INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA DE INTELIGÊNCIA ARTIFICIAL




# Aplicação técnicas aprendizagem automática no cancro da mama

**JOSÉ CARLOS CORDEIRO ANDRADE SANTOS**
Outubro de 2022

P.PORTO


# Mammo3d

## Aplicação técnicas aprendizagem automática no cancro da mama

**José Carlos Cordeiro Andrade Santos**

**Aluno nº: 1140921**

**Dissertação para obtenção do Grau de**

**Mestre em Engenharia de Inteligência Artificial**

**Orientador(a): Prof.ª Goreti Marreiros**

**Co-orientador: Prof. Pedro Pereira Rodrigues**

**Júri:**

Presidente: Doutor António Constantino Lopes Martins, Professor Adjunto do Instituto Superior de Engenharia do Instituto Politécnico do Porto

Vogais:

Doutor Victor Manuel Rodrigues Alves, Professor Auxiliar da Universidade do Minho

Doutora Maria Goreti Carvalho Marreiros, Professora Coordenadora com Agregação do Instituto Superior de Engenharia do Instituto Politécnico do Porto

Porto, Outubro 2022

# Resumo

O cancro da mama continua atualmente a ser um importante problema de saúde pública a nível internacional e nacional pelo que a problemática da sua abordagem continua a ter todo o interesse. Em Portugal, anualmente são detetados cerca de 7.000 novos casos de cancro da mama, e 1.800 mulheres morrem com esta doença.

De acordo com a Norma da Direção-Geral da Saúde para abordagem imagiológica da mama feminina, todas as mulheres assintomáticas com idade compreendida entre 50 e 69 anos, devem realizar uma mamografia de rastreio a cada dois anos. Na presença de alterações morfológicas ou em mulheres com risco moderado a elevado de cancro da mama, o médico assistente pode sugerir antecipar a realização da mamografia e complementar a investigação diagnóstica com os métodos que achar necessários. Se o cancro for detetado precocemente, a probabilidade de o tratamento ser eficaz e bem-sucedido é muito mais elevada.

A ressonância magnética é um exame de alta sensibilidade e especificidade moderada, sugerida em pacientes jovens, com aumento substancial do risco, i.e., que apresentam predisposição genética ou história familiar da doença. Este exame utiliza uma tecnologia à base de ondas de radiofrequência num forte campo magnético a fim de obter imagens mais detalhadas dos tecidos internos da mama, no entanto, o seu uso é limitado pela indisponibilidade (imediata) comparada com outros exames e preço associado e contraindicado em pessoas com claustrofobia, dispositivos metálicos como *pacemakers* ou próteses ou reações ao meio de contraste.

Assim, esta tese tem como objetivo desenvolver uma ferramenta de aprendizagem automática com recurso a Redes Adversariais Generativas Cíclicas, capaz de converter uma imagem de mamografia numa semelhante ao produto de uma ressonância magnética, com o intuito de proporcionar uma melhor perceção do campo cirúrgico e aumentar os ganhos em saúde.

O conjunto de dados foi cedido pelo Centro Hospitalar Universitário de São João e continha volumes de cortes transversais sucessivos de mamas. Neste caso, o corte seccional com área transversal máxima era o único com interesse para estudo, por isso, extraímos todas as localizações dos cortes para obter os cortes mediais respetivos das mamas.

As Redes Adversariais Generativas são pares de sistemas de Inteligência Artificial treinados para criar conteúdo e realizar tarefas mais rapidamente do que um único sistema. Nesta tese, estas realizam a tradução para uma imagem com base noutra singular não emparelhada, ou seja, uma imagem semelhante ao produto de uma ressonância magnética com base numa mamografia, sem imagem de ressonância magnética correspondente.

As ferramentas métricas de Medida do Índice de Similaridade Estrutural e de Relação Sinal-Ruído de Pico foram usadas para avaliar a qualidade da imagem sintetizada em relação à imagem real. Com o valor de 0.69667, o valor obtido pela medida do índice de similaridade estrutural indica alta similaridade da imagem criada com a de referência. Quanto à relação sinal-ruído de pico obtida de 31.805 dB, usada para quantificar a qualidade da imagem reconstruída a partir de uma imagem original que sofreu compressão, encontra-se dentro do intervalo de valores típicos.

Embora as ferramentas métricas forneçam um resultado quantitativo do desempenho, a melhor resposta que obtivemos foi visual. As imagens sintéticas obtidas apresentam uma aparência visualmente realista, embora seja possível detetar nestes alguns artefactos, devido à diferente forma de captação de imagem pelos diferentes exames e definição inferior dos exames originais usados como base em comparação com a ressonância magnética.

Em conclusão, a partir de um conjunto de dados com 57 imagens obtidas por mamografia, em perfil cefalo-caudal, foi possível gerar imagens sintéticas da estrutura mamária semelhantes ao produto da ressonância magnética baseadas em mamografia implementando e testando modelos de rede adversarial generativa, usando dados não emparelhados, como demonstrado pelas diversas métricas e verificações gráficas.

**Palavras-chave:** Mamografia, Ressonância Magnética, Redes Generativas Adversariais

# Abstract

Breast cancer has become a serious public health problem world-wide. According to European standards, all asymptomatic women aged 50 years and over should have breast cancer screenings, i.e., a mammogram. Magnetic resonance imaging produces extremely detailed pictures, facilitating a more accurate diagnosis. However, its' use is limited.

Thus, this thesis aims to implement an efficient deep learning method using Cycle-consistent Generative Adversarial Networks to synthesize artificial thoracic magnetic resonance imaging exams from mammograms, with the purpose of providing a better perception of the surgical field and increasing health gains.

The dataset was provided by Centro Hospitalar Universitário de São João and contained successive cross-sections of breast volumes.

Structured Similarity Indexing Method and Peak Signal to Noise-Ratio were the reference-based metrics used to evaluate the quality of the synthesized images. The structured similarity indexing method value rounds up to 0.69667, meaning the output images present high similarity to the reference one. The peak signal to noise-ratio equals 31.805 dB, which falls within the typical values' interval. Via perceptual study, we consider the output images to have a visually realistic appearance, when compared to real ones.

In conclusion, we were able to generate synthetic magnetic resonance-like images based on mammogram by implementing and testing generative adversarial network models, using unpaired data, as demonstrated by the several metrics and graphical checks. Although, it is worth noting that the product images have visually detectable artifacts and lack the definition of real exams.

**Keywords**: Mammogram, Magnetic Resonance Imaging, Generative Adversarial Networks

# Acknowledgments

Agradeço primeiramente à Prof.ª Goreti Marreiros, enquanto orientadora do presente trabalho, pela sua competência e todas as recomendações dadas.

Agradeço também ao Prof. Pedro Pereira Rodrigues, enquanto co-orientador do presente trabalho, a quem pretend, desde já, agradecer pela sugestão deste tema, que se revelou extremamente interessante, e que me deu uma grande satisfação desenvolver.

Agradeço também à Dr.ª Inês Moreira por ter cedido o conjunto de imagens a partir das quais foi possível elaborar este projeto, para além da disponibilidade em prestar esclarecimentos técnicos e científicos, cruciais para a elaboração e compreensão deste trabalho.

Por último, agradeço à minha namorada pelo apoio contínuo, amor incondicional, bem como pela sua colaboração a nível de conhecimentos e disponibilidade para construir esta tese.

A elaboração deste trabalho teve o contributo, direto e indireto, de muitos intervenientes que não poderiam ficar esquecidos e a quem expresso o meu profundo agradecimento. Finalizada esta etapa, quero agradecer a todos aqueles que me apoiaram durante todo o percurso porque, por detrás das realizações pessoais, escondem-se as contribuições, apoios, sugestões, comentários e críticas, vindos de muitas pessoas. A sua importância assume uma mais-valia tão preciosa que, sem elas, com toda a certeza, teria sido muito difícil chegar a este resultado.

# Content

# List of figures

# List of Tables

# List of Code snippets

# Abbreviations

| | |
|---|---|
| **BCT** | Breast-conserving therapy |
| **CC** | Cephalocaudal |
| **CHUSJ** | Centro Hospitalar Universitário de São João |
| **CT(s)** | Computer Tomography(ies) |
| **CycleGAN(s)** | Cycle Generative Adversarial Network(s) |
| **GAN(s)** | Generative Adversarial Network(s) |
| **LMO** | Lateral-medial oblique |
| **MG(s)** | Mammogram(s) |
| **MRI** | Magnetic Resonance Imaging |
| **MRi(s)** | Magnetic Resonance Image(s) |
| **PSNR** | Peak Signal to Noise-Ratio |
| **ROLL** | Radioguided Occult Lesion Localisation |
| **RSL** | Radioactive Seed Localisation |
| **sMRi(s)** | Synthetic Magnetic Resonance Image(s) |
| **SSIM** | Structured Similarity Indexing Method |

# 1 Introduction

This thesis project was conducted at Instituto Superior de Engenharia do Porto. Its goal was to investigate image-to-image translation using the deep learning method generative adversarial networks. This chapter introduces the background behind the problem we aim to tackle and enlightens about the purpose and motivation of the thesis.

## 1.1 Background of the Problem

Breast cancer, a relatively common type of cancer, even fatal, with higher incidence among women, has become a serious public health problem world-wide. [1] According to Portuguese Screening Guidelines for women, all asymptomatic women aged 50 years and older should be screened with a mammogram (MG) once every two years. Statistically, up to 25% of breast cancers are non-palpable at the diagnosis stage. [2]

There are several treatment options, and many patients even endure more than one type of treatment. A successful breast surgical treatment program relies on image guidance tools and the skills of a multidisciplinary team. The goal is to remove the lesion with safe and adequate surgical margins and prevent unnecessary reoperations. [3]

Nevertheless, using image guidance modalities as complementary tools helps surgeons to visualize the lesion within the breast in a more reliable way improve their performance and the outcome. [4]

While the MG is the only exam validated for breast cancer screening in the general population, a Magnetic Resonance (MR) Imaging is a highly sensitive technique that produces extremely detailed pictures, allowing the physician to make a more accurate diagnosis and facilitating a better treatment. [5] However, its use is limited for several reasons, thus it would be extremely convenient to obtain these high-detailed images while bypassing the MR.

## 1.2 Aims and Objectives

Thus, this thesis aims to develop a machine learning tool able to convert a MG to a Magnetic Resonance Image (MRi), serving as a guiding thread for a larger project, to be put into practice in the future, with the purpose of converting a two-dimensional image into a three-dimensional image that will allow a better perception of the surgical field to the overworked health professional and increase health gains by reducing women's hospitalizations following surgical interventions. [6]

To achieve such a feat, it's necessary to deconstruct the larger and more complex project into simpler concepts and isolated steps. Therefore, this thesis will contemplate the initial step: converting an original image generated by MG into a synthetic Magnetic Resonance Imaging (MRI)-like image, striving for it to be as equivalent and coincident as possible to an original one produced by MRI. The next step, not contemplated in this thesis, will overtake into the three-dimensional domain.

As the MRI offers excellent contrast resolution for soft tissues, the idea of generating MRis from other exams not as detail-oriented, such as a MG, for thorough tasks such as determining the exact size and location of the cancer is appealing. [7] At first glance, converting an image generated by a MG into one generated by MRI may seem more challenging than if performed the other way around, as MRI creates an image with greater contrast and detail. However, machine learning methods which proved to be quite competent in the treatment of non-linear data have already been studied, making this research proposal plausible.

The purpose of this research is to implement an efficient Deep Learning method using generative models to synthesize artificial thoracic MRI exams from MG, to overcome the data scarcity that has been delaying the clinical use of automated algorithms to detect breast lesions. Therefore, it will enrich the scientific community by providing a deeper analysis into the synthesized image metric for evaluation of the generative models.

However, as it is a recent development, there are no published articles yet that specifically report on converting a MG into an MRI-like image. Hence, this thesis will pave the way for a more detailed exploration of this concept and be an important asset if able to determine the feasibility of such hypothesis.

The following questions are addressed in this thesis:

> Can Generative Adversarial Networks (GANs) produce synthetic an MRI-like images based on MG?
>
> How visually realistic can the synthetic images turn out?
>
> How does the implemented GAN model perform when its' results are compared to ground truth data?

## 1.3 Structure & Layout

The presented document is organized into five chapters. The next instalment outlines crisply the context and challenges related to our research question, followed by a survey of scholarly knowledge on the topic. The methodology chapter highlights the philosophical underpinnings of the research and outlines the specific research design choices we've made. Afterwards, the resulting data is presented and described. The findings are, then, interpreted and linked to prior research, as well as to our own research questions.

# 2 Theory and related work

## 2.1 Background

Breast cancer is a neoplasm (abnormal and uncontrolled growth or proliferation of a certain tissue in the body) that originates in the breast tissue. [1]

### 2.1.1 Epidemiology

Breast cancer has become a serious public health problem world-wide. In 2020, there were 2.3 million women diagnosed with breast cancer and 685 000 deaths globally. As of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past 5 years, making it the world's most prevalent cancer. There are more lost disability-adjusted life years (DALYs) by women to breast cancer globally than any other type of cancer. [8]

Furthermore, it's predicted that the number of new cases each year will increase from 10 million in 2002 to 15 million in 2025, with 60% of these occurring in developing countries, where the incidence has increased by about 5% each year. [1] The incidence of breast cancer is higher in developed and developing countries, rather than in non-developed countries, and this may be associated with the higher average life expectancy of the former. [3]

Comparatively, in Portugal, there are around 7,000 new breast cancer cases and 1,800 deaths in Portugal every year. [9] Therefore, screening programs and women's awareness are crucial to reducing this mortality rate, leading to earlier detection of breast cancer and smaller lesions. [10]

### 2.1.2 Screening

According to Portuguese Breast Cancer Screening Guidelines for women, all asymptomatic women aged 50 years and older should be screened with a MG once every two years. In the presence of one or more symptoms, the health professional may recommend undergoing imaging methods earlier. The decision to start screening with MG in women prior to age 50 years should be an individual one, if these women place a higher value on the potential benefit than the potential harms (due to a genetic mutation or positive family history of breast or ovarian cancer, etc.). For women aged 69 years and older, screening with a MG is recommended once every two to three years. [1]

Statistically, up to 25% of breast cancers are non-palpable at the diagnosis stage (size smaller than 10 mm). In developed countries, less than 10% of breast cancers are diagnosed in advanced stages of the disease, contributing to the fact that most patients (about 80%) are treated by complete surgical removal. [3]

### 2.1.3 Management

There are several treatment options such as surgery, chemotherapy, radiation therapy, hormone therapy and targeted therapies and many patients even endure more than one type of treatment. In most cases, the imperative factor in choosing the treatment is the stage of the disease.

Breast-conserving therapy (BCT) refers to breast-conserving surgery, that is, tumor excision while leaving as much normal breast as possible, typically followed by moderate-dose radiation therapy to eradicate any microscopic residual disease and has become a standard treatment option for women with early-stage invasive breast cancer. Long-term survival following BCT is equivalent to mastectomy, providing also an acceptable cosmetic outcome and low morbidity. [3] The preoperative localization of these non-palpable lesions guided by breast imaging is an important and required procedure for BCT. [11]

A successful breast surgical treatment program relies on image guidance tools and skills of a multidisciplinary team between the surgeon, the radiologist, and other specialists. The goal is to remove the lesion with safe and adequate surgical margins, prevent unnecessary resection of healthy tissue and provide a good aesthetic outcome. [12]

Nonetheless, there are intrinsic variables to these image guidance tools beyond the radiologist's control and expertise, which will have impact on the outcome. We must take notice of the relative position of the patient during the different procedures: while performing an MRI, the patient is lying face-down (ventral decubitus) whereas, during the surgical procedure, the patient is lying on their back (dorsal decubitus); furthermore, another crucial detail is the deformation of the breast, integral to the routine of the MG, this one done while the patient is standing (orthostatic position). Since the breast is highly pliable, its manipulation, either

gravitational or intentional, will produce different representations of the soft tissue architecture, making it even more difficult to accurately determine the lesion's localization.

### 2.1.4 Preoperative localisation techniques

Since the 1970s, preoperative localizations for breast lesions and axillary lymph nodes have been achieved by the placement of a tiny wire which guides the surgeon to the correct area in the breast during the operation. But, given the guide-wire limitations, other techniques were developed and emerged as potential alternatives. [13]

While the wire remains a safe and tested technique, Radioguided Occult Lesion Localisation (ROLL) and Radioactive Seed Localisation (RSL) could be offered as a reliable alternative. Since an involved margin mostly leads to a reoperation, both are additionally related outcomes of the intervention. When comparing to wire, the excision of a lesion can be more effective using RSL and ROLL, with less involved margins and, therefore, fewer reoperations. In general, recent techniques seem to be potential alternatives to guide-wire, due to the quick learning curve, better scheduling, and management issues. [4]

On the other hand, radioactive material uses for RSL, and ROLL techniques require specific institutional protocols for recovery and disposal and all personnel involved with handling must have radiation safety training. To overcome this problem, most recent techniques use non-radioactive material but seem to be MRI conditional, which limits their use. [4]

Nevertheless, using these localization techniques as complementary tools helps surgeons to visualize the lesion within the breast in a more reliable way and therefore improving their performance and the outcomes described above.

### 2.1.5 Breast Imaging Tests

These procedures can be performed under ultrasound (US) guidance, stereotactic or MRI and, less frequently, through Computer Tomography (CT). Thus, the radiologist selects the image guidance modality according to the best lesion visibility and patient's body features. [11]

Breast imaging is one of the keystones in the diagnosis of breast pathologies, facilitating a better diagnosis and consequent treatment of the patient. [10]

A MG is a low-dose x-ray done with the woman standing (orthostatic position) which can evaluate the breast parenchyma, namely, to detect masses or clusters of microcalcifications, while the breast is being flattened between two plates. [11] Being the only exam validated for breast cancer screening in the general population, its goal is to diagnose subclinical tumors, that is, tumors that are not palpable and do not show other signs or symptoms. Considering the similar radiological behaviors of the breast components, the technique presents high sensitivity assessing the size, location, and characteristics of small nodules, in cephalocaudal and lateral-medial oblique views. [11]

Breast ultrasound (US) is a non-invasive imaging method that uses high-frequency sound waves and their echoes to produce pictures of the internal structures of the breast. [14] The US can be used as a complementary resource in cases of high tissue density or breast implants, but also to complete the mammographic study when the latter is not recommended or when the results are not elucidative enough. Although considered as unreliable to detect lesions smaller than 1 cm, in the case of malignant lesions screened with a MG, an ultrasound is performed to characterize the tumor extension and to assess the presence of multifocality and lymph node involvement.

MRI is a highly sensitive technique that uses a powerful magnetic field and radio waves, thus considered harmless by using non-ionizing radiation, to produce extremely detailed pictures, between 10 to 100 times greater than that obtained by X-rays [15], allowing the physician to make a more accurate diagnosis. It is particularly effective in detecting occult tumors, assessing multifocality and multicentricity, distinguishing between scar tissue and a recurrent mass, assessing the cancer's response to neoadjuvant chemotherapy and staging. The use of breast MRI is indicated in younger women or women with a high breast tissue density, that is, in whom there is a significantly higher amount of glandular tissue in the conventional mammographic image, otherwise reducing the number of false positives that would erroneously increase the rate of detection of breast cancer. [4]

However, the use of MRI is limited because of its high cost and significant operating and processing time. More importantly, MRI is contraindicated for some patients with claustrophobia or cardiac pacemakers due to the possibility of injury. In contrast, CT scans are much less expensive, are faster, and do not face the same limitations.

A CT scan uses multiple X-rays, taken at different angles to produce cross-sectional images of an area such as internal organs, fractures, or head trauma. This test often involves injection of iodine-based contrast materials intravenously to enhance the images. It's not used routinely to evaluate the breast. During a breast CT, the patient must lie face down on a table while a CT scanner rotates around the breast. The total dose of radiation is the same as in a conventional MG. [16]

## 2.2 State of Art

The recent and constant exponential growth of machine learning has great potential and provides countless opportunities for the creation and development of automatic, accurate and consistent responses to the infinite unknowns of today's society.

As it is an innovative concept, it arouses the curiosity of experts to reassess the imaging methods currently used for diagnosis. However, as it is a recent development, the existing and available literature on the subject is scarce. Since there are no published articles that specifically report on converting a MG into an MRi, the literature review was expanded to include projects whose goal described modelling a synthetic MRI-like image (sMRi) after an image generated by CT scan.

This thesis presents a qualitative literature review. In order to define the inclusion and exclusion criteria for studies with effects on the constitution of the sample, we considered the research question previously formulated. Thus, we included experimental studies carried out in humans without limitation of age group, with discrimination of the diagnostic imaging technique used and which reported on converting medical images using neural network.

The review of the literature was carried out in several databases such as IEEE xPLore,Google School, Sciencedirect and Springer Nature with the query (mammogram and synthetic MR and artificial intelligence or GAN or CycleGAN or adversarial neural network or Deep Learning or Medical Imaging), without language restrictions. The review includes studies mainly from the year 2019, dating the most recent one to 2020, considering it being a very recent subject. We also checked gray literature and preprint databases, with no results available. References of included studies were checked for additional relevant studies.

### 2.2.1 Generating Lumbar Spine MRi from CT scan

Cheng-Bin et al. proposed a method for estimating lumbar spine MRis based on CT images using a novel objective function and a dual Cycle-consistent Generative Adversarial Network (CycleGAN) with semi-supervised learning ($DC^2$Anet). [17]

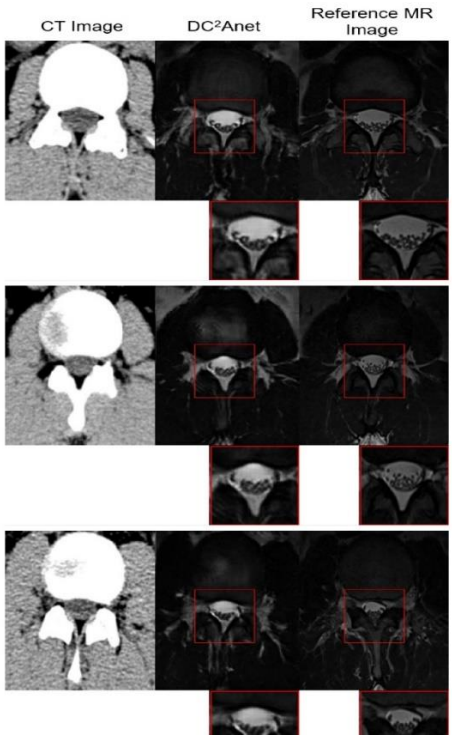


Figure 1 - sMRis of the lumbar spine.

Figure 1 presents three examples of synthetic images produced by the proposed $DC^2$Anet method, alongside the corresponding CTs and MRis. From left to right: input CT image, the

proposed DC$^2$Anet and the reference MRi. The spinal cord region in the central area of the image is enlarged to evaluate the reconstruction capability.

DC$^2$Anet learned to differentiate between different structures with similar intensity values, such as a vertebra, fat tissue and disc signals, in CT images but not in MRis. The disc signal and thecal sac in the sMRi looked like the MRi of reference. However, the proposed method exhibited limitations in the reconstruction of the degree of disc protrusion and the degree of stenosis.

### 2.2.2 Spine CT to MRi Synthesis

GANs were trained to transform spine CT image slices into spine Magnetic Resonance T2 weighted (MRT2) axial image slices by combining adversarial loss and voxel-wise loss. Experiments were performed using 280 pairs of lumbar spine CT scans and MRT2 images. Jung-Hwan Lee et al. excluded CTs and MRis of severe lumbar spine pathologies, such as tumor, infection, or fracture, although they included images of degenerative diseases. [18]

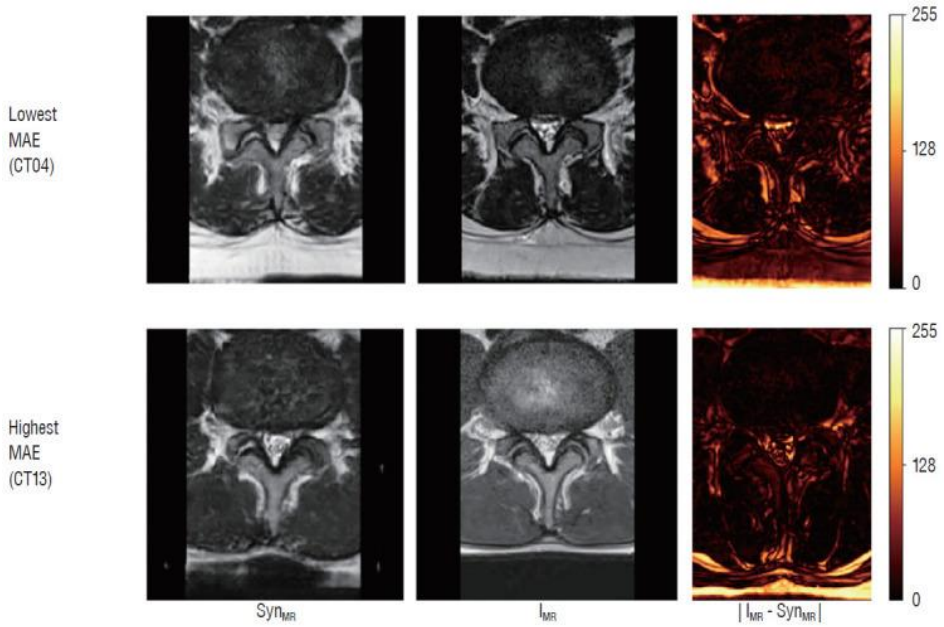


Figure 2 - sMRis of the lumbar spine.

Figure 2 shows examples of mean absolute error (MAE) between the real ($I_{MR}$) and synthesized ($Syn_{MR}$) MRis.

### 2.2.3 Transform 2D brain CT image slices into 2D brain MRi slices

Cheng-Bin et al. proposed a synthesis method based on generative adversarial networks (GANs) to estimate an MRi from a CT image. For that, they required both the CT scans and the corresponding MRis to estimate the MRI-like image, which they refer to as paired data.

However, obtaining these is challenging, since patients had to undergo both a CT scan and an MRI exam within a short interval of time. As for single CT or MRI data, these were referred to as unpaired data.

The proposed approach uses paired and unpaired data to overcome the context-misalignment issue of unpaired training, and to alleviate the registration and blurry results of paired training. Therefore, the MRI-GAN had two structures to simultaneously train different data.

Figure 3 depicts from (left) to (right): The input CT scan ($I_{CT}$), the sMRi ($Syn_{MR}(I_{CT})$), the MRi used as reference ($I_{MR}$), and the absolute error between the reference and synthesized MRis($|Syn_{MR}(I_{CT}) - I_{MR}|$). The results of the MRI-GAN using paired and unpaired data together indicate a better performance according to the quantitative and qualitative evaluations, rather than using one type of data only. [19]

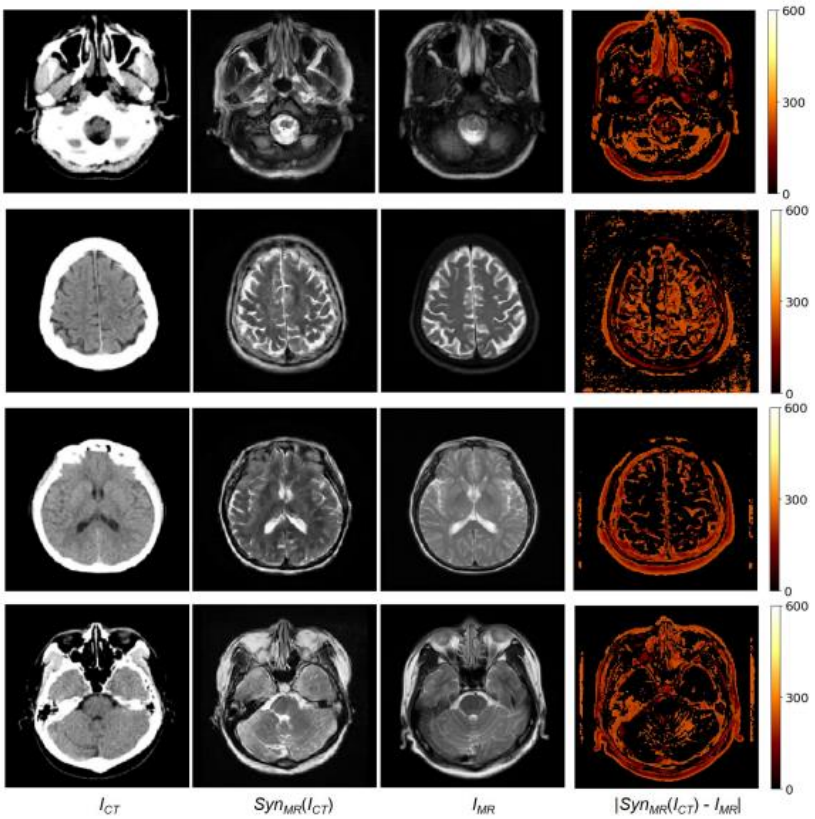


Figure 3 - sMRi of the brain.

Accurate tumor segmentation is a requirement for MRI-based radiotherapy but is currently labor-intensive and observer-dependent.

### 2.2.4 Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumour segmentation

Jiang et al. aimed to compute a robust deep learning model for segmenting tumors from MRis. Nevertheless, the lack of large expert annotated MRI datasets makes training deep learning

models difficult. Therefore, a cross-modality (CT-MRI) deep learning segmentation approach that augments training data using sMRis produced by transforming expert-segmented CT images was developed.

To overcome the issues of the existing methods, which couldn't accurately model the anatomical characteristics of atypical structures, including tumors, a tumor-attention loss that regularizes the GAN model and produces sMRis with well-preserved tumor structures was introduced. [20]

In this study, 81 T2-weighted MRI scans from patients with non-small cell lung cancers were analyzed. This model augmented training data arising from 6 expert-segmented T2w MR patient scans with 377 sMRis originated from non-small cell lung cancer CT patient scans obtained from the Cancer Imaging Archive. This approach produced highly similar estimations of tumor growth as an expert.

Figure 4 illustrates example sMRis generated from a representative CT image (A) using (C) the state-of-the-art CycleGAN, (D) the Unsupervised Image-to-Image Translation (UNIT) adaptation method and (E) the proposed approach. The corresponding T2w MRi, acquired within a week of the CT image, is also shown alongside for comparison (B).

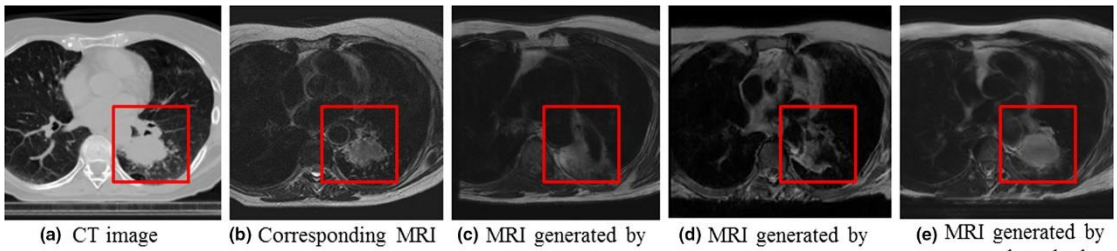


Figure 4 - sMRis of the lungs.

### 2.2.5 sMRI-aided multi-organ segmentation on male pelvic CT

Dong et al. obtained sMRis by CycleGANs to offer superior soft-tissue information for male pelvic CT images.

A deep attention strategy was integrated into the Dual pyramid networks (DPNs) to select and extract the most relevant features from both the CTs and the sMRis to identify the organs' boundaries.

The proposed method was trained and evaluated using datasets from 140 patients with prostate cancer and were then compared against state-of-art methods. [21]

Figure 5 juxtaposes the proposed method on three additional patients. Column (1) shows the CT image, (2) the sMRi, (3) the manual contour of the bladder (white), the prostate (yellow) and

the rectum (red), and (6) the multi-organ segmentation by DAUnet trained on sMRi data. Lines (a), (b) and (c) exhibit the segmentation results on three different patients.

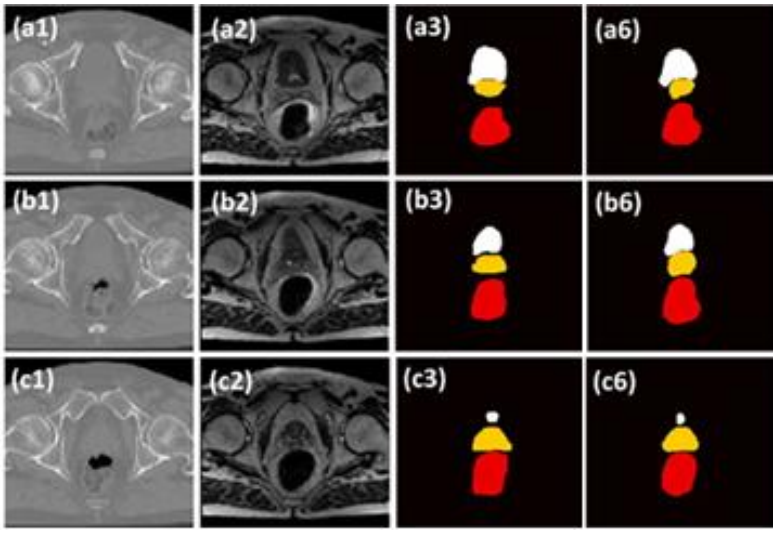


Figure 5 - sMRis of the male pelvis and sMRI-aided multi-organ segmentation.

# 3 Materials and Methods

The Dataset used in this work was provided by Centro Hospitalar Universitário de São João (CHUSJ), after being previously authorized and approved by the Hospital's Committee of Ethics to be used on the project mammo3d [22]. The Dataset consist of two type of views, craniocaudal (CC) and mediolateral oblique (MLO). Both views are very different and, since MLO view contains part of the arm pit muscle which could become a constraint to the method, this last view was discarded. Due to a low paired image count between MGs and MRis, and since CycleGAN does not require those images to be paired, we chose two separate sources for our MG and MR images. The dataset used for MGs is CBIS-DDSM [22], containing 50 images, and the one used for MRis is QIN [16], with a total of 836 images. These datasets contain 3D volumes of several breasts but, for computational power constraint, we will work only on one specific slice of the breast for the MRi. Figure 6 shows an example of an extracted slice for both modalities.

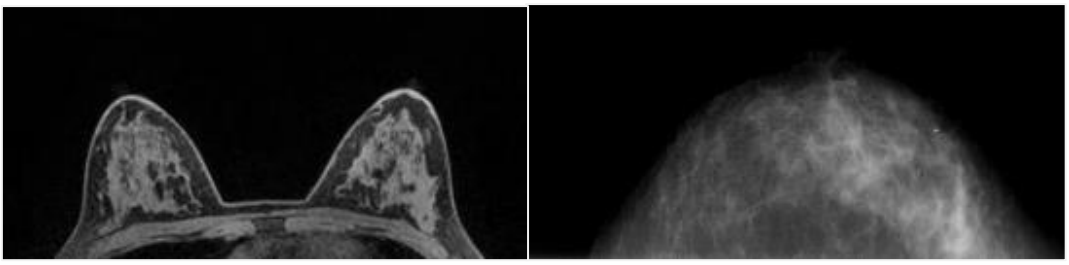

Figure 6 - Examples of images resulting of an MRI exam (left) and a MG (right).

In terms of pre-processing, some crucial steps were undertaken. The single images generated by MRI, also called slices, are successive horizontal cross-sections of the breast, whereas to our

process only the slice with the maximum cross-sectional area mattered (i.e, with the longer distance from the edge of the "semisphere" to the center), routinely called middle of the breast.

Therefore, we resorted to dicom library and extracted all the slices' locations to obtain the middle slice of the breast. Afterwards, the original slice was cropped into two separate images (left and right breasts) and then resized to 256x256.

## 3.1 Implementation details

CycleGAN is an extension of the Generative Adversarial Network architecture that performs image-to-image translation with unpaired images, with unpaired meaning single MG data with no corresponding MRi.

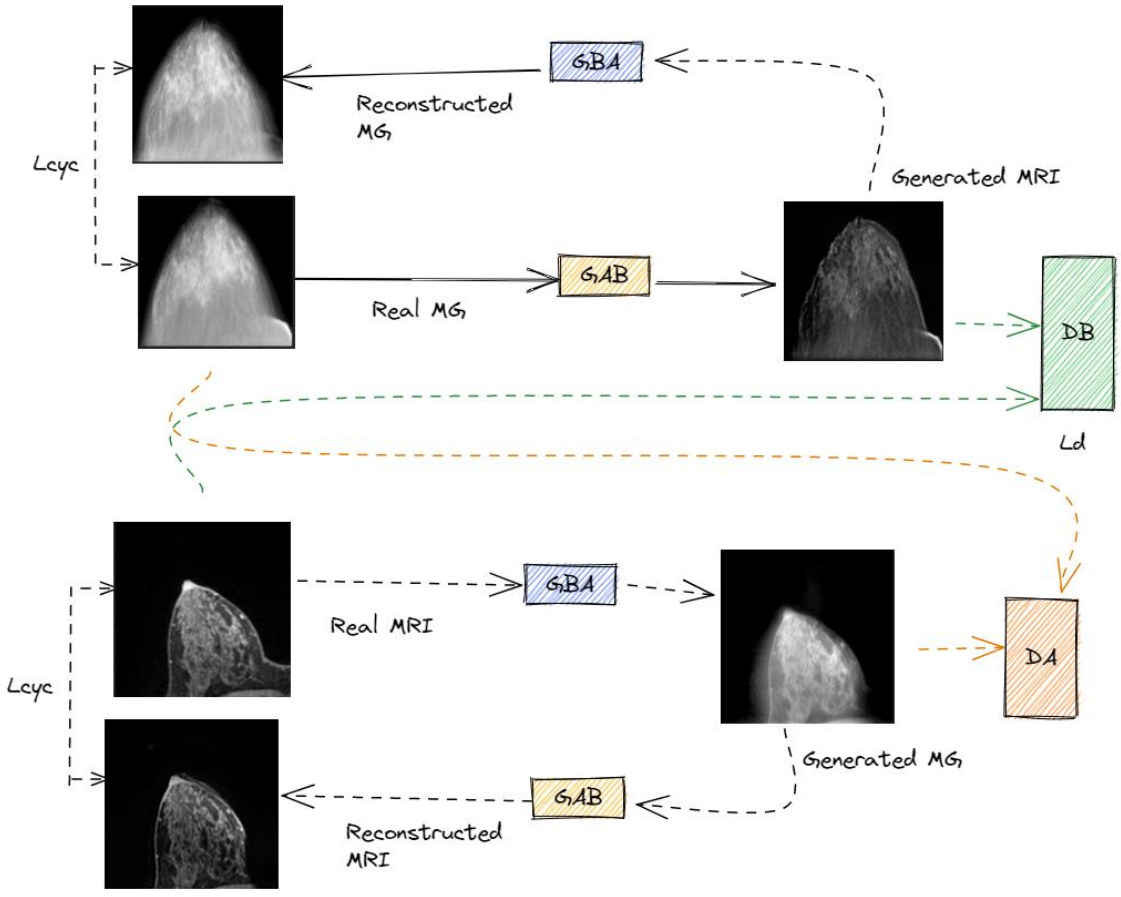


Figure 7 - CycleGAN implementation Overview.

CycleGAN is fed two domains X and Y, given unpaired training samples x ∈ X and y ∈ Y. [23] While the domain A is composed only of MG breast images, the images in domain B are exclusively breast MRi.

The mapping functions G : X → Y and F : Y → X are called generators. As shown in Figure 7, a CycleGAN framework for the synthesis of a breast MRi includes two separate generators: one

forward generator from MG images to MRis (GAB), and one backward generator from MRis to MG images (GBA).

For the generators, we used U-NET, a convolutional neural network architecture. Since the resolution of images required by U-NET was 256 × 256 pixels, our original images were modified to fit the standard, as explained above. In order to increase the training speed, we also enriched our generators by adding pre-trained U-NET networks. The pre-trained U-NET models were provided by the segmentation-model python library [24].

```
1. def generator(baseline = 'resnet18', weight = 'imagenet', batchnorm = True,
   activation = nn.ReLU):
2. return sm.Unet(baseline, encoder_weights = weight, decoder_use_batchnorm =
   batchnorm, activation = activation, in_channels=1)
```

Code snippet 1 – Pretrained generators with U-NET network.

Then, two discriminators, DX and DY, aim to distinguish between the real and the created images. While $D$A is the discriminator that differentiates generated and real MG images, $D$B is the discriminator that differentiates the synthesized from the real MRis. In order to increase the training speed, we also enriched our discriminators by adding pre-trained data.

The loss computation used for the discriminators is the binary cross entropy loss, Lcyc, which reports how accurately the discriminator can differentiate the generated from the real images. During the several attempts, the brand-new images generated were quite similar amongst them. To bypass this, a new loss tool was added, i.e., Ld, a pretrained features extractor (VGG19) with three layers.

```
vgg19 = torchvision.models.vgg19(pretrained = True)

first_features = nn.Sequential(*list(vgg19.children())[0][:6]).cuda()
second_features = nn.Sequential(*list(vgg19.children())[0][:15]).cuda()
third_features = nn.Sequential(*list(vgg19.children())[0][:26]).cuda()
first_features.eval()
second_features.eval()
third_features.eval()
first_features.requires_grad_(False)
second_features.requires_grad_(False)
third_features.requires_grad_(False)
feature_extractors = [first_features, second_features, third_features]
feature_weights = [0.5, 0.5, 0.5]
```

Code snippet 2 – VGG19 extractor 3 layers.

VGG model basically follows the Convolution Neural Network Architecture. This structure is a successor of CNN architecture that was built to increase performance [25].

Multiple trials were done, while training the CycleGAN. We first started with 100 epochs and noticed that the learning rate remained constant throughout the succeeding trials, even though it was supposed to be incrementing. Only with 200 epochs were we able to finally start seeing

results. Since heavy computation was required for the algorithm to run, we used GoogleColab to tackle this hazard.

# 4 Results and discussion

After attempting different hyperparameters, we stored the model, that was built upon the whole dataset, which obtained some acceptable visual results:

- Learning ratio k: adds up the number of iterations that the generator was trained for before training the discriminator for the one step. The generator tries to find the best image to fool the discriminator, but not too quickly, otherwise there is no feedback to be taken from the discriminator. If the discriminator gets too good too quickly, then the generator is never able to converge.
- Learning rates: With two different learning rates, both the generator and the discriminator were tweaked accordingly till their individual performances were optimal. During our trials, we obtained the best results by setting a $1e^{-4}$ learning rate for the generator and $1e^{-3}$ learning rate for the discriminator.
- Loss weights: The aim of the cycle perpetual loss was to avoid mode collapse, which happened when we used the cycle consistency loss on its' own.

Following that, we intended to juxtapose these results with other studies' results, even though none reported on converting a MG to and MRi. Therefore, since the main purpose is to understand the difference between these synthetic images and the real ones, we focused on projects which aimed to generate synthetic images by CycleGAN also exclusively using unpaired data, to compare its results to ours.

## 4.1 Evaluation

Choosing a decent metric to determine how well a GAN works against another one is particularly challenging since there is not a specific way to do it. In some cases, validating the resulting images by visual inspection could be sufficient to access if a model seems to work efficiently. However, in the case of a generative algorithm in the medical domain, we are not proficient at judging if the generated image has retained the key features a health professional

would hypothetically need to continue their clinical reasoning. Moreover, since the algorithm does not consider paired images, we do not have the real MG image corresponding to the MRi.

Structured Similarity Indexing Method (SSIM) and Peak Signal to noise-ratio (PSNR) were the reference-based metrics used in this thesis to evaluate the quality of a synthesized image against the real image.

SSIM is a perception-based model that considers image degradation as the perceived change in structural information, while also incorporating important perceptual phenomena including both luminance masking and contrast masking terms. [26] It's a method of image quality assessment and reflects the structural similarity between two images. Many projects have used the SSIM for solving problems in several fields, such as image compression, image restoration and pattern recognition. The resultant SSIM index is a decimal value between -1 and 1, where 1 indicates perfect similarity, 0 indicates no similarity, and -1 indicates perfect anti-correlation.

PSNR block computes the peak signal-to-noise ratio, in decibels, between two images. PSNR is most commonly used to measure the quality of reconstruction of lossy compression codecs (e.g., for image compression). The signal in this case is the original data, and the noise is the error introduced by compression. Typical values for the PSNR in lossy image compression are between 30 and 50 dB, where higher is better. [27]

Table 1 details quantitative evaluations of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) data across all brand-new images, synthesized using CycleGAN.

Table 1 - Quantitative comparison of the image quality analysis.

| AUTHOR | METHOD | SSIM | PSNR |
|---|---|---|---|
| | CycleGAN | 0.69667 | 31.805 |
| **HUIXIAN ZHANG ET AL** | CycleGAN | 0.747 | 31.671 |

Although metrics provide a quantitative result of the performance of our CycleGAN, the best feedback we had during the hype parameterization was visually, since it allowed us to see if the transformations were similar to the original image. Every 4 epochs we plotted the image to check on the progress. The images below show the results on the validation set after 100, 150 and 200 epochs of training.

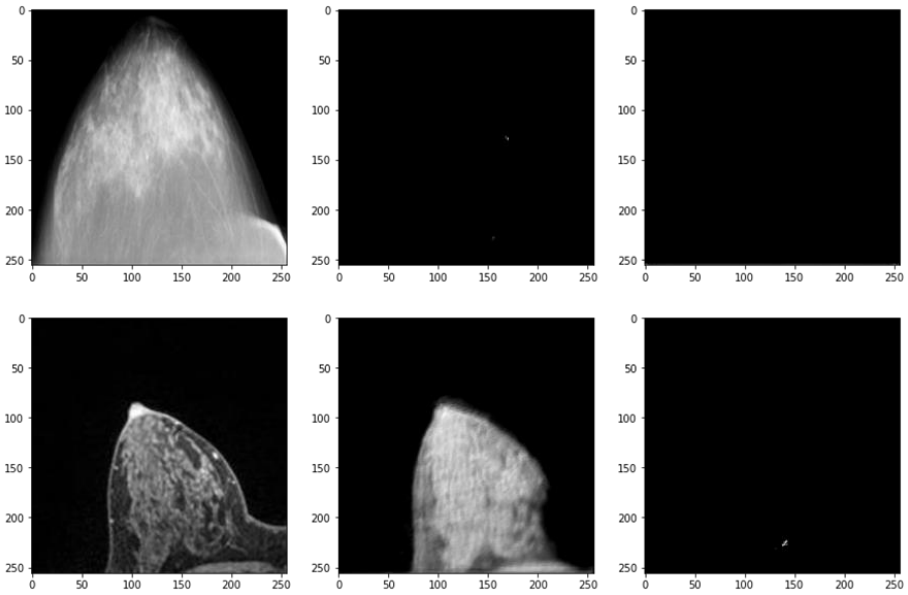

Figure 8 - Results on epoch 100.

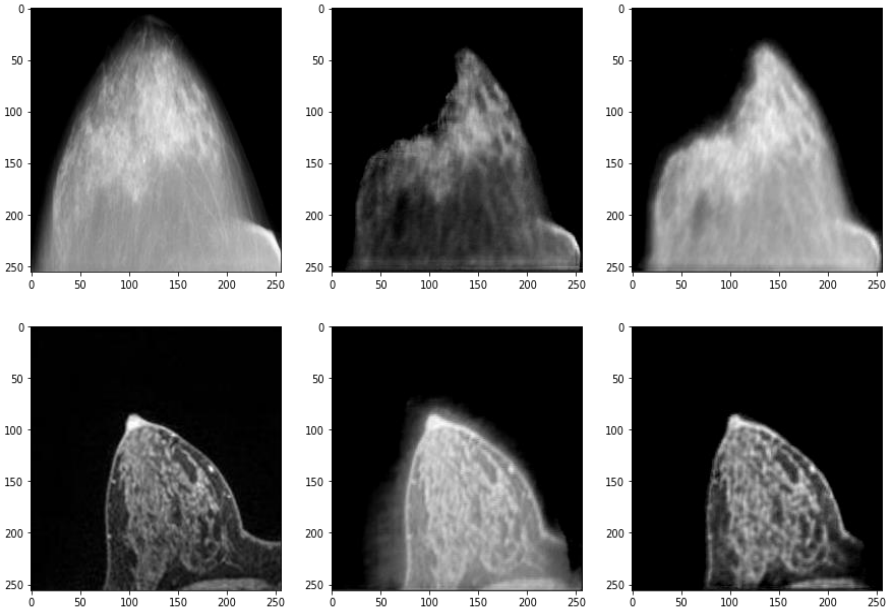

Figure 9 - Results on epoch 150.

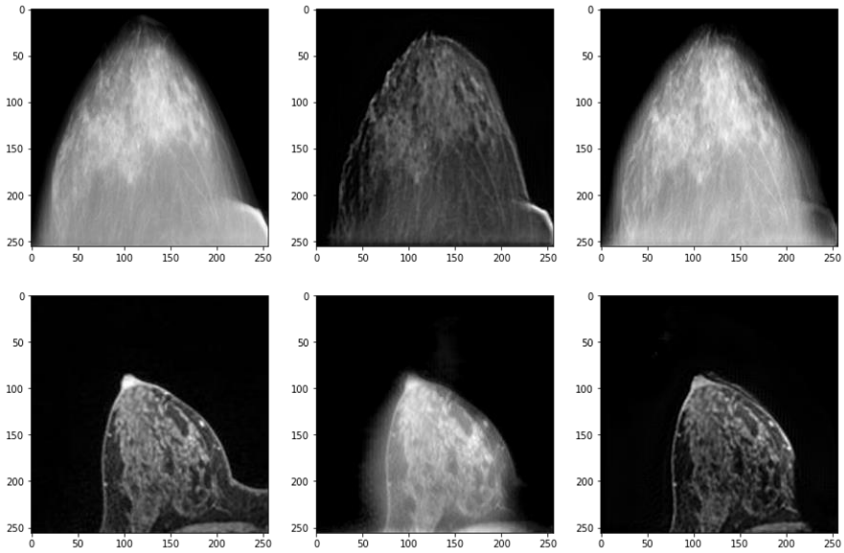

Figure 10 - Results on epoch 200.

The choice of implementing and testing two GAN models was made early on, during the review of the literature related to the research question (on Chapter 2.2), since it had been proved to be the best strategy to synthesize realistic images.

## 4.2 Visualization

In figure 6, we juxtapose the results generated by CycleGAN. At first sight, it is worth pointing out the distinct size between both images, which can be explained by the compression factor intrinsic to the breast MG technique. On the other hand, we can identify details on sMRi as preserved from the original MRi.

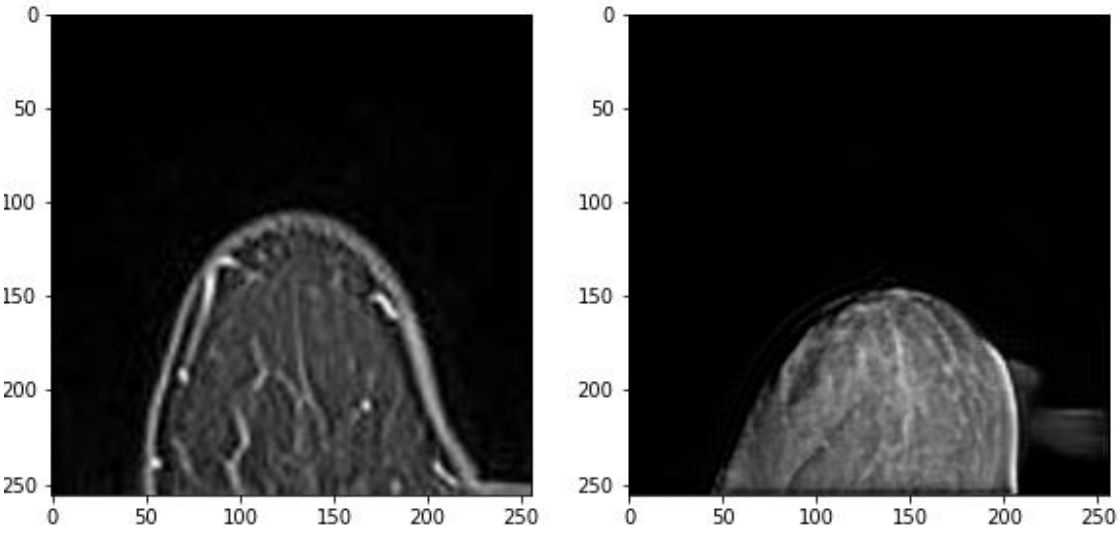

Figure 11 - Original breast MRi and sMRi, respectively.

## 4.3 Research Findings

Can GANs be used to generate synthetic images from data provided? Synthetic images can be generated.

1. Can Generative Adversarial Networks (GANs) produce synthetic an MRI-like images based on MG?

After attempting different hyperparameters, the model was able to generate synthetic MRI-like images based on MG by implementing and testing two GAN models, using unpaired data, providing adequate results as demonstrated by the several metrics and graphical checks.

2. How visually realistic can the synthetic images turn out?

Based on each epoch and individual breast image, different sets of results are obtained from an original image, besides that several slices present with artifacts, which multiplies variability. Nonetheless, the output sMRis have a visually realistic appearance, when compared to real MRis. Although, it is worth noting that this was proven via a perceptual study and the concept of realistic appearance in images is subjective, therefore it is difficult to specify how realistic the attained results are.

3. How does the implemented GAN model perform when its' results are compared to ground truth data?

SSIM is a method for predicting the perceived quality of pictures, measuring the similarity between two images. Being that our SSIM value rounds up to 0.69667, where 1 indicates perfect similarity and 0 no similarity, the output images present high similarity to the reference one.

PSNR is commonly used to quantify reconstruction quality for images subject to lossy compression. Typical values for the PSNR are between 30 and 50 dB, where higher is better. With our PSNR rounding up to 31.805 dB, it falls within the aforementioned interval.

## 4.4 Limitations

As a delimitation of this thesis, we identify the lack of optimization of the hyperparameters, while investigating the models, focusing instead on architectural changes and model modifications. Because of the factor of time limitation, we made the conscious choice to exclude hyperparameter searches and rely on the suggested settings from the original implementations. If, for example, the hyperparameter grid searches had been included for the different models, better results might have been obtained.

Furthermore, the small size of the provided dataset can also be a limitation to the obtained results. According to [28], one of the biggest dilemmas on Artificial intelligence Training Breakthrough is, in fact, the small data. Since the discriminator coaches the generator, giving

pixel-by-pixel feedback to help it improve the realism of its synthetic images, it typically takes 50,000 to 100,000 training images to train a high-quality GAN. [28]

Another limitation of learning-based methods is how unpredictable performance can be, when applied to datasets that are very different from the training sets. These differences may be attributed to unusual or abnormal anatomy or images with degraded quality due to severe artifacts and noise.

The existing and available literature on the subject is insufficient which became a challenge, leading us to investigate a similar thought process, i.e. generating synthetic images by CycleGAN also exclusively using unpaired data, but on a different body part. We chose to adapt the same metrics (SSIM and PSNR) that Huixian Zhang, et al., used to evaluate the quality of a synthesized image to facilitate the comparison of results. Nonetheless, we must conjecture that a bias may have been introduced by comparing the brain, an organ that serves as the center of the nervous system, whose main component is nervous tissue, with the breast, that is mainly adipose tissue.

# 5 Conclusion

The use of deep neural networks with two pairs of generators and discriminators makes the CycleGAN training often difficult, since the deep neural networks should be training data simultaneously, whereas a generator must be good at fabricating data and a discriminator must be good at identifying these data. Ultimately, we must find an unstable equilibrium between these two objectives.

At the end, the model provides adequate results as demonstrated by the several metrics and graphical checks. The process of training a CycleGAN is also a time-consuming one, requiring a lot of computation power. Limiting the training attempts on previously identified breast slices and performing model and parameters selection on a restricted amount of data were vital to the outcome, seeing as the computation on the full dataset for 200 epochs already took 8 hours on Google Colab.

## 5.1 Future line of work

An alternative strategy yet to be explored could be to implement even more models and compare the results amongst them. These other models could, hypothetically, synthesize more visually realistic images than the implemented ones, but the drawback would be a lower number of versions of the CycleGAN model would have been implemented due to time limitation.

This dissertation has shown that, although we were able to generate breast MRis from MG images, the synthetic images have visually detectable artifacts and lack the definition of real exams. To overcome this issue, and therefore improve the quality of the image and the precision of the localization of the tumor, Convolutional Neural Networks (CNNs) should be used for multi-class image classification and segmentation of Digital Mammogram. Once the data is acquired, the project would be to generate three-dimensional (3D) random

anthropomorphic voxelized phantoms of the human female breast using the Virtual Imaging Clinical Trials for Regulatory Evaluation (VICTRE) project of the Food and Drug Administration (FDA) software tool. Using this tool, realistic anatomical structures can be generated by specifying different virtual-patient characteristics that include breast type, shape, granularity, density, and size obtained previously. [10]

# Bibliography


[1] “cancro-online,” cancro-online, 05 2020. [Online]. Available: https://www.cancro-online.pt/cancro-da-mama/informacao-basica/o-que-e-o-cancro-da-mama/.

[2] Direção Geral da Saúde., “Abordagem Imagiológica da Mama Feminina,” 2011.

[3] Atkins, et al. “Positive margin rates following breast-conserving surgery for stage I-III breast cancer: palpable versus nonpalpable tumors,” *Positive margin rates following breast-conserving surgery for stage I-III breast cancer: palpable versus nonpalpable tumors,* 2012.

[4] I. C. e. a. Moreira, “Preoperative localisation techniques in breast conservative surgery: A systematic review and meta-analysis,” *Preoperative localisation techniques in breast conservative surgery: A systematic review and meta-analysis,* 12 2020.

[5] Baert A, et al. “Digital Mammography. Berlin: Springer,” 2010.

[6] J. Sá, “Serviços de Urgência e Segurança Clínica,” *Sociedade Portuguesa de Medicina Interna,* 2021.

[7] P. Metcalfe, et al. “The Potential for an Enhanced Role for MRI in Radiation-Therapy Treatment Planning,” *The Potential for an Enhanced Role for MRI in Radiation-Therapy Treatment Planning,* p. 429–446, 2013.

[8] “https://www.who.int/news-room/fact-sheets/detail/breast-cancer,” 26 3 2021. [Online]. [Accessed 22 07 2022].

[9] “https://www.ligacontracancro.pt/cancro-da-mama/,” [Online]. [Accessed 24 07 2022].

[10] M. ldo Badano, et al. “Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial,” *Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial,* 2018.

[11] N. Perry, et al. “European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis,” *European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis,* 2006.

[12] N. Perry. *Multi-disciplinary aspects of quality assurance in the diagnosis of breast disease.*

[13] B.K.Y. Chan, et al. “Localization techniques for guided surgical excision of non-palpable breast lesions”.*Localization techniques for guided surgical excision of non-palpable breast lesions.*

[14] “World Cancer Research Fund/American Institute for Cancer Research, Diet, nutrition, physical activity and breast cancer,” World Cancer Research Fund, 2018.

[15] L. J.P, “Técnicas de Diagnóstico com raios X. Aspectos Físicos e Biofísicos”.*Técnicas de Diagnóstico com raios X. Aspectos Físicos e Biofísicos.*

[16] “cancerimagingarchive,” [Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/QIN+Breast+DCE-MRI.

[17] J. I. L. ,. a. Cheng-Bin Jin,et al. “DC2Anet: Generating Lumbar Spine MR Images from CT Scan Data Based on Semi-Supervised Learning,” *Web of Science,* p. 24, 2019.

[18] K. Jung Hwan Lee,et al. “Spine Computed Tomography to Magnetic Resonance Image Synthesis Using Generative Adversarial Networks : A Preliminary Study,” *The National Center for Biotechnology Information,* 2020.

[19] W. J. ,. a. Cheng-Bin Jin, et al. “Deep CT to MR Synthesis Using Paired and Unpaired Data,” *Multidisciplinary Digital Publishing Institute,* p. 19, 2019.

[20] D. V. Jue Jiang,et al. “Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumor segmentation from small MR datasets,” *American association,* p. 13, 2019.

[21] P. a. Xue Dong, et al. “Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network,” *sciencedirect,* p. 10, 2019.

[22] “cancerimagingarchive,” [Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM. [Accessed 12 04 2022].

[23] a. Zhou Wang, et al. “Image quality assessment: from error visibility to structural similarity.,” in *IEEE*.

[24] P. Yakubovskiy, “Segmentation Models Pytorch,” GitHub, 2020. [Online]. Available: https://github.com/qubvel/segmentation_models.pytorch.

[25] R. M. T. Abed, et al. “"Alzheimer's Disease Prediction Using Convolutional Neural Network,” 2020.

[26] a. Z. Wang,et al. “Image quality assessment: From error visibility to structural similarity,” *Image quality assessment: From error visibility to structural similarity,* pp. 600-612, 2004.

[27] MathWorks, “mathworks,” [Online]. Available: https://www.mathworks.com/help/vision/ref/psnr.html.

[28] I. SALIAN, “NVIDIA Research Achieves AI Training Breakthrough Using Limited Datasets”.

[29] “World Health Organization. Guidelines for management of breast cancer,” pp. 11-56, 2006.

[30] “Groot M. et al. Costs and Health Effects of Breast Cancer Interventions in Epidemiologically,” *2006,* pp. 81-90.

[31] “Barber MD.; Thomas JST, et al.” *Clinical Publishing,* 2008.

[32] M. Ferlay J,et al. “IARC,” in *(ECN and EUNICE projects, 2007)*, 2008.

[33] M. C., “Determinantes Genéticos no Cancro da Mama,” APAMCM. [Online]. [Accessed 2022].

[34] K. K. Hunt, et al. “ Schwartz's Principles of Surgery,” 2020.

[35] Y. W. L.-Y. Lawrence N. Rothenberg, et al. “Report No. 149”.*A Guide to Mammography and Other Breast Imaging Procedures.*

[36] “Cadernos da ARP,” *Pascoal A. Especificidades da Técnica Mamográfica,* pp. 17-21, 2008.

[37] H. P. M. D. Smith AP., *Emerging Technologies in Breast Cancer Detection,* pp. 16-24, 2006.

[38] M. P. G. L. Paulsen K., " Alternative Breast Imaging – Four Model-Based," *Alternative Breast Imaging – Four Model-Based,* 2005.

[39] K. D. Hawkes, et al. "MRI to X-ray mammography registration using a volume-preservingaffine transformation," 2012.

[40] 1. V. A. Neutorgasse 9/2, "The new EU General Data Protection regulation: What the radiologist should know," *The new EU General Data Protection regulation: What the radiologist should know,* 2017 .

[41] J. Geis, A. Brady, C. Wu, J. Spencer, et al., "Ethics of AI in Radiology: Joint European and North American Multisociety Statement," *Ethics of AI in Radiology: Joint European and North American Multisociety Statement,* 2020.

[42] D. B. D. B. M. L. a. Q. F. B. Deng, "Characterization of structural-prior guided optical tomography using realistic breast models derived from dual-energy x-ray mammography," *Biomedical Optics Express,* 2015.

[43] B. G. Matthew Clark, "Reproducing two-dimensional mammograms with three-dimensional printed phantoms," *Journal of Medical Imaging,* 2018.

[44] A. Creswell, T. White, V. Dumoulin,et al. "Generative Adversarial Networks: An Overview," *Deep learning for visual understanding,* 2018.

[45] R. A. Johan Andersson, "Refining Synthetic Images with GANs: An Automated Production of Object Detection Training Data," 2019.

[46] P. F. a. T. B. O. Ronneberger, "U-net: Convolutional networks for biomedical image segmentation".*U-net: Convolutional networks for biomedical image segmentation.*

[47] Y. L. Michael Mathieu, et al. "Deep multi-scale video prediction Deep multi-scale video prediction," in *Xiv preprint arXiv*.

[48] T. R. M. T. Abed, et al. ""Alzheimer's Disease Prediction Using Convolutional Neural Network," 2020.

[49] Frutuoso J., et al. "Mamografia: Recursos Materiais e Técnicas Emergentes," *Revista de Ciências da Saúde da ESSCVP,* pp. 32-45, 2009.

[50] Gongbo Sun, et al., "End-to-End Rubbing Restoration Using Generative Adversarial Networks".