakthrough in technology as it was the first to use deep learning to assess the aesthetic outcome of breast cancer treatments. In addition, the results presented were superior to those presented in the state of the art.

The methods in the literature use traditional computer vision algorithms, which can be too complex and time-consuming to compute. To get around these problems, Gonçalves et al. [44] tried to develop a new algorithm for detecting key points that is efficient but at the same time manages to maintain or increase accuracy (see Figure 3.4).

Gonçalves et al. [44], applied image segmentation into the pipeline of esthetic evaluation of breast cancer surgery outcomes. Image segmentation is the process of dividing an image into specific parts, or regions, so that each region contains pixels with similar attributes. This process is essential for identifying and isolating regions of interest in images. An innovative approach has been proposed to detect breast contours in digital images by combining image segmentation techniques with the keypoint detection model developed by [34]. in a structured pipeline. First, a deep image segmentation model had to be trained, and the U-Net++ architecture was chosen due to its good previous results in similar tasks. Due to the small size of the dataset (about 200 images), data augmentation techniques such as rotations, reflections, and masks were used to train the algorithm. In the next step, the endpoints predicted by Silva et al. keypoint detection algorithm were projected onto the contours of the masks, minimizing the euclidean distance between each keypoint on the mask contour and the predicted keypoint. The end result was a new set of 34 breast contour keypoints, in addition to the nipples and sternum notch predicted by the Silva et al.

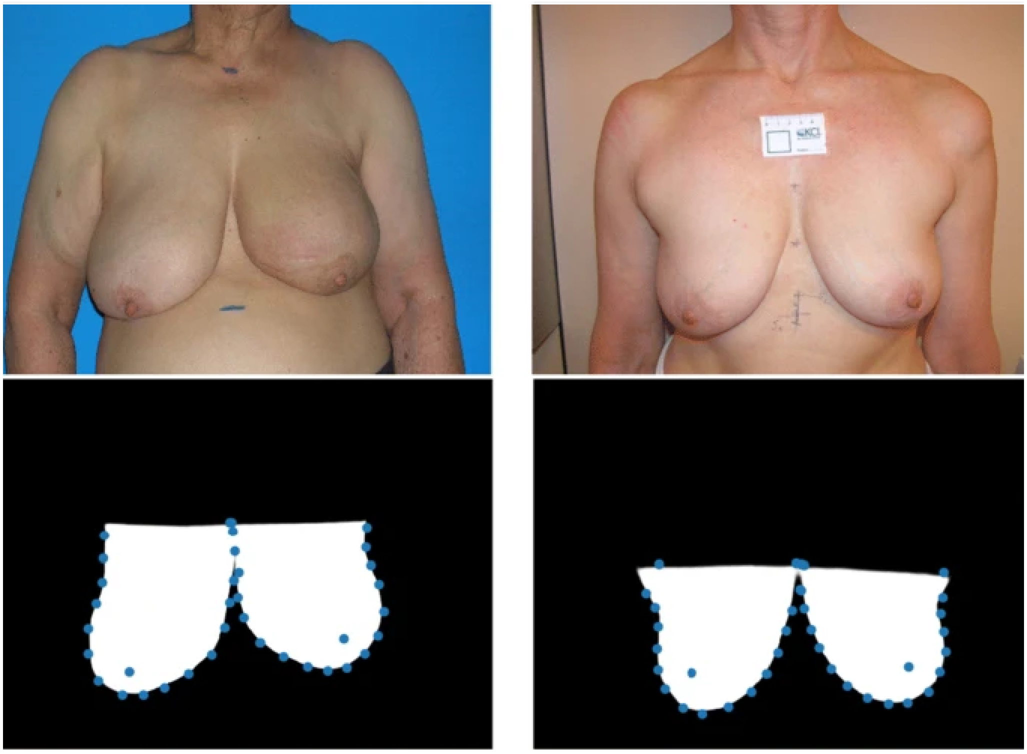


Figure 3.4: Breast image segmentation and keypoints detection
, from [44].

algorithm. This approach represents a significant advance in breast contour detection, successfully integrating deep image segmentation and keypoint detection.

Freitas et al. [4] introduced a hybrid approach that represents a significant advancement in the methodology of breast contour detection. The approach consists of two main modules: Deep Edge Detection and Shortest-path Estimation. It is derived from the conventional method developed by Cardoso and Cardoso [49], with a major modification, it replaces the Sobel Edge operator with DL models that can be optimized for the task at hand (see Figure 3.5).

The Deep Edge Detection module is crucial in identifying edges between skin regions and enhancing the edge detection process. Two models were explored for this purpose:

1. RINDNet, a state-of-the-art, complex, and computationally intensive model designed to tackle intricate problems by learning and refining edge detection capabilities. It has over 59 million parameters, making it well-suited for applications that require high precision in edge detection. With over 59 million parameters, RINDNet is well-suited for applications requiring high precision in edge detection.

2. SobelU-Net is a simpler and more lightweight alternative that integrates Sobel filters within a U-Net architecture. This model leverages the advantages of conventional edge operators while incorporating the learning capabilities of deep networks.

The hybrid model, particularly the SobelU-Net, demonstrated superior performance across various datasets, showcasing robust generalization capabilities not seen in fully automatic deep models.

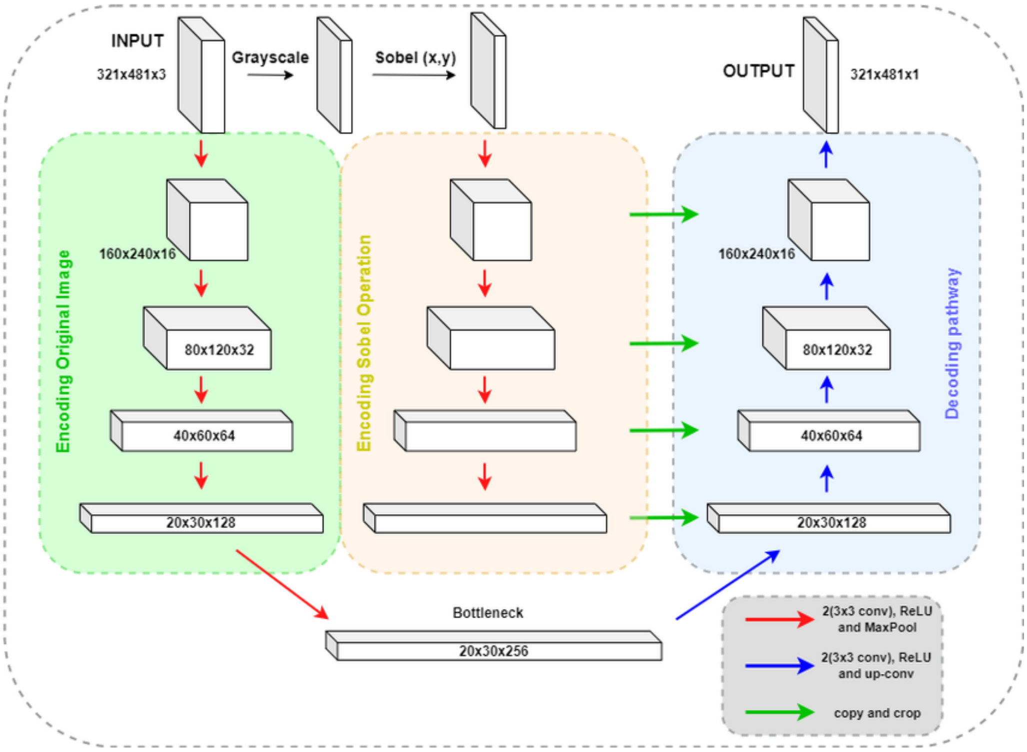


Figure 3.5: SobelU-Net architecture [4].

The approach excels in adapting to new datasets with minimal adjustment, thanks to its innovative integration of edge filtering operations. The main contribution of this study is its hybrid framework, which combines conventional methods' reliability with deep learning's sophistication to achieve state-of-the-art results in breast contour detection. This framework ensures adaptability and generalizability to limited data environments. This methodology provides a practical solution for detecting breast contours, which can pave the way for future exploration into more advanced deep learning and graph convolution networks. However, this is contingent upon the availability of more extensive datasets.

3.5 Conclusion

This research highlights the significance of advancing CBIR systems in the medical field. Recent studies, including the application of convolutional neural networks (CNNs), interpretability-guided networks, and attention mechanisms [44, 1, 51, 4], demonstrate the potential to enhance the accuracy and effectiveness of medical image retrieval. There has been a significant increase in the use of deep learning in retrieval of BCCT outcomes, more specifically in the field of image retrieval, leading to major advances. The studies developed by Cardoso et al. [3, 49, 51] produced a remarkable impact upon this area, serving as the basis for new studies such as the case of [4] and [12] which constitute the state of the art at the moment.

Silva et al. [12] with a DNN that automates the evaluation of aesthetic results, yet has the limitation that there is still intermediate supervision in the detachment of keypoints and the integration of pre-defined functions to translate those keypoints to asymmetry measures; and Freitas et al. [4] with a hybrid model in which he uses a U-net architecture with an integrated Sobel filter, obtaining good results in general, but performing poorly under very specific data points in which the breast contour edges were not as prominent.

Chapter 4

Image Segmentation

4.1 Image Segmentation: An overview

Image segmentation is a sub-field of computer vision that involves detecting ROIs in an image, with the aim of simplifying the image in question in order to facilitate analysis. This technique gained prominence with Serra's [52] proposal of an axiomatic definition for the concept of segmentation in image processing, based on the idea of a maximum partition. In the field of medical image segmentation, the most significant reference is the model proposed by Ronneberger [5] in 2015, which comprises a novel DNN for biomedical image segmentation, the U-Net. This network was specifically designed for image segmentation tasks, particularly in medical contexts, such as the segmentation of tissues or organs in MRI and CT images.

4.2 Application of U-Net Architecture in Image Segmentation

The U-Net structure comprises a contraction path, an expansion path and skip connections. The contraction path (encoder) follows the conventional architectural framework of a CNN, comprising multiple convolutional layers, activation functions such as rectified linear units (ReLU), and pooling operations, typically max pooling. These operations aim to reduce the spatial resolution of the image while increasing its depth, thereby capturing the most complex features. In contrast, the expansion path (decoder) employs upsampling operations to enhance the spatial resolution of the extracted image features. The skip connections between the encoder and decoder facilitate the utilization of the encoder's high-resolution information in the decoder, thereby enhancing the accuracy of the segmentation process. These connections assist the network in retaining the essential details of the original image, which is crucial for the accurate reconstruction of the segmented image.

The process of image segmentation is of paramount importance in this project, with the objective of detecting and isolating the patient's breasts through the creation of masks. It is, however, important to note that this segmentation process is a means of achieving a broader goal, namely the extraction of features from the torso region, which can then be used in the image retrieval process.

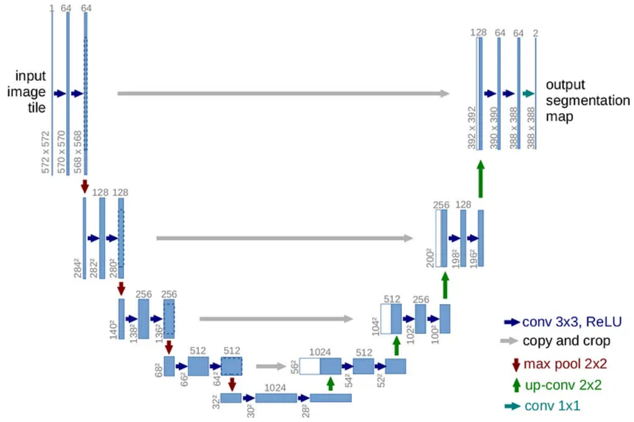


Figure 4.1: U-net architecture by [5].

4.3 Implementation and Results

To enable this process to be effective, previously created masks were used to train the model. In this case, a U-Net with some modifications to the original one described in ([5]) was employed (see Figure 4.1). Due to the satisfactory results of this model and its simplicity (when compared to other, more advanced and modern models), it was selected as the starting point of the project. The structure described by François Porcher[53] was used and later adapted to the needs of this project. In relation to the structure of the reference model, normalization layers were added in batches after each convolution, with the aim of speeding up training, having greater stability in learning, mitigating problems of gradient explosion or disappearance and increasing the robustness of the model. Dropout is also added in the downsampling layers (in the last one), in the bottleneck and in the first upsampling layer to prevent overfitting and make the network more general. Finally, instead of using transposed convolutions for upsampling, Upsample uses bilinear mode. The sigmoid function was selected as the activation function for the output of the U-Net network, given that the task at hand is a binary segmentation.

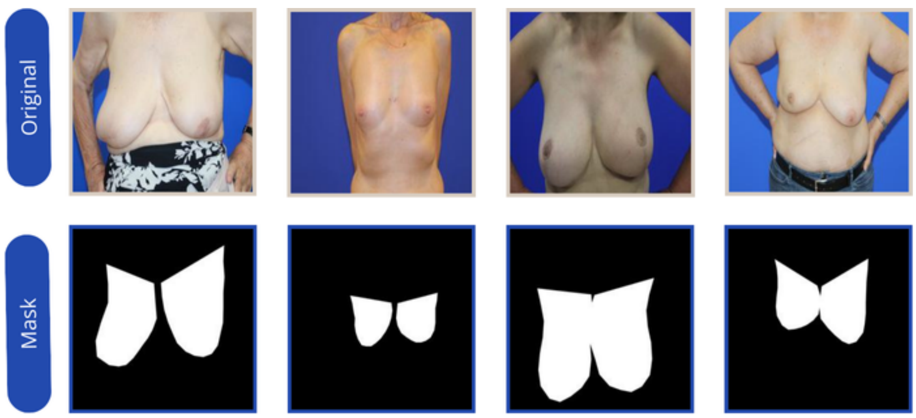


Figure 4.2: Exemples of the dataset orignal patient photos and the corresponding masks.

A dataset consisting of 1781 images with their corresponding masks was available (see Figure

4.2) and divided into 70% for training (1247 images), 15% for the validation dataset (267 images) and 15% for the test dataset (267 images). All the images were resized to 256×256 pixels and the color channels were normalized using the averages and standard deviations calculated from the dataset as shown in Table 4.1:

Channel	**R**	**G**	**B**
Averages	0.4506	0.4514	0.4939
Standard Deviations	0.2421	0.1717	0.1641

Table 4.1: Channel averages and standard deviations.

The evaluation of the model's performance was conducted using two primary metrics, the Dice coefficient (DC) and the mean squared error (MSE). The Dice coefficient (Equation 4.2) is utilized to assess the extent of overlap between the predicted and actual masks, and has been established as an effective measure for segmentation tasks. The MSE (Equation 4.1) is employed to quantify the mean difference between predicted and actual mask values, providing a supplementary perspective on the accuracy of the model.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.1}$$

where:

- n is the number of observations,
- y_i is the actual value,
- $\hat{y}_i$ is the predicted value.

The Dice Coefficient (DC) is given by the formula:

$$\text{DC} = \frac{2\sum_{i=1}^{n} p_i g_i}{\sum_{i=1}^{n} p_i + \sum_{i=1}^{n} g_i} \tag{4.2}$$

where:

- p_i is the predicted value,
- g_i is the ground truth value,
- n is the number of observations.

The combined loss function employed during model training incorporates both the Dice Coefficient (DC) and the binary cross-entropy (BCE) (Equation 4.3) , thereby ensuring a robust approach to optimization.

The Binary Cross-Entropy (BCE) is given by the formula:

$$\text{BCE} = -\frac{1}{n} \sum_{i=1}^{n} [g_i \log(p_i) + (1 - g_i) \log(1 - p_i)] \tag{4.3}$$

The combined loss function (Loss) is a weighted sum of DC and BCE:

$$\text{Loss} = \alpha \cdot \text{BCE} + \beta \cdot (1 - \text{DC}) \tag{4.4}$$

where:

- α and β are weighting factors that balance the contributions of BCE and DC, respectively.

The model was trained with 110 epochs, using the adaptive moment estimation (Adam) optimization function and implemented with Pytorch. In order to make the model more robust and more diverse, data augmentation techniques were used, namely "random rotation" and "color jitter". In the case of "random rotation", in which the image is rotated randomly between different angles, rotation values were limited to between -10 and 10. In the second transformation, a variation in image brightness and contrast is applied, and specific parameters were defined for these variations. The color jitter transformation makes it possible to adjust the brightness and contrast of the original image. The brightness is adjusted by a factor chosen at random within the interval [0.8, 1.2], and the contrast is adjusted by a factor within the interval [0.8, 1.2]. In image segmentation, a threshold is used to convert the probabilities predicted by the model into binary classifications (0 or 1). Initially, the threshold was set to 0.5, but this led to completely white segmented images, indicating that almost all pixels were classified as belonging to the class of interest. In order to gain a deeper insight into the behavior of the model's outputs, auxiliary functions were employed, which demonstrated that the majority of pixels exhibited probability values above 0.5. Following the analysis of these outputs, the threshold was adjusted to 0.6. The newly implemented threshold resulted in a reduction in the number of pixels classified as 1, thereby leading to more accurate and realistic segmentation.

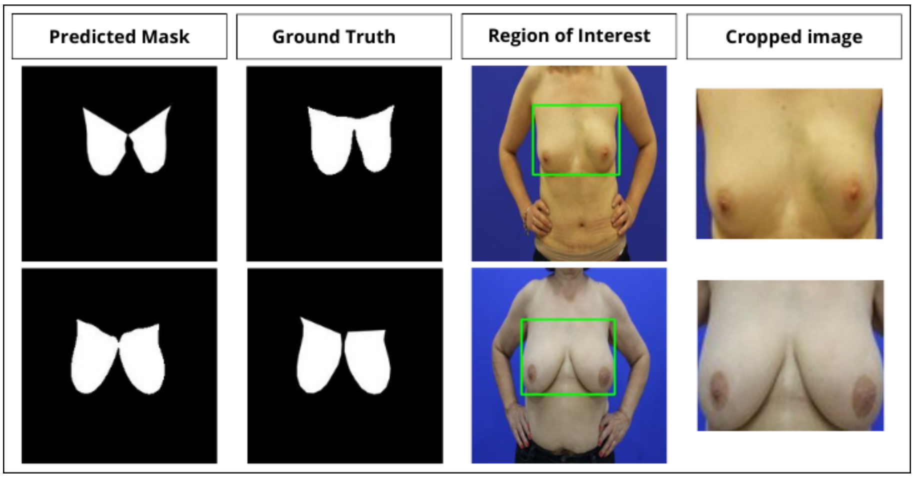


Figure 4.3: Examples of predicted segmentation masks, ROI and cropped images from ROI.

The results of the analysis demonstrated that adjusting the threshold proved to be an effective method for improving the segmentation results (see Figure 4.3). This highlights the importance of correctly adjusting this parameter based on an analysis of the model's outputs. The predicted segmentation masks were used to insert rectangular bounding boxes, which defined the region of interest (ROI). This region was selected as the area from which features were to be extracted in the subsequent step. To prevent the occurrence of multiple bounding boxes within a single image, due to the potential variation in breast spacing, and to ensure the inclusion of the entire breast region within the ROI, the extreme left, right, top, and bottom boundaries of the image were defined based on the contours identified. This resulted in the creation of a single bounding box encompassing all the contours detected. The image was then divided by the boundaries of the ROI, resulting in the final image.

As illustrated in Figure 4.3, the segmentation process yielded satisfactory results in the generation of the masks, which were in close alignment with the ground truth. In addition to the visual outcomes, the evaluation metrics also demonstrated favorable values in test (see Table 4.2):

Metrics	Score
Dice Score	0.9351
MSE	1.2865

Table 4.2: Results of evaluation metrics in test.

To assess the efficacy of the image segmentation model, two primary metrics were employed: DC and MSE. These metrics offer a comprehensive overview of the model's performance in terms of similarity and accuracy. The results of the evaluations are presented in Table 4.2, which represents the average of the test set results. The Dice Score, which assesses the similarity between the areas segmented by the model and the reference areas, reached a value of 0.9351, indicating a high degree of agreement between the model's predictions and the actual data. In contrast, the MSE metric, which assesses the accuracy of the predictions by calculating the average of the squares of the errors, obtained a value of 1.2865, reflecting the overall accuracy of the model. It was not the intention of this model to provide an exact representation of the breast region; rather, it was designed to facilitate the creation of masks with a certain degree of precision. However, the results demonstrate that this segmentation model is both effective and suitable for practical use.

4.3.1 Additional Applications of the Segmentation Model

The segmentation model was employed in a work authored by Pinheiro et al. [54], currently under review. Our model achieved competitive results in the task of detection of breast endpoints, which will be detailed in a paper currently under review. The detection of breast endpoints using this model was done by defining the endpoints as the points closest to the upper corners of the image. The application of the U-net model yielded satisfactory results, with a Dice score of 0.82 when generating the masks.

4.4 Conclusion

The introduction of image segmentation into various medical imaging pipelines has resulted in the enhancement of existing results [34, 44]. Although there are more sophisticated architectures than U-Net [34], it demonstrated its efficacy in our segmentation task, which involved identifying the patient's ROI and subsequently creating masks of that same region. This was achieved after training the network with a dataset of 1247 images. The results were achieved in terms of both the evaluation metrics used (DC and MSE) and at a visual level, where the similarities between the original masks (ground truth) and the masks generated by U-Net are considerable when they are compared.

Chapter 5

Experimental Work

In the context of breast cancer treatment aesthetic outcomes, integrating heterogeneous data can significantly enhance the precision of individualized medicine. This thesis aims to address the complexity and richness of breast cancer data by combining detailed tabular information (e.g., age, height, breast dimensions), with torso photographs of patients. Through this comprehensive approach, the goal is to find the most appropriate set of images for each patient, so that they can identify themselves whenever they look at the retrieved images [1].

5.1 Computer Vision Approach - Segmentation-based

The methodology involves analyzing medical images of breast cancer using advanced computer vision algorithms, specifically image segmentation techniques and feature-based methods. The goal is to develop a system that can identify and present images of similar cases, assisting healthcare professionals in making prognoses and communicating expectations of aesthetic and clinical results to patients. In order to achieve this goal, it is necessary to develop an effective segmentation model that allows the breast region to be identified precisely, so that the extraction of features only considers this area for image retrieval.

The proposed model's architectural design (see Figure 5.1) is intended to integrate image processing and clinical data in order to identify and return images of patients exhibiting similar characteristics. The process commences with a set of image data, which serves as input for the segmentation model. The segmentation model processes the input images to predict masks, which highlight regions of interest (ROIs) in the images. Based on the mask predictions, ROIs (i.e., the breasts) are identified in the images and the corresponding sections are cropped to isolate these regions. This component serves as a conduit between the images and the patient data, facilitating the matching of images to similar patient data. Subsequently, features are extracted from the processed images and the model utilizes pre-defined similarity criteria to identify analogous patients or images.

The final output of the model is a set of images that meet the specified similarity criteria, representing patients with similar characteristics based on the integrated analysis of image and clinical data.

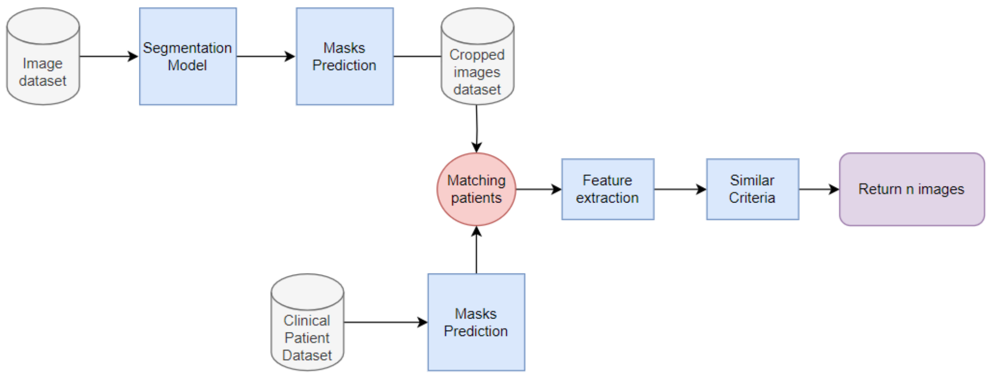


Figure 5.1: The proposed model architecture.

5.1.1 Advanced Feature Extraction Techniques for Image Retrieval

Once the ROI of each image has been identified and new images created, cropped by the limits of the ROI, the next step is to extract features. The aim is to determine whether the retrieval results improve with the extraction of features from the ROI. Feature extraction is the fundamental step that transforms raw image data into meaningful numerical representations, known as feature vectors. These representations capture essential information about the visual properties of images, such as colour, texture, shape and complex patterns. At the heart of this transformation are deep learning models such as convolutional neural networks (CNNs) and visual transformers (ViTs), which have revolutionized the way visual features are extracted and used. Zhou et al.[55] used a complex SIFT-based model for feature extraction and image matching, combining a colour histogram, a local directional pattern (LDP) used to describe texture and geometric shapes, and a Bag of Features (BoF) model based on SIFT to capture specific details of an image. The SIFT descriptor detects points of interest in the image and extracts their features, which are highly discriminative. In the BoF model, these local features are mapped to a code of words, creating a histogram of the occurrences of these words. A study by Ahmad[56] tested different feature extraction techniques for CBIR systems, specifically applied to computed tomography (CT) images of the brain. The aim was to identify the techniques that most effectively extract critical information to aid medical diagnosis, retrieval of similar types of disease and monitoring of patient progress. The study was conducted on a dataset of 3,032 CT images of the human brain. Each technique was evaluated based on its ability to retrieve relevant images from the database. The authors conclude that no single feature extraction technique is universally superior. Instead, a combination of methods that capture different aspects of the images (e.g., texture, shape, intensity) provides the most effective results for image retrieval in medical applications, improving the system's ability to support medical specialists in diagnosing and monitoring brain disease. In order to extract features, pre-trained models were employed, which, as evidenced by various articles [57, 58], offer a number of advantages in feature extraction, including enhanced performance on smaller datasets, greater efficiency in computer vision tasks, and improved accuracy in areas such as medical imaging. These models are capable of leveraging extensive pre-trained data and reducing computational load, rendering them a valuable tool in modern image processing. Pre-trained models are neural networks that have been previously trained on large datasets, such as ImageNet, and can be reused for specific tasks, such as feature extraction, without the need to train the model from scratch. The utilization of such models expedites the development process and enhances accuracy by capitalizing on the insights gleaned during the initial training phase.

In order to take advantage of these proven capabilities, ResNet50, Densenet121 and a ViT were employed. The Resnet50 neural network is distinguished by its residual network architecture (see Figure 5.2) , comprising 50 layers. This depth facilitates the learning of deeper networks. The network's innovative feature is the residual connections, which enable the network to learn a set of residual functions that map the input to the desired output. These residual connections permit the network to learn much deeper architectures than was previously possible, without the problem of vanishing gradients.

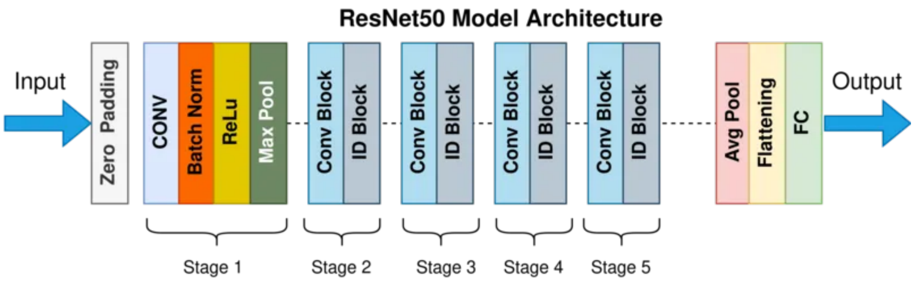


Figure 5.2: Traditional ResNet50 architecture.

Densenet121, on the other hand, has been designed to overcome the limitations of traditional deep networks. It features enhanced connectivity between layers, whereby each layer is connected to all previous layers. This feature promotes extensive reuse of features and allows for more efficient gradient flows during training, leading to improved performance and faster convergence.

In recent years, ViT have been employed extensively in medical imaging. In 2020, Dosovitski [59] proposed the ViT model for image classification, based on the transformer's self-attention mechanism, which achieved state-of-the-art (SOTA) results. Subsequently, ViT was also employed to analyze medical images [60], demonstrating a notable capacity to extract long-range features. This feature is of particular value in medical contexts, where the detection and accurate interpretation of subtle, scattered patterns in an image are crucial for accurate diagnoses. ViT employs the transformer's self-attention mechanism to capture distant spatial relationships within images, a capability that traditional convolution-based models may not possess to the same extent. This capacity enables ViT to identify anomalies and critical features that may be dispersed throughout the image, thereby rendering it a valuable tool for enhancing the accuracy and efficiency of medical analysis and diagnosis. Feng et al.[61] present an improved ViT architecture, ViT-Patch, for medical imaging, with the aim of identifying malignant breast ultrasound images. This article serves to illustrate once more the innovative capacity of VITs in this field.

In accordance with the methodology proposed by Martel et al.[42], the similarity of images, in this work, is quantified by means of the Euclidean distance, which is calculated between the feature vector of the query image and the feature vectors of all the other images in the dataset. This calculation results in a matrix of distances, where each element represents the Euclidean distance between a pair of images. The Euclidean distance (Equation 5.1) between two points $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ is given by:

$$d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{5.1}$$

This calculation results in a matrix of distances, where each element represents the Euclidean distance between a pair of images. The images are then ordered according to the calculated distances, with the smallest distances indicating the greatest similarity and the largest distances indicating the least similarity. In this instance, the eight most analogous and the eight most disparate images were identified.

The evaluation metrics employed for image recovery were Precision@K, Recall and NDCG (Normalized Discounted Cumulative Gain). Precision@K is defined as the proportion of items that are correctly identified among the first K items retrieved, according to the following formula (Equation 5.2):

$$\text{Precision@K} = \frac{\sum_{i=1}^{K} \text{relevance}(i)}{K} \tag{5.2}$$

The Recall (Equation5.3) statistic represents the proportion of relevant items retrieved up to position k in relation to the total number of relevant items available in the dataset:

$$\text{Recall} = \frac{\sum_{i=1}^{K} \text{relevance}(i)}{\text{Total number of relevant items}} \tag{5.3}$$

In recent articles [1, 42], NDCG has been employed as a metric to assess the efficacy of medical image retrieval. It evaluates the quality of the sorting of the results up to position K, taking into account not only the relevance of the items retrieved but also their position in the results list. The greater the relevance and the closer to the top of the list, the greater the gain.

The DCG (Equation 5.4) evaluates the quality of the sorting of results up to position K. It takes into account not only the relevance of the retrieved items, but also their position in the results list.

The higher the relevance and the closer to the top of the list, the greater the gain.

$$\mathrm{DCG} = \sum_{i=1}^{K} \frac{2^{\mathrm{rel}_i} - 1}{\log_2(i+1)} \tag{5.4}$$

Subsequently, the IDCG (Ideal Discounted Cumulative Gain) Equation 5.5 is calculated, which represents the DCG for an optimal ordering of the items.

$$\mathrm{IDCG} = \sum_{i=1}^{K} \frac{2^{\mathrm{rel}_i^{\mathrm{ideal}}} - 1}{\log_2(i+1)} \tag{5.5}$$

Finally, the NDCG (Equation 5.6) is defined as the ratio between DCG and IDCG:

$$\mathrm{NDCG} = \frac{\mathrm{DCG}}{\mathrm{IDCG}} \tag{5.6}$$

NDCG ranges from 0 to 1, where 1 indicates a perfect ordering of the results up to position K.

5.1.2 Implementation and Results

The pre-trained models were implemented in Python using PyTorch and the images resized according to the requirements of each model. Regarding the Resnet-50, the pre-trained version was employed with weights 'IMAGENET1K V1' and the modified network architecture, which involved the removal of the final classification layer and the retention of the convolutional layers responsible for extracting visual feature. Similarly to ResNet, the final layer of the DenseNet121 was removed and an adaptive pooling layer was added to adjust the dimensions of the output. The pre-trained version with 'IMAGENET1K V1' weights was also used. In light of the outcomes demonstrated by ViT, we proceeded to implement a ViT, also a pre-trained model from the "Timm" library, `vit_base_patch16_224`. The classification head was removed in order to permit the extraction of the features that had been learned during the pre-training phase.

The experiments involved the use of both original images and images cropped by ROI. The models were tested in accordance with the following methodology. Firstly, the euclidean distance matrix between the feature vectors of all the images in the dataset was calculated. This distance matrix allowed for the identification of the most similar and least similar images for each query image. Secondly, for each image in the dataset, the eight most similar images and the eight least similar images were determined, with the exclusion of the image itself from the results list. The indices of the most similar and least similar images are obtained by sorting the distance values in ascending order and selecting the first eight indices. The results of the most similar and least similar images for each image are then stored in a dictionary list, which is subsequently converted into a Dataframe.

Subsequently, evaluation metrics for the analogous images were calculated utilizing the Precision, Recall and NDCG metrics. For each image in the dataset, a comparison is made between the list of truly similar images (ground truth) and the list of similar images obtained by the algorithm. The Precision, Recall and NDCG metrics were calculated for a value of k = 8.

In contrast to the approach taken with the analogous images, the evaluation metrics could not be employed in the case of the less analogous images, as no images had been catalogued with the ranking of the less analogous ones. Consequently, the evaluation had to be conducted visually, by comparing the query image with the retrieved images.

Finally, the average of the evaluation metrics is calculated for all images in the dataset of the most similar patients. The results of the averages of the Precision, Recall and NDCG metrics are presented in order to provide an overview of the performance of the similarity algorithm (see Table below 5.1).

Models	**Precision**	**Recall**	**nDCG**	**Image type**
Resnet-50	0.0104	0.0073	0.0111	Cropped
Densenet-121	0.0117	0.0084	0.0124	Cropped
ViT	**0.0145**	**0.0101**	**0.0167**	Cropped
Resnet-50	0.0087	0.0060	0.0086	Original
Densenet-121	0.0124	0.0090	0.0127	Original
ViT	**0.0137**	**0.0094**	**0.0149**	Original

Table 5.1: Results for the Retrieval Experiments Conducted - Similar Images. The bold results represent the best model for each evaluation metric.

Table 5.1 demonstrate that when Euclidean distance is employed solely for the purpose of measuring similarity through feature extraction, the resulting outcomes are not particularly meaningful. Nevertheless, the results suggest that, although the outcomes are relatively low, there is a discernible difference in the results obtained when images are cropped by ROI after segmentation and the original images except for Densenet121. In terms of image similarity, ViT is the model that demonstrates the most favorable results, both with cropped images and with Precision 0.0145, Recall 0.0101 and nDCG 0.0167. Furthermore, when original images are used, the model achieves Precision 0.0137, Recall 0.0094 and nDCG 0.0149. This demonstrates that the first case has the most favorable evaluation metrics. Conversely, Densenet121 shows the best results after extracting features from the original images, albeit with a very slight advantage. In the absence of a suitable method for comparing images of disparate characteristics, we employed image comparison (see image 5.3) to ascertain the plausibility of the results. The comparison revealed notable discrepancies in the query patient's skin structure and tone between the dissimilar images retrieved, particularly when compared to the similar images.

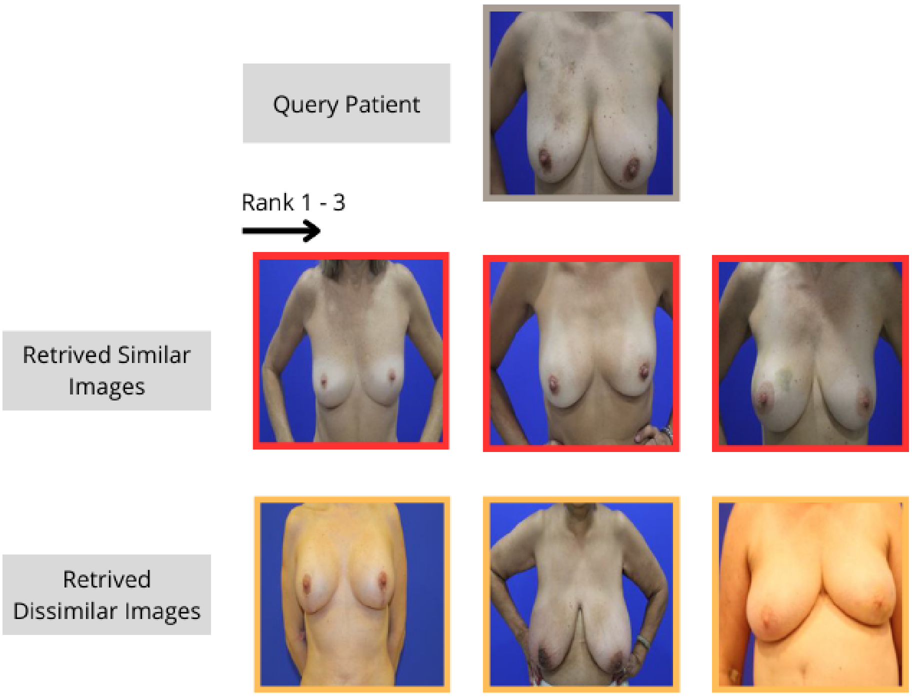


Figure 5.3: Retrieval results by ViT model with cropped images.

The Figure 5.3 illustrates the results of image retrieval after feature extraction by the ViT model, which produced the best results. However, as previously mentioned, these results were not in accordance with the "ground truth", which is the ranking and ordering of images by clinicians. The image at the top is the image of the patient that is being used as a reference to find similar and non-similar images. The images below are the three retrieved images that are most similar to the query patient. The images in the red box do not conform to the ground truth. The images in the bottom row are dissimilar retrieved images and have been included to demonstrate the model's ability to distinguish between similar and dissimilar images. Despite not conforming to the ground truth, the retrieved images show a significant visual correspondence with the query image, while the dissimilar ones show a significant visual distance to the same image. The discrepancy between the images identified as similar by ground truth and retrieved by the model indicates a potential discrepancy between the similarity metrics and the models' opinions with respect to the similarity criteria.

5.2 Machine Learning for Tabular Data

In a preliminary approach, with the objective of establishing a baseline for this project, tabular patient data (including age, height, etc.) was employed to predict patient similarity and dissimilarity. This approach also enables the identification of patterns and correlations in the patient images in the database. The processing and analysis of the tabular data was carried out with the aim of establishing significant relationships between the patient biometric data and the expectations of the results after treatment [1].

The dataset encompasses a range of patient-related attributes, including:

- Patient height: Patient's height in centimeters;
- Patient's weight: Patient's weight in kilograms;
- Cancer Side: The side of the body affected by the cancer;
- Bra Size Number: Numerical representation of the patient's bra size;
- Mapped Bra Size: Numerical mapping of the patient's bra size;
- Age: Patient's age.

Given the diverse nature of these attributes, it is of critical importance to ensure that they are on the same scale through the process of normalization. This is essential for the accurate calculation of distances, as attributes with larger numerical ranges have the potential to disproportionately influence the results. The MinMaxScaler function was employed to normalize the data, with each attribute transformed into a range between 0 and 1. To assess the similarity between patients, the Euclidean distance between the normalized attribute vectors was utilized. This distance provides a quantitative measure of similarity, with smaller distances indicating greater similarity.

To quantify the similarity between patients, the Euclidean distance between their normalized attribute vectors was calculated. The Euclidean distance provides a quantitative measure of similarity, with smaller distances indicating greater similarity. To assess the accuracy of our similarity retrieval process, we use three main metrics: Precision@k, Recall@k and NDCG@k. The results of the evaluation metrics obtained are significantly inferior to those of the previous results that employed segmentation (see Table 5.2).

	Precision	**Recall**	**nDCG**
Similarity	0.0125	0.0084	0.0104

Table 5.2: Tabular data results

Similarly, in the tabular data model, the similarity of patients was assessed using metrics, as was the case with the computer vision model. In order to facilitate comparison between patients with the least similarity, images are employed for the objective and subjective assessment of the retrieval of tabular data.

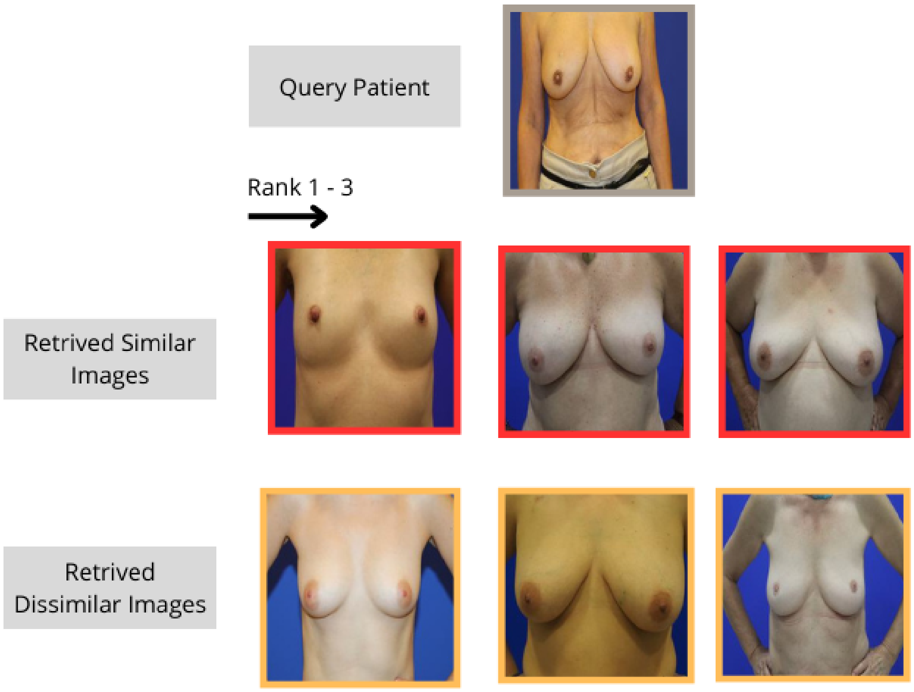


Figure 5.4: Tabular model retrieval results.

The Figure 5.4 depicts the retrieval result of a single patient and the three closest and three most distant patients identified by the tabular model. With regard to the dissimilar images, it is evident that they do appear to differ from the original image. However, this is merely a subjective visual assessment, as there is no method of comparison. The most distant image does indeed exhibit a markedly reduced breast volume in comparison to the image of the query patient. With regard to the other recovered images, it is possible that factors such as age and height may have had a greater impact on the distance calculation. With regard to the similar images, it appears that they are not as closely aligned with the query patient as the computer vision model. This discrepancy may be attributed to the influence of additional factors, such as age, height, or weight, which have a greater impact on the distance calculation. Consequently, it can be concluded that this model is less accurate than the computer vision model.

5.3 Retrieval Framework

To further test the theory of image segmentation as a means of extracting more interesting features for image retrieval, we used the framework developed by Zolfagharnasab et al. [62], currently under review. This framework is designed to train and evaluate deep learning models using the triplet loss technique, which is effective in metric learning tasks [63, 64]. This equation (see Equation 5.7) is used to ensure that the distance between a point (a) and a positive point (p) (belonging to the same class) is smaller than the distance between a point (a) and a negative point (n), belonging to a different class, by the margin of α, the equation being given by:

$$\mathscr{T}(a,p,n) = \frac{1}{2N}\sum_{i=1}^{N}\max\left[0, m + |a_i - p_i|^2 - |a_i - n_i|^2\right] \tag{5.7}$$

where a_i is the anchor, p_i is the positive example, n_i is the negative example, $|\cdot|^2$ represents the squared Euclidean distance, m is the margin, and N is the number of triplets in the batch.

As can be seen in this equation, configurations where the anchor point is closer to the negative point are penalized, thus encouraging the network to learn where the points of the same class are closer.

The triplet loss is integrated in the current deep-learning technique to maximize retrieval results. The combination of triplet loss and deep neural networks provides a robust approach to learning discriminative representations of data. Pre-trained models such as ResNet50, DeiT and BEiT were used, which, having been trained on large datasets, already capture rich, hierarchical features of images, making them excellent starting points for various computer vision tasks. These models are used to extract features in the form of vectors that capture the visual essence of the image. Such vectors are then used to compute distances and similarities between images. In accordance with the methodology employed in the preceding experiment, the segmentation stage was incorporated as a preliminary step prior to utilizing this framework, with the objective of evaluating the potential added value of this stage. Consequently, both the original images and the cropped images following the identification of the ROI were employed to feed the pre-trained models. The images were resized in accordance with the input criteria of each of the models. The learning rate was set at 0.0001, Adam was selected as the optimization function, the datasets were divided into a ratio of 80/20 (train and test), and the images cataloged as "Excellent/Good" and "Fair/Poor" were used initially.

The metrics employed were those of Accuracy and Adjusted Normalized Discounted Cumulative Gain (AnDCG). As illustrated in the Tables 5.3 and 5.4, the models trained with the "Cropped" images once again demonstrated the most optimal performance in the test. For the "Excelent/Good" outcomes the model with a transformer architecture, BEiT, demonstrated the most favourable results with both image types, achieving an accuracy of 61% and AnDCG of 87.8 in the cropped images and an accuracy of 56% and AnDCG of 86.5 in the original images. Also, the ResNet-50 model demonstrated a 2% higher accuracy between the models trained with cropped images and those trained with original images. The test conducted with the extraction of features using the DeiT model demonstrated that there was no discernible difference in the results obtained when images were used. Regarding the "Fair/Poor" outcomes the results also demonstrate that by using segmented images the framework can benefit from it. The ResNet shows an accuracy of 0.53 and AnDCG of 84.5 for the segmented images while the same model shows an accuracy of 0.51 and AnDCG of 83. for the original images. Deit shows that does not benefits from the segmented images as it demonstrates the same results for both image types with accuracy of 0.52 and AnDCG of 85. The best model was, once more, the BeiT model with an accuracy in the ”Cropped" images 1% higher (0.61) then the original images (0.60) and AnDCG 1.5 points higher comparing with the original images.

Models	Accuracy	AnDCG	Image type
Resnet-50	0.56	86.6	Cropped
DeiT	0.51	84.8	Cropped
BEiT	**0.61**	**87.8**	Cropped
Resnet-50	0.54	85.6	Original
DeiT	0.51	84.8	Original
BEiT	**0.56**	**86.5**	Original

Table 5.3: Retrieval Framework Results "Excelent/Good".

Models	Accuracy	AnDCG	Image type
Resnet-50	0.53	84.5	Cropped
DeiT	0.52	85	Cropped
BEiT	**0.61**	**82.4**	Cropped
Resnet-50	0.51	83.6	Original
DeiT	0.52	85	Original
BEiT	**0.60**	**80.9**	Original

Table 5.4: Retrieval Framework Results "Fair/Poor".

These findings provide empirical support for the hypothesis that feature extraction with ROI-cropped images is advantageous and improves the results. This suggests that image segmentation to identify the ROI in the first phase is an important process.

5.4 The dataset

In the development of this study, the dataset utilized plays a crucial role in training and validating the proposed computer vision models. This dataset is divided into two primary categories based on the aesthetic outcomes of breast cancer treatment: "Excellent/Good" and "Fair/Poor". The dataset comprises information collated by clinicians on the most similar patients to each patient, categorized according to the clinical assessment of the patient's overall performance. The number of similar patients for each patient varies between 8 and 12. It should be noted that some patients have incomplete information. In addition to the aforementioned information, the dataset also contains five patient attributes: age, height, weight, bra cup, and bra size. Following the cleansing of the dataset, the final number of patients with complete information was 180.

5.5 Conclusion

In the image segmentation phase, the U-Net model demonstrated satisfactory performance, producing masks that were highly representative of the torso region in the original image. This enhanced the potential for success in the subsequent stage of the experiment, namely the extraction of features. Although the U-Net model is not the most recent iteration and there are more sophisticated models available, it remains a highly competent model and is more accessible for implementation. Regarding the second phase, utilization of an image segmentation model, such as U-Net, has been observed to have a positive impact on image retrieval, although not to the extent that was initially anticipated. The task of ranking similar images using only image feature extraction is a complex one. The objective of this experiment was to ascertain whether there is any advantage in using image segmentation for feature extraction. The results of this experiment indicate that there is some evidence of this advantage. It is crucial to highlight that the outcomes, despite diverging from the ground truth, indicate the presence of both visual similarity and dissimilarity in the images. Specifically, the images retrieved to exemplify dissimilarity exhibit a notable divergence from the ground truth. This indicates that there may be a discrepancy between the concept of similarity as perceived by the clinicians and that ascribed to the model images (where similarity is gauged by euclidean distance). As previously demonstrated in the literature [56], the extraction of features from images is a valuable approach in the field of medical image retrieval, as is the utilization of pre-trained models ([57, 58]). It would be of interest to apply a trainable model to optimize feature extraction or train image retrieval. Regarding the tabular data model, it indicates that this option is not an optimal choice, despite the relatively minor discrepancy in evaluation metrics compared to the model employing image segmentation. The retrieved images of the patients appear to deviate significantly from the expected visual outcome. Finally, the we could test our concept by applying the segmentation phase to a framework pipeline that uses the triplet loss to train and evaluate the model. This framework could achieve better results than our models in terms of evaluation metrics and it also shows that the segmentation integrated in the pipeline, where it uses the ROI images, has better results than using the original images.

Chapter 6

Conclusion and Future Work

6.1 Report Overview

This thesis addresses the capabilities of deep image segmentation with the aim of improving the retrieval of medical images, with a specific focus on the detection of the torso region in images of breast cancer patients. Two approaches were trialled: one using the patient's tabular data (biometric data), such as age, breast size, height, weight, etc., and the other using advanced computer vision techniques to detect the torso region and then create masks for that area. In the computer vision approach, a segmentation model capable of identifying the breast region was developed and trained, ensuring that feature extraction was limited exclusively to that specific area. In order to take advantage of the capabilities of pre-trained deep learning models (ViT, ResNet, DenseNet), feature extraction was performed using these models [57, 58]. Consequently, image retrieval was enhanced by focusing solely on ROIs, leading to enhanced accuracy in identifying patients with similar characteristics. The experimental results demonstrated that integrating the segmentation model into the image retrieval pipeline led to a notable improvement in precision, recall, and NDCG compared to methods that did not utilize segmentation. In both approaches, the Euclidean distance was employed to quantify the degree of similarity or dissimilarity between the feature/attribute vectors. These findings not only validate the effectiveness of image segmentation for retrieving medical images, but also suggest potential avenues for the continuous improvement of this technology. Finally, a framework developed for a paper still under submission was tested, which used triplet loss to train and evaluate the models. In this framework, we implemented the segmentation task in its pipeline and the results demonstrated that this process is advantageous for image retrieval, with an increase in the results of the evaluation metrics compared to the same task without the use of image segmentation. In conclusion, this work contributes to the advancement of medical image retrieval techniques by providing a more accurate and objective analysis of images. It is hoped that this will assist healthcare professionals and breast cancer patients with more accurate image retrievals that are more similar to their own characteristics. Further research may investigate the potential of more sophisticated deep learning techniques and the integration of additional imaging modalities to enhance the precision of medical image retrieval.

6.2 Future Work

As previously stated, there are several areas for improvement in the image retrieval process. For future research, it would be beneficial to investigate ways to enhance the alignment between the evaluation metrics and the clinicians' opinions. Currently, the images retrieved by our algorithm appear to be similar to those of the consultation patient. The use of triplet or quadruplet loss could assist in adjusting the algorithms to align more closely with the expectations and perceptions of healthcare professionals. Furthermore, a pipeline that employs tabular data in conjunction with the extraction of visual characteristics would be an intriguing experiment, which could potentially enhance the outcomes of image retrieval. The integration of structured information (such as clinical and demographic data) with unstructured data (images) can provide a more comprehensive and accurate representation, significantly enhancing the efficacy of image retrieval systems.

References

[1] W. Silva, T. Gonçalves, K. Härmä, E. Schröder, V. C. Obmann, M. C. Barroso, A. Poellinger, M. Reyes, and J. S. Cardoso, "Computer-aided diagnosis through medical image retrieval in radiology," *Scientific Reports*, vol. 12, p. 20732, Dec. 2022. Number: 1 Publisher: Nature Publishing Group.

[2] T. Forster, C. Köhler, M. Dorn, M. F. Häfner, N. Arians, L. König, S. B. Harrabi, I. Schlampp, E. Meixner, V. Heinrich, N. Weidner, M. Golatta, A. Hennigs, J. Heil, H. Hof, D. Krug, J. Debus, and J. Hörner-Rieber, "Methods of Esthetic Assessment after Adjuvant Whole-Breast Radiotherapy in Breast Cancer Patients: Evaluation of the BCCT.core Software and Patients' and Physicians' Assessment from the Randomized IMRT-MC2 Trial," *Cancers*, vol. 14, p. 3010, Jan. 2022. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

[3] J. S. Cardoso and M. J. Cardoso, "Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment," *Artificial Intelligence in Medicine*, vol. 40, pp. 115–126, June 2007.

[4] N. Freitas, D. Silva, C. Mavioso, M. J. Cardoso, and J. S. Cardoso, "Deep Edge Detection Methods for the Automatic Calculation of the Breast Contour," *Bioengineering*, vol. 10, p. 401, Apr. 2023. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015. arXiv:1505.04597 [cs].

[6] M. Noguchi, Y. Saito, Y. Mizukami, A. Nonomura, N. Ohta, N. Koyasaki, T. Taniya, and I. Miyazaki, "Breast deformity, its correction, and assessment of breast conserving surgery," *Breast Cancer Research and Treatment*, vol. 18, pp. 111–118, May 1991.

[7] K. H. Hwang, H. Lee, and D. Choi, "Medical Image Retrieval: Past and Present," *Healthcare Informatics Research*, vol. 18, pp. 3–9, Mar. 2012. Publisher: Korean Society of Medical Informatics.

[8] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, pp. 262–282, Jan. 2007.

[9] S. A. Napel, C. F. Beaulieu, C. Rodriguez, J. Cui, J. Xu, A. Gupta, D. Korenblum, H. Greenspan, Y. Ma, and D. L. Rubin, "Automated Retrieval of CT Images of Liver Lesions on the Basis of Image Similarity: Method and Preliminary Results," *Radiology*, vol. 256, pp. 243–252, July 2010. Publisher: Radiological Society of North America.

[10] M. Owais, M. Arsalan, J. Choi, and K. R. Park, "Effective Diagnosis and Treatment through Content-Based Medical Image Retrieval (CBMIR) by Using Artificial Intelligence," *Journal*

of Clinical Medicine, vol. 8, p. 462, Apr. 2019. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[11] F. Fitzal, W. Krois, H. Trischler, L. Wutzel, O. Riedl, U. Kühbelböck, B. Wintersteiner, M. J. Cardoso, P. Dubsky, M. Gnant, R. Jakesz, and T. Wild, "The use of a breast symmetry index for objective evaluation of breast cosmesis," *Breast (Edinburgh, Scotland)*, vol. 16, pp. 429–435, Aug. 2007.

[12] W. Silva, M. Carvalho, C. Mavioso, M. J. Cardoso, and J. S. Cardoso, "Deep Aesthetic Assessment and Retrieval of Breast Cancer Treatment Outcomes," May 2022. arXiv:2205.12611 [cs].

[13] U. Veronesi, P. Boyle, A. Goldhirsch, R. Orecchia, and G. Viale, "Breast cancer," *The Lancet*, vol. 365, pp. 1727–1741, May 2005. Publisher: Elsevier.

[14] J. Ma and A. Jemal, "Breast Cancer Statistics," in *Breast Cancer Metastasis and Drug Resistance: Progress and Prospects* (A. Ahmad, ed.), pp. 1–18, New York, NY: Springer, 2013.

[15] Y. Feng, M. Spezia, S. Huang, C. Yuan, Z. Zeng, L. Zhang, X. Ji, W. Liu, B. Huang, W. Luo, B. Liu, Y. Lei, S. Du, A. Vuppalapati, H. H. Luu, R. C. Haydon, T.-C. He, and G. Ren, "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis," *Genes & Diseases*, vol. 5, pp. 77–106, June 2018.

[16] N. Harbeck, F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang, and F. Cardoso, "Breast cancer," *Nature Reviews Disease Primers*, vol. 5, pp. 1–31, Sept. 2019. Number: 1 Publisher: Nature Publishing Group.

[17] BCRF, "Types of Breast Cancer | BCRF," Aug. 2023.

[18] E. J. Watkins, "Overview of breast cancer," *JAAPA*, vol. 32, p. 13, Oct. 2019.

[19] G. N. Sharma, R. Dave, J. Sanadya, P. Sharma, and K. K. Sharma, "VARIOUS TYPES AND MANAGEMENT OF BREAST CANCER: AN OVERVIEW," *Journal of Advanced Pharmaceutical Technology & Research*, vol. 1, no. 2, pp. 109–126, 2010.

[20] B. Weigelt, F. C. Geyer, and J. S. Reis-Filho, "Histological types of breast cancer: How special are they?," *Molecular Oncology*, vol. 4, pp. 192–208, June 2010.

[21] A. G. Waks and E. P. Winer, "Breast Cancer Treatment," *JAMA*, vol. 321, p. 316, Jan. 2019.

[22] L. Wang, "Early Diagnosis of Breast Cancer," *Sensors*, vol. 17, p. 1572, July 2017. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[23] O. J. Scully, B.-H. Bay, G. Yip, and Y. Yu, "Breast Cancer Metastasis," *Cancer Genomics & Proteomics*, vol. 9, pp. 311–320, Sept. 2012. Publisher: International Institute of Anticancer Research.

[24] L. M. Apantaku, "Breast Cancer Diagnosis and Screening," *American Family Physician*, vol. 62, pp. 596–602, Aug. 2000.

[25] Roberto Eduardo Guerra Estrada, Andrés Felipe Díaz Muñoz, Fred David Delgado Ricardo, Astrid Carolina Saavedra Andrade, María Morera-Esquivel, Angela María Piamba Valencia, Omar Avendaño Solano, and Oscar Ivan Avendaño Solano, "Use of mammography with contrast for the diagnosis of breast cancer," *World Journal of Advanced Research and Reviews*, vol. 14, pp. 277–283, Apr. 2022.

[26] B. L. Sprague, L. Ichikawa, J. Eavey, K. P. Lowry, G. Rauscher, E. S. O'Meara, D. L. Miglioretti, S. Chen, J. M. Lee, N. K. Stout, J. S. Mandelblatt, N. Alsheik, S. D. Herschorn, H. Perry, D. L. Weaver, and K. Kerlikowske, "Breast cancer risk characteristics of women undergoing whole-breast ultrasound screening versus mammography alone," *Cancer*, vol. 129, no. 16, pp. 2456–2468, 2023. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cncr.34768.

[27] S. M. Rosenberg, L. S. Dominici, S. Gelber, P. D. Poorvu, K. J. Ruddy, J. S. Wong, R. M. Tamimi, L. Schapira, S. Come, J. M. Peppercorn, V. F. Borges, and A. H. Partridge, "Association of Breast Cancer Surgery With Quality of Life and Psychosocial Well-being in Young Breast Cancer Survivors," *JAMA Surgery*, vol. 155, pp. 1035–1042, Nov. 2020.

[28] M. Riis, "Modern surgical treatment of breast cancer," *Annals of Medicine and Surgery*, vol. 56, pp. 95–107, Aug. 2020.

[29] K. L. Maughan, M. A. Lutterbie, and P. S. Ham, "Treatment of Breast Cancer," *American Family Physician*, vol. 81, pp. 1339–1346, June 2010.

[30] A. Luini, G. Gatti, V. Galimberti, S. Zurrida, M. Intra, O. Gentilini, G. Paganelli, G. Viale, R. Orecchia, P. Veronesi, and U. Veronesi, "Conservative Treatment of Breast Cancer: Its Evolution," *Breast Cancer Research and Treatment*, vol. 94, pp. 195–198, Dec. 2005.

[31] H. P. Oliveira, J. S. Cardoso, A. Magalhaes, and M. J. Cardoso, "Methods for the Aesthetic Evaluation of Breast Cancer Conservation Treatment: A Technological Review," *Current Medical Imaging Reviews*, vol. 9, pp. 32–46, Apr. 2013.

[32] K. Feng, Z. Jia, G. Liu, Z. Xing, J. Li, J. Li, F. Ren, J. Wu, W. Wang, J. Wang, J. Liu, and X. Wang, "A review of studies on omitting surgery after neoadjuvant chemotherapy in breast cancer," *American Journal of Cancer Research*, vol. 12, pp. 3512–3531, Aug. 2022.

[33] R. P. Gruber, R. A. Kahn, H. Lash, M. R. Maser, D. B. Apfelberg, and D. R. Laub, "Breast Reconstruction Following Mastectomy: A Comparison of Submuscular and Subcutaneous Techniques," *Plastic and Reconstructive Surgery*, vol. 67, p. 312, Mar. 1981.

[34] W. Silva, E. Castro, M. J. Cardoso, F. Fitzal, and J. S. Cardoso, "Deep Keypoint Detection for the Aesthetic Evaluation of Breast Cancer Surgery Outcomes," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, (Venice, Italy), pp. 1082–1086, IEEE, Apr. 2019.

[35] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, pp. 198–211, June 2007.

[36] H. K. Huang, "Utilization of medical imaging informatics and biometrics technologies in healthcare delivery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, pp. 27–39, June 2008.

[37] T. Gevers and A. W. M. Smeulders, "Content-based image retrieval by viewpoint-invariant color indexing," *Image and Vision Computing*, vol. 17, pp. 475–488, May 1999.

[38] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1349–1380, Dec. 2000.

[39] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-Based Medical Image Retrieval: A Survey of Applications to Multidimensional and Multimodality Data," *Journal of Digital Imaging*, vol. 26, pp. 1025–1039, Dec. 2013.

[40] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori, "Automated Storage and Retrieval of Thin-Section CT Images to Assist Diagnosis: System Description and Preliminary Assessment," *Radiology*, vol. 228, pp. 265–270, July 2003. Publisher: Radiological Society of North America.

[41] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, Nov. 2017.

[42] W. Silva, A. Poellinger, J. S. Cardoso, and M. Reyes, "Interpretability-Guided Content-Based Medical Image Retrieval," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, eds.), vol. 12261, pp. 305–314, Cham: Springer International Publishing, 2020. Series Title: Lecture Notes in Computer Science.

[43] J. R. Harris, M. B. Levene, G. Svensson, and S. Hellman, "Analysis of cosmetic results following primary radiation therapy for stages I and II carcinoma of the breast," *International Journal of Radiation Oncology*Biology*Physics*, vol. 5, pp. 257–261, Feb. 1979.

[44] T. Gonçalves, W. Silva, M. J. Cardoso, and J. S. Cardoso, "A novel approach to keypoint detection for the aesthetic evaluation of breast cancer surgery outcomes," *Health and Technology*, vol. 10, pp. 891–903, July 2020.

[45] R. D. Pezner, M. P. Patterson, L. R. Hill, N. Vora, K. R. Desai, J. O. Archambeau, and J. A. Lipsett, "Breast retraction assessment: an objective evaluation of cosmetic results of patients treated conservatively for breast cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 11, pp. 575–578, Mar. 1985.

[46] E. Van Limbergen, E. van der Schueren, and K. Van Tongelen, "Cosmetic evaluation of breast conserving treatment for mammary cancer. 1. Proposal of a quantitative scoring system," *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology*, vol. 16, pp. 159–167, Nov. 1989.

[47] L. I. Tsouskas and I. S. Fentiman, "Breast compliance: A new method for evaluation of cosmetic outcome after conservative treatment of early breast cancer," *Breast Cancer Research and Treatment*, vol. 15, pp. 185–190, May 1990.

[48] J. S. Cardoso and M. J. Cardoso, "Breast Contour Detection for the Aesthetic Evaluation of Breast Cancer Conservative Treatment," in *Computer Recognition Systems 2* (M. Kurzynski, E. Puchala, M. Wozniak, and A. Zolnierek, eds.), Advances in Soft Computing, (Berlin, Heidelberg), pp. 518–525, Springer, 2007.

[49] M. J. Cardoso and J. Cardoso, "Automatic Breast Contour Detection in Digital Photographs," 2008.

[50] N. Udpa, M. P. Sampat, M. S. Kim, G. P. Reece, and M. K. Markey, "Objective assessment of the aesthetic outcomes of breast cancer treatment: toward automatic localization of fiducial points on digital photographs," in *Medical Imaging 2007: Computer-Aided Diagnosis* (M. L. Giger and N. Karssemeijer, eds.), vol. 6514, p. 651420, International Society for Optics and Photonics, SPIE, 2007.

[51] J. S. Cardoso, I. Domingues, and H. P. Oliveira, "Closed Shortest Path in the Original Coordinates with an Application to Breast Cancer," *International Journal of Pattern Recognition and Artificial Intelligence*, Jan. 2015. Publisher: World Scientific Publishing Company.

[52] J. Serra, "Image segmentation," in *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, vol. 1, pp. I–345, Sept. 2003. ISSN: 1522-4880.

[53] F. Porcher, "Cooking your first U-Net for Image Segmentation," Sept. 2023.

[54] N. Freitas, M. Veloso, C. Mavioso, M. J. Cardoso, H. P. Oliveira, and J. S. Cardoso, "Endpoint Detection in Breast Images for Automatic Classification of Breast Cancer Aesthetic Results," 2024.

[55] J. Zhou, X. Liu, W. Liu, and J. Gan, "Image retrieval based on effective feature extraction and diffusion process," *Multimedia Tools and Applications*, vol. 78, pp. 6163–6190, Mar. 2019.

[56] W. S. H. M. W. Ahmad and M. F. Ahmad Fauzi, "Comparison of different feature extraction techniques in content-based image retrieval for CT brain images," in *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pp. 503–508, Oct. 2008.

[57] E. d. S. Puls, M. V. Todescato, and J. L. Carbonera, "An evaluation of pre-trained models for feature extraction in image classification," Oct. 2023. arXiv:2310.02037 [cs].

[58] A. Ahmed, "Pre-trained CNNs Models for Content based Image Retrieval," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, Sept. 2021. Number: 7 Publisher: The Science and Information (SAI) Organization Limited.

[59] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," June 2021. arXiv:2010.11929 [cs].

[60] L. Song, G. Liu, and M. Ma, "TD-Net:unsupervised medical image registration network based on Transformer and CNN," *Applied Intelligence*, vol. 52, pp. 18201–18209, Dec. 2022.

[61] H. Feng, B. Yang, J. Wang, M. Liu, L. Yin, W. Zheng, Z. Yin, and C. Liu, "Identifying Malignant Breast Ultrasound Images Using ViT-Patch," *Applied Sciences*, vol. 13, p. 3489, Jan. 2023. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

[62] M. H. Zolfagharnasab, N. Freitas, T. Gonçalves, E. Bonci, C. Mavioso, M. J. Cardoso, H. P. Oliveira, and J. S. Cardoso, "Predicting Aesthetic Outcomes in Breast Cancer Surgery: A Multimodal Retrieval Approach," 2024.

[63] A. D'Innocente, N. Garg, Y. Zhang, L. Bazzani, and M. Donoser, "Localized Triplet Loss for Fine-grained Fashion Image Retrieval," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (Nashville, TN, USA), pp. 3905–3910, IEEE, June 2021.

[64] T. Bui, L. Ribeiro, M. Ponti, and J. Collomosse, "Compact descriptors for sketch-based image retrieval using a triplet loss convolutional neural network," *Computer Vision and Image Understanding*, vol. 164, pp. 27–37, Nov. 2017.