



# UNIVERSITÀ DEGLI STUDI DI PADOVA

Corso di Laurea Magistrale in Medicina e Chirurgia

Dipartimento di Imaging e Radioterapia, IOV – IRCCS

UOC Radiologia Senologica

Direttore: Dott.ssa Francesca Caumo

## TESI DI LAUREA

### **Minimizing Patient Risk in Contrast Enhanced Mammography (CEM): Deep Learning Techniques for Virtual Contrast Enhancement**

Relatore: Prof. Manfredo Atzori

Correlatrice: Dott.ssa Gisella Gennaro

Laureando: Harel Kotler

Matricola: 1178779

Anno Accademico 2023/2024



# INDEX

1. INTRODUCTION.....	1
1.1. Breast cancer.....	1
1.2. Mammography.....	4
1.3. Contrast-Enhanced Mammography (CEM).....	9
1.4. Breast Imaging Reporting and Data System.....	15
1.5. Generative AI in medical imaging.....	17
1.6. Contrast generation in breast imaging.....	20
1.7. Generative adversarial networks and pix2pix.....	21
2. THESIS PURPOSE.....	24
3. MATERIALS AND METHODS.....	25
3.1. Study Population.....	25
3.1.1. Image Acquisition.....	28
3.1.2. Image Preprocessing.....	28
3.2. Train and Validation Distribution.....	28
3.3. GAN model.....	29
3.3.1 Stacked Channel Combinations.....	31
3.3.2. Oversampling.....	32
3.3.3. Model training.....	32

3.4. Model Performance Metrics.....	34
3.5. Statistical analysis .....	35
4. RESULTS .....	37
4.1. Study Population Characteristics.....	37
4.2. Whole-Image Performance Comparison Across Training and Validation .....	39
4.3. Visual Comparison of Model Outputs .....	41
4.4. Model performance on the segmented breast.....	45
4.5. Model performance in the “Mass Only” category.....	49
5. DISCUSSION.....	51
6. CONCLUSION .....	54
7. REFERENCES .....	55
8. RINGRAZIAMENTI .....	61

# **ABSTRACT**

## **Background**

Breast cancer is the most common malignancy among women and a leading cause of mortality worldwide. Early detection significantly improves prognosis and is the rationale for screening programs. Screening programs typically consist of biennial mammography, with more comprehensive annual protocols for high-risk women. However, mammography has lower sensitivity in patients with dense breasts. Contrast-enhanced mammography (CEM) addresses this limitation by improving mass visibility, offering sensitivity similar to MRI but at a lower cost with faster results. In recent years, there has been increasing interest in using generative models in many imaging tasks including contrast generation in CT and MRI. However, this area remains unexplored in CEM.

## **Motivation**

Improve diagnostic performance and reduce risks associated with contrast media by utilizing generative models to create CEM images from mammography without the use of contrast agents.

## **Objective**

Develop a generative model to produce dual-energy subtracted (DES) images of CEM from low-energy (LE) ones, enhancing the visibility of tissues and masses without the use of contrast agents.

## **Materials and Methods**

This retrospective study at Veneto Institute of Oncology (IOV) analyzed 8073 breast images from 557 women from an intermediate and high-risk screening group and a consultation group. The images were split into train and validation

groups and subsequently categorized as "With Mass" or "Without Mass". A third category labeled "Mass Only" was created based on the mass segmentation of the "With Mass" images. A Generative Adversarial Network (GAN) model was developed and trained on multiple raw and processed image channel combinations. The Validation group was divided into adipose breasts (BI-RADS density a-c) and dense breasts (BI-RADS density d). Mean Absolute Error (MAE) of both the whole segmented breast and the mass alone were used to evaluate each model.

## Results

In the segmented breast images, the model trained on both raw and processed images exhibited the best performance in both "Without Mass" and "With Mass" categories across all subgroups ( $MAE_{Breast}$ : "Without Mass" Overall = 6.2; "With Mass" Overall = 8.2). In the "Mass Only" category, the model trained on oversampled raw and processed images achieved the best performance across all subgroups ( $MAE_{Mass}$ : Adipose Breast = 39.5, Dense Breast = 35.1, Overall = 37.3), with the lowest  $MAE_{Mass}$  observed in the dense breast subgroup. Models trained only on processed images performed the worst.

## Conclusions

This study presents a novel approach to generating DES images from mammographic images using GANs. The models showed promise across various categories, suggesting the possibility of offering CEM diagnostic performance for breast cancer diagnosis while minimizing patient risks associated with radiation and contrast agents.

# RIASSUNTO

## Background

Il cancro della mammella è la forma più comune di tumore tra le donne a livello mondiale e una delle principali cause di mortalità legata al cancro. La diagnosi precoce migliora notevolmente la prognosi, rendendo cruciale l'identificazione tempestiva. I programmi di screening del cancro al seno sono rivolti principalmente a donne di età compresa tra 40 e 74 anni, attraverso mammografie biennali e/o annuali. Tuttavia, la mammografia standard ha una bassa sensibilità nel rilevare il cancro nelle donne con tessuto mammario denso. La mammografia con contrasto (CEM) affronta questa limitazione attraverso l'utilizzo del mezzo di contrasto iodato e dei raggi X a due energie, migliorando significativamente la visibilità delle lesioni con una sensibilità paragonabile alla risonanza magnetica, ma a un costo inferiore.

## Obiettivo

L'obiettivo di questo studio è di migliorare la performance diagnostica della mammografia attraverso l'utilizzo di modelli generativi, creando immagini a contrasto virtuale da mammografie, senza l'uso di agenti di contrasto reali. Questo approccio mira a ridurre o eliminare i rischi associati ai mezzi di contrasto.

## Materiali and Metodi

Lo studio ha analizzato retrospettivamente 8073 immagini mammografiche di 557 donne presso l'Istituto Oncologico Veneto. Le immagini sono state suddivise nei gruppi di training e validazione e successivamente categorizzate in "With Mass" o "Without Mass". Una terza categoria, denominata "Mass Only" è stata creata utilizzando le segmentazioni delle masse della categoria "With

Mass”. Il modello generativo, basato su una rete avversaria generativa (GAN), è stato allenato su diverse combinazioni di canali di immagini grezze e processate. Il gruppo di validazione è stato suddiviso in base alla densità mammaria in mammelle adipose (BI-RADS a-c) e dense (BI-RADS d). La Mean Absolute Error (MAE) è stata usata come la metrica principale per la valutazione di ciascun modello.

## **Risultati**

I risultati hanno mostrato un'alta consistenza tra le prestazioni del training e della validazione, con il modello migliore ottenuto combinando immagini grezze e processate nelle categorie “Without Mass” e “With Mass” ( $MAE_{\text{Breast}}$ : “Without Mass” Overall = 6.2; “With Mass” Overall = 8.2). Nella categoria “Mass Only”, il miglior modello ottenuto è stato quello allenato sull’oversampling delle immagini grezze e processate ( $MAE_{\text{Mass}}$ : Adipose Breast = 39.5, Dense Breast = 35.1, Overall = 37.3). I modelli allenati solo sulle immagini processate sono risultati i meno performanti.

## **Conclusioni**

I modelli dimostrano la capacità potenziale di ottenere le stesse performance della CEM senza l’uso del mezzo di contrasto, minimizzando i rischi per le pazienti associate alla radiazione e ai mezzi di contrasto.





# 1. INTRODUCTION

## 1.1. Breast cancer

Breast cancer (BC) is an important global health concern, representing the most common malignancy in women and a significant cancer related mortality. It is estimated that in 2024, there will be about 2 million new cancer cases and 600,000 cancer related deaths in the United States alone, with BC contributing considerably to these figures <sup>1</sup>.

BC is a multifactorial disease influenced by hormonal, environmental, lifestyle, and genetic risk factors. Hormonal risk factors include prolonged exposure to endogenous estrogens, which can result from early menarche, late menopause, or nulliparity. External risk factors involve hormone replacement therapy, obesity, alcohol consumption, and sedentary lifestyle <sup>2</sup>.

As for genetic risk factors, 5-10% of BC are hereditary and primarily driven by genetic mutations in high-penetrance genes, such as breast cancer gene 1 (BRCA1) and breast cancer gene 2 (BRCA2) <sup>2,3</sup>. These mutations lead to a lifetime risk of 70-90% of BC as well as an increased risk of early-onset BC, bilateral disease, ovarian cancer and pancreatic cancer <sup>4,5</sup>.

The stages of BC range from 0 to IV. Stage 0 is non-invasive, also known as carcinoma in situ. Stage I depicts a tumor not exceeding 2 cm in size with no involvement of the lymph nodes. Stage II tumors are greater than 2 cm in size or may affect one to five nearby lymph nodes. Stage III suggests a locally advanced tumor, involving more distant lymph nodes. Stage IV is characterized by metastasis to distant organs such as bones, liver, or lungs. Early detection is important, as patient survival rates decrease with each stage.

13.4% of the patients are diagnosed at stage 0. 42% at stage I, 32.2% at stage II, 8.7% at stage III, and 3.8% at stage IV <sup>6</sup>.

BC is categorized into four main subtypes based on molecular and histological features, as presented in Figure 1. The luminal A and luminal B subtypes are the most common, accounting for 70-80% of the cases. These subtypes are estrogen or progesterone receptor positive. The HER2-enriched subtype represents 10-20% of all cases and is characterized by the overexpression of human epidermal growth factor receptor 2 (HER2). Triple-negative breast cancer (TNBC) accounts for 10-20% of cases. This subtype lacks the expression of estrogen or progesterone receptors and does not overexpress HER2 receptors. TNBC is associated with a more aggressive disease and a worse prognosis <sup>4,7</sup>.

Molecular subtype	Luminal A	Luminal B	HER2	TNBC
ER/PR	+		-	
HER2	-	+	-	
Frequency <sup>a</sup>	50-60%	30%	10%	10-20%
Grade <sup>b</sup>	Low		High	
Prognosis <sup>c</sup>	Good		Poor	
5-y survival rate <sup>d</sup>	94.3%	90.5%	84.0%	76.9%
Treatment	Endocrine therapy		Anti-HER2 therapy	
	Chemotherapy			

**Figure 1: Characteristics of BC molecular subtypes.** Abbreviations: BC, breast cancer, ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; PR, progesterone receptor; TNBC, triple-negative breast cancer. Source: *Breast Cancer Treatments: Updates and New Challenges*. A Burguin et al. (2021).

Histologically, BC can be classified as ductal carcinoma in situ (DCIS, 20%) or invasive BC (80%). The most common invasive types are invasive ductal carcinoma (IDC, 64%) and invasive lobular carcinoma (ILC, 8%)<sup>1,6</sup>.

The most common symptom of BC is a palpable lump, recognized by up to 83% of women. Other signs include nipple changes (tenderness, inversion, discharge), skin alterations (redness, scaling, "orange peel" texture), and unexplained breast size changes or pain<sup>8</sup>. The therapeutic approach to BC incorporates a combination of surgery, radiation, chemotherapy, endocrine therapy, targeted therapy, and immunotherapy, tailored to the specific characteristics of the disease<sup>7,9</sup>.

Most BC cases occur between the ages of 55 and 74 years, with the highest incidence among women aged 70 to 74 years, at about 470 cases per 100,000 women<sup>10</sup>. The incidence of BC has also been gradually increasing at a rate of about 0.6% per year since the mid-2000s<sup>4</sup>. These trends emphasize the need for effective screening, diagnostic capabilities, and prevention strategies to address the growing burden of BC.

Despite the rising incidence, mortality rates have greatly improved, with a 42% decrease from 1989 to 2021. This reduction has prevented over 500,000 deaths and is primarily attributed to advancements in early detection and increased awareness, as well as improvements in treatment modalities<sup>4</sup>.

Breast cancer screening is a vital tool for early detection, allowing for timely and more effective interventions by identifying cancer at a preclinical stage, often before symptoms appear<sup>11</sup>. To effectively apply it, the United States Preventive Services Task Force recommends bi-annual mammography between the ages of 40 and 74<sup>4</sup>. In the European union, the general recommendation for bi-annual mammography is between the ages 50 and 69. In some European union countries, such as Italy and Sweden, it extends to 74.

In recent years, personalized screening strategies are becoming increasingly relevant, as risk factors such as genetic predisposition, breast density, and family history can significantly influence an individual's likelihood of developing breast cancer.

Available screening methods include breast self-examination, clinical breast examination, mammography, and digital breast tomosynthesis (DBT) <sup>12</sup>.

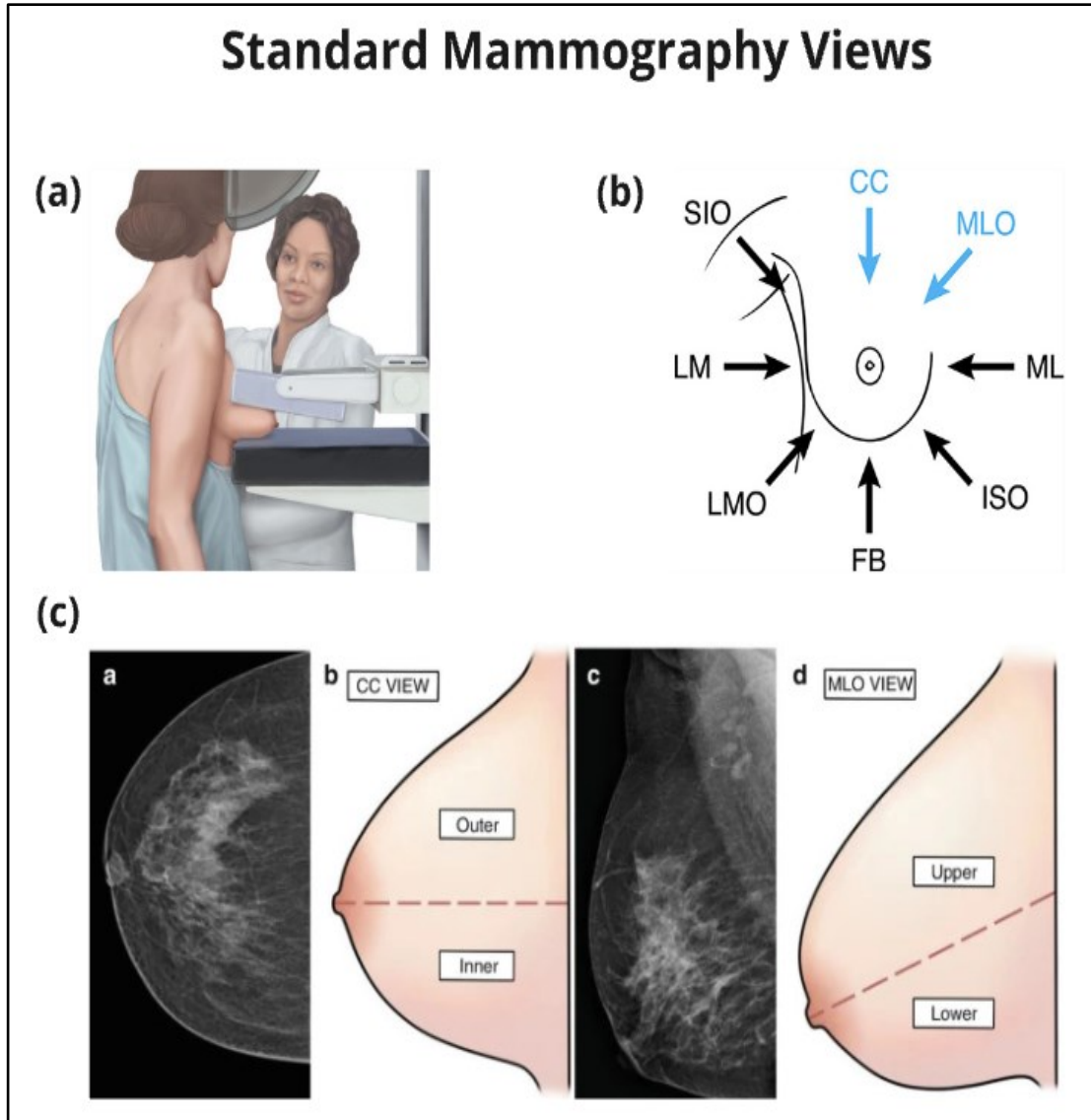
## **1.2. Mammography**

Mammography is a well-established soft tissue imaging technique used for the early detection of BC. As the primary method for mass screening, mammography is one of the most thoroughly investigated areas of medicine, with robust evidence supporting its role in mortality reduction by up to 50% <sup>13</sup>.

Mammography has a high diagnostic accuracy, ranging from 79% to 97% <sup>14,15</sup>. Furthermore, it is designed to meet several key objectives: early BC detection, acceptable side effects for the screened population, reproducibility of results, and cost-effectiveness when applied at regular intervals <sup>16</sup>.

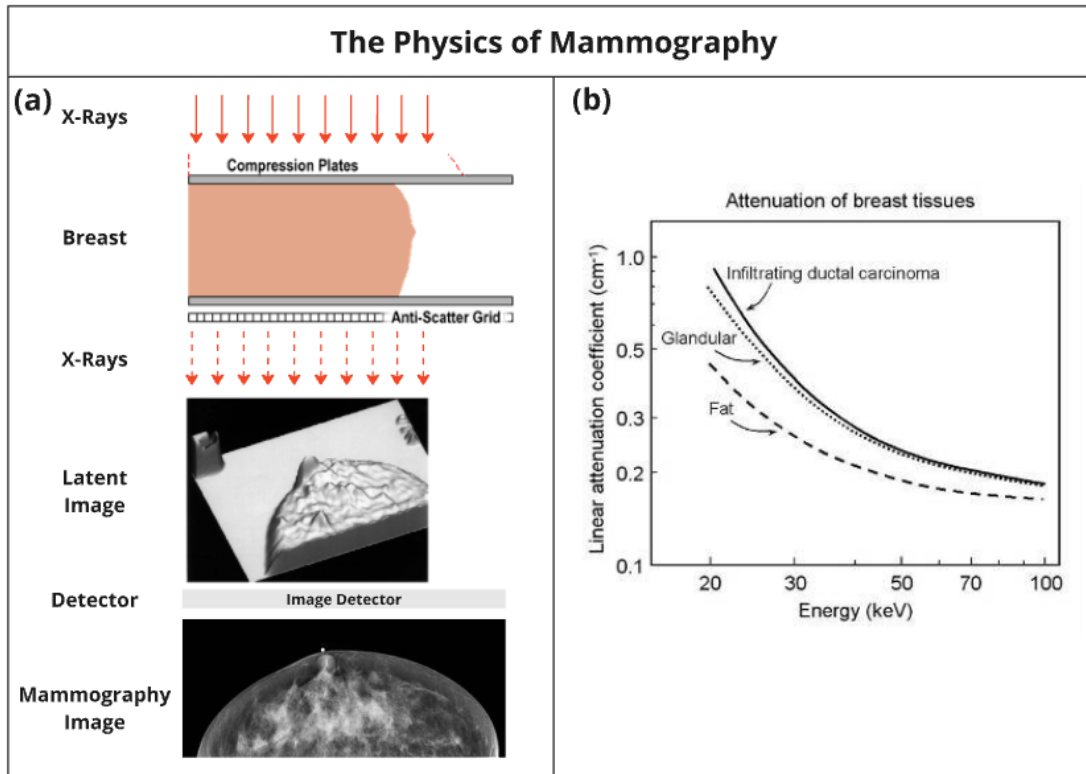
The standard views in mammography include craniocaudal (CC) and mediolateral-oblique (MLO), which help in visualizing different parts of the breast and improve the detection of abnormalities near the chest wall <sup>17</sup>. Another frequently captured one is the mediolateral (ML) view.

Figure 2 illustrates the process of positioning the breast for the correct mammographic view, the entire list of views, and the resulting images paired with the corresponding breast view illustration.



**Figure 2: Mammographic views.** (a) The process of positioning the breast for the correct mammographic view; (b) all mammographic view orientations; (c) the resulting images paired with the corresponding breast view illustration. Sources: (a) *Breast Imaging: Mammography, Digital Tomosynthesis, Dynamic Contrast Enhancement*. Ebrahimi et al. (2019); (b) *Instituto de Física Corpuscular, topics, diagnose breast*; (c) *Radiological Anatomy*, Gnanavel et al. (2023). Abbreviations: CC, craniocaudal; FB, from below; ISO, inferosuperior oblique; LM, lateromedial; LMO, lateromedial oblique; ML, medio lateral; MLO, medio lateral oblique; SIO, superoinferior oblique.

The physics behind mammography involves the attenuation of X-rays as they pass through different breast tissues, as demonstrated in figure 3, where glandular and adipose tissues exhibit different attenuation coefficients <sup>18</sup>.



**Figure 3: physics of digital mammography.** (a) X-rays passing through compressed breast tissue, creating a latent image which is then detected by the detector and converted into a mammography image. (b) Attenuation of different breast tissues. Abbreviations: keV, Kilo electron volt. Source: Bushberg, et al. *The Essential Physics of Medical Imaging*, 2nd ed., p. 193.10 (2002).

In mammography, contrast is an important factor that can determine the visibility of breast tissue abnormalities such as tumors. The main source of contrast in mammographic images is the variations in X-ray attenuation between the glandular and the adipose tissues. Glandular tissue appears relatively denser; hence, contrast should be optimized to better distinguish between healthy and potentially malignant regions <sup>17</sup>. Lower energy X-rays

produce greater differential attenuation between the tissues, increasing the visibility of small abnormalities<sup>18</sup>. However, contrast optimization can also introduce noise that reduces the quality of the image as it conceals subtle details. To reduce noise, advanced image processing techniques that balance noise reduction with the preservation of diagnostic information are applied. The effects of contrast levels and noise levels are illustrated in Figure 4.

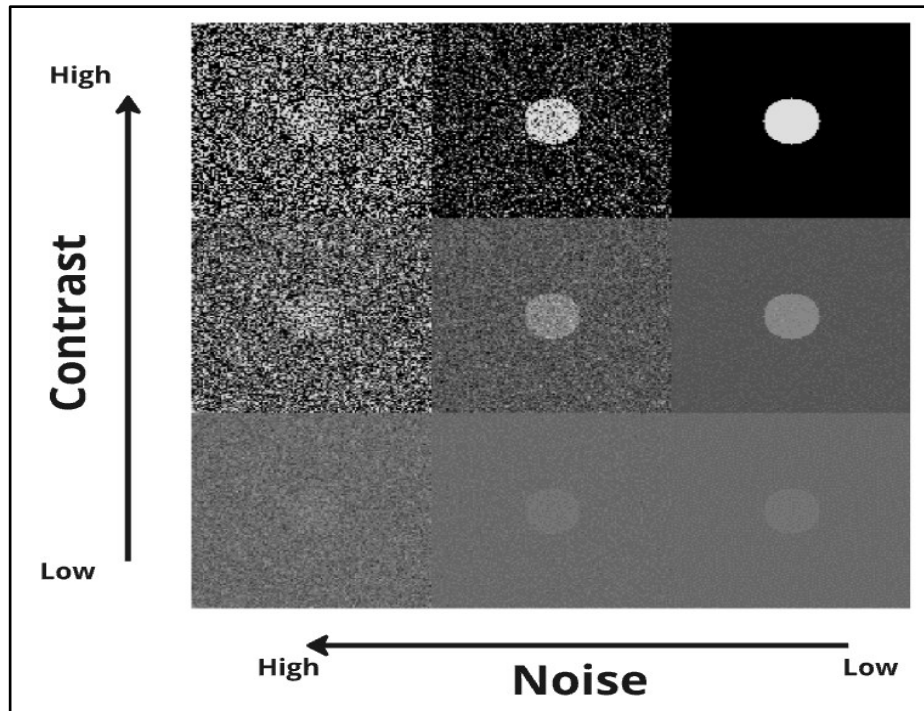
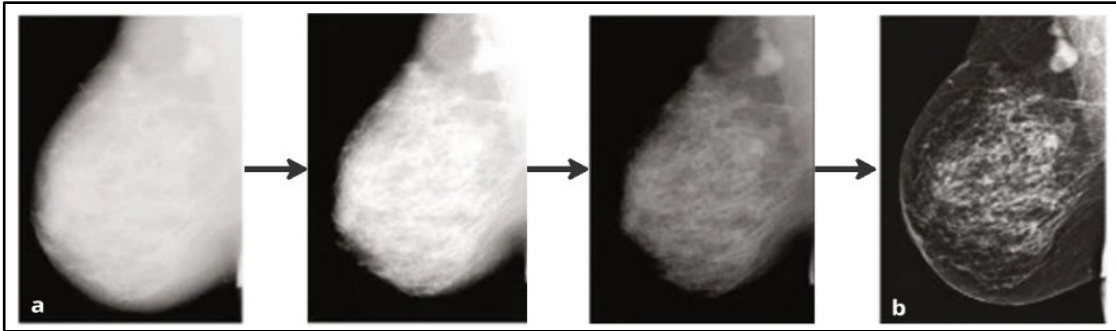


Figure 4: The effect of contrast and noise on mammographic image quality.

Processing is a distinct feature of mammography because it enables improved visualization of breast structures. Obtained mammography images **are initially raw** with low intrinsic contrast. Processing algorithms are applied to these images to improve contrast, highlight areas of interest, and adjust for differences in breast tissue density, as demonstrated in figure 5. This leads to **a processed image** with clearer visualization of both peripheral and internal regions. This procedure ensures simpler abnormalities identification by radiologists<sup>17</sup>.





**Figure 5: The effect of different processing algorithms on mammography, from raw (a) to fully processed (b) images.** Source: *Breast Imaging Physics in Mammography (Part II)*. Fico et al. (2023).

Mammography, however, has some limitations in that it uses a two-dimensional (2D) image for a three-dimensional object (3D), leading to overlapping tissues of different densities. **This can lower the sensitivity, particularly in women with dense breast tissue as the glandular tissues may obscure underlying tumors or create artificial ones.** These limitations are more apparent in screening compared to diagnostic settings, where other modalities such as ultrasound (US) can add more information <sup>15</sup>.

Increased breast density significantly impacts mammography performance and BC risk. Women with dense breasts have a 4-6 times higher risk of BC <sup>19</sup>. Mammographic sensitivity greatly decreases in denser breasts, leading to a 10-20% reduction in both sensitivity and specificity <sup>20,21</sup>. The impact of density on cancer risk and detection persists for at least 8 years and is more pronounced in younger women <sup>19</sup>.

To overcome this limitation, newer technologies were developed including contrast-enhanced magnetic resonance imaging (MRI), DBT, and contrast-enhanced mammography (CEM).

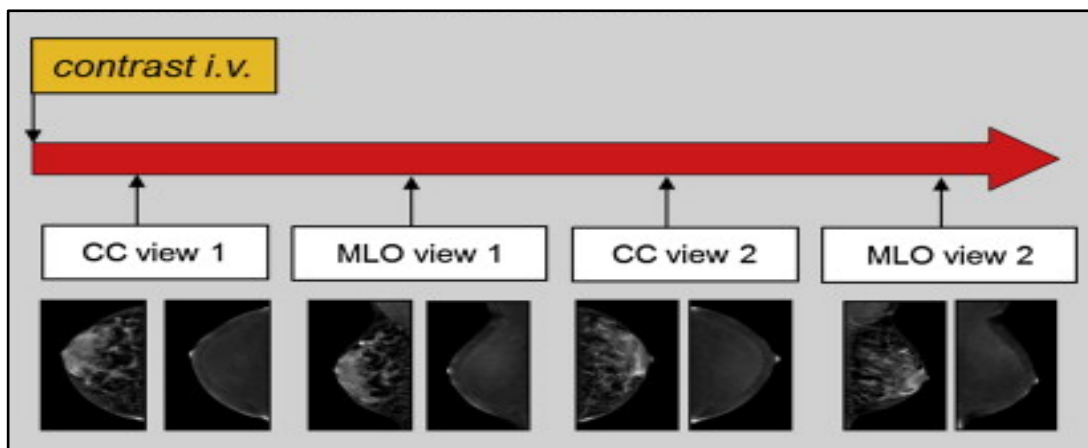
DBT is based on geometric subtraction, where several X-ray images are taken from various angles and then reconstructed into thin slices. Temporal

subtraction is used in contrast-enhanced MRI by comparing images taken at different times. CEM uses intravenous contrast agents and energy subtraction to better represent abnormal vascularization, which is commonly associated with malignancies; thus, it is particularly strong in detecting tumors on women who have dense breast tissue <sup>17</sup>.

### 1.3. Contrast-Enhanced Mammography (CEM)

CEM is an advanced imaging technique designed to improve BC detection, specifically in women with dense breast tissue <sup>22</sup>. This technique merges morphological and functional information by using dual-energy subtraction to increase the visibility of potential malignancies.

A CEM exam begins with the intravenous injection of an iodinated contrast agent 2 minutes before the image acquisition begins. Next, as shown in Figure 6, two views are taken for each breast, producing the right CC view, right MLO view, left CC view, and left MLO view.



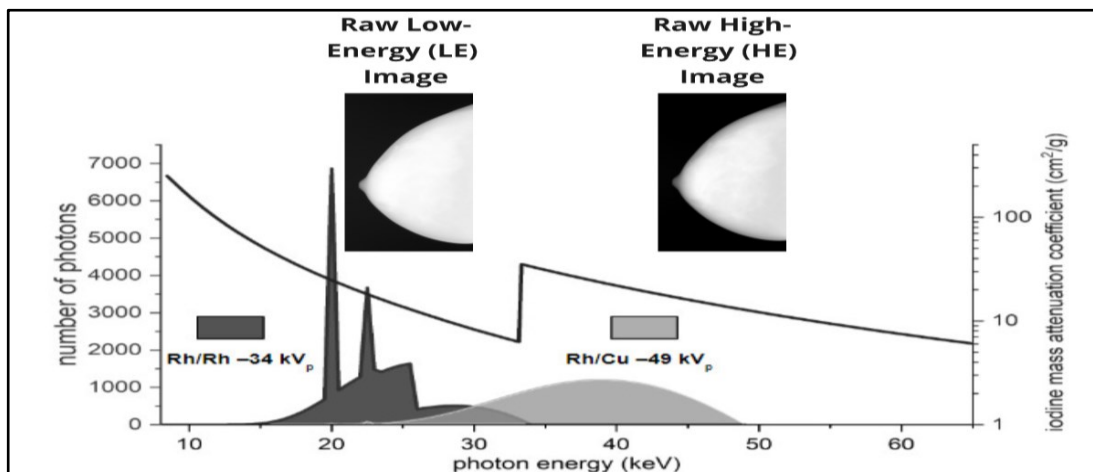
**Figure 6: CEM examination.** A single contrast agent injection followed by the capturing of two views (craniocaudal, Medio-lateral obliques) for each breast (1 and 2). Abbreviations: CC, craniocaudal; CEM, contrast-enhanced mammography; MLO, medio-lateral oblique. Source: Contrast enhanced mammography: Techniques, current results, and potential indications. M.B.I Lobbes et al. (2013).

The absorption peak of iodine is approximately 33.2 KeV, which is compatible with the 25-49 kV range used in mammography. CEM utilizes two distinct X-ray energy beams for improved contrast between tissues. The first beam, with a lower energy, typically falls below the iodine peak absorption. The second one, with a higher energy, surpasses the iodine absorption peak <sup>23</sup>.

For each mammography view, these two beams are captured and used to produce two images:

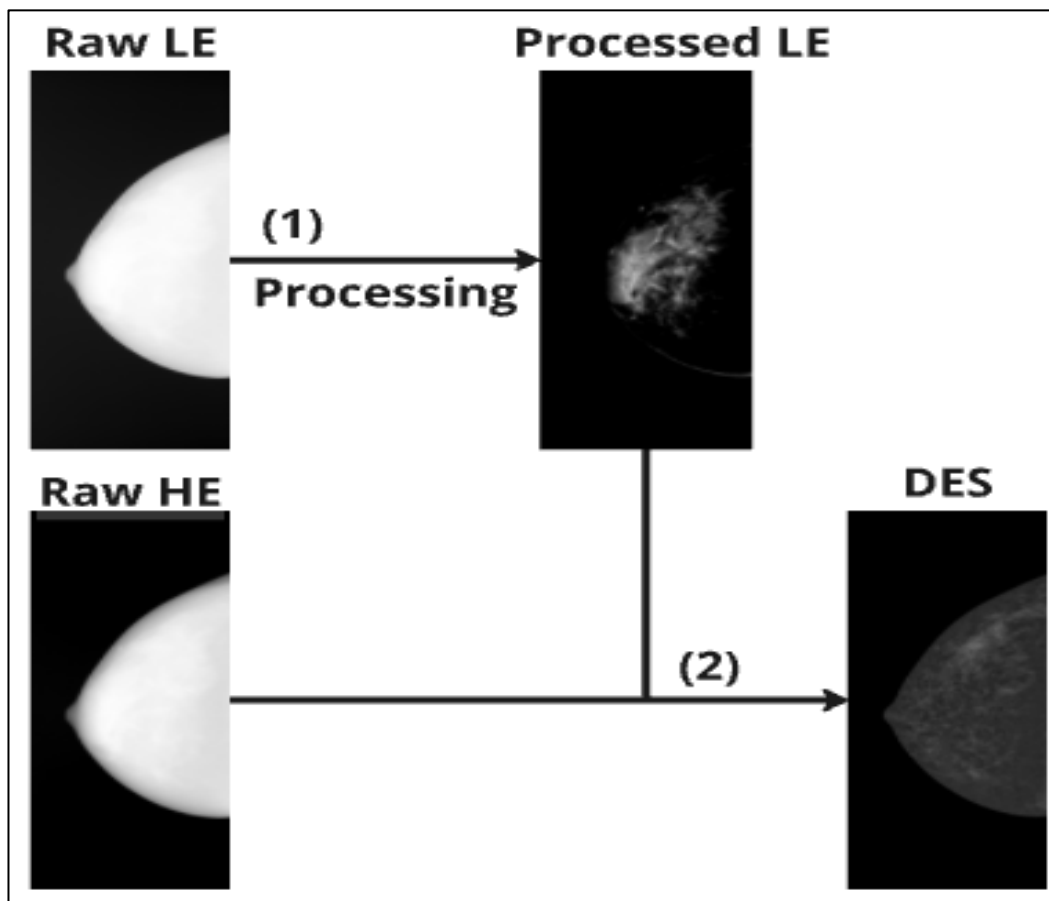
1. Raw low-energy (LE) image - similar to a standard mammography, for X-ray energy spectrum below the iodine absorption peak at (33.2 keV).
2. Raw high-energy (HE) image - accounting for X-ray energy above 33.2 keV to maximize absorption by the contrast agent.

The two images and the corresponding X-ray spectrums are described in Figure 7.



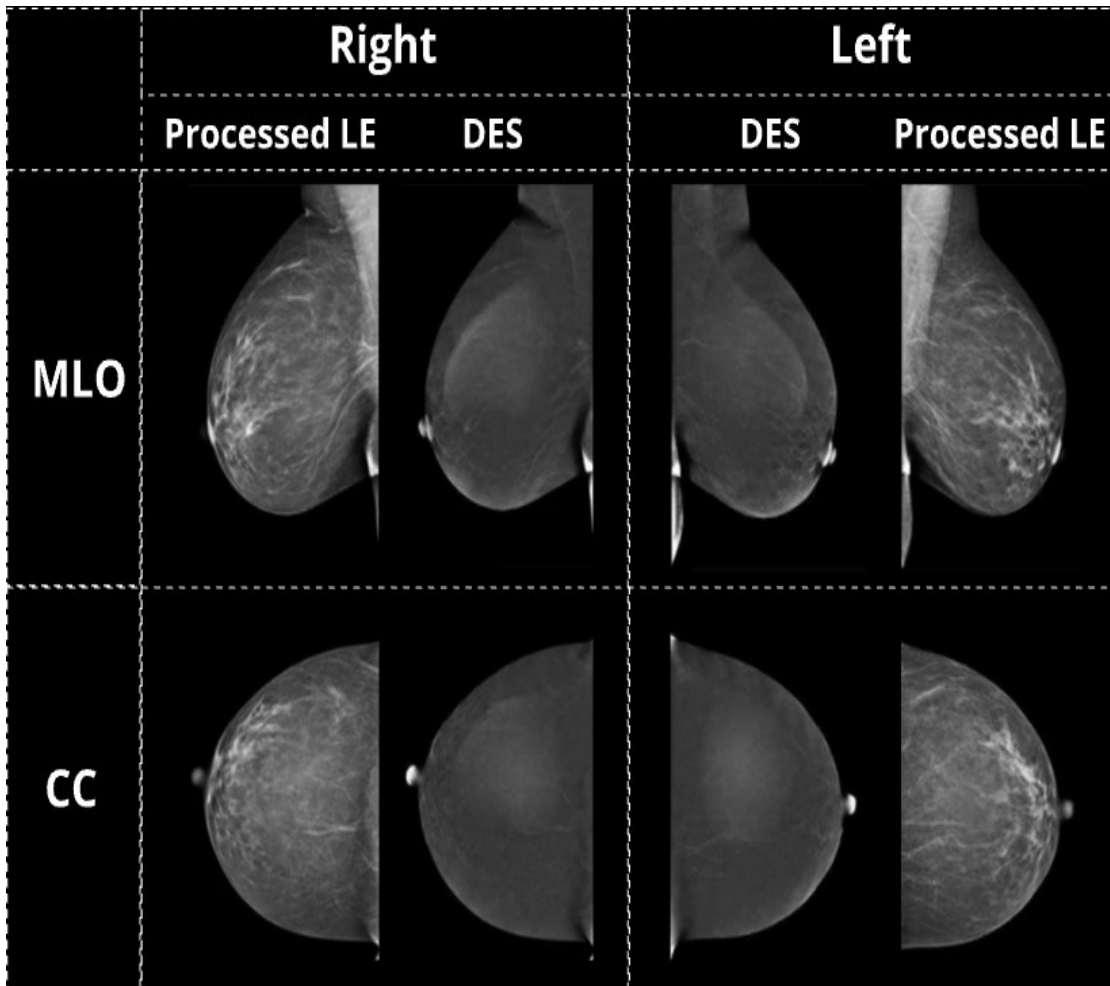
**Figure 7: Dual energy spectrums divided by the absorption of iodine with the resulting captured raw images.** The darker gray curve represents photons in the lower energy spectrum absorbed by a rhodium (Rh/Rh) filter and translated to the raw LE image. The lighter gray curve represents photons in the higher energy spectrum absorbed by rhodium/copper (Rh/Cu) filter and translated to the raw HE image. The black line represents the iodine absorption, with the peak resulting at approximately 33.2 KeV. Abbreviations: HE, high-energy; LE, low-energy; keV, kilo-electron volts.

The raw LE image is first processed, creating a processed LE image equivalent to standard mammography image<sup>24</sup>. Subsequently, the processed LE image and raw HE images are combined through a process called dual-energy subtraction (DES), creating a hybrid image denominated by the same name (DES) that reduces anatomical noise caused by overlapping tissues and improves the detection of malignancies<sup>25</sup>. The steps from acquisition to the creation of the final DES image are described in Figure 8.



**Figure 8: CEM, from raw LE to DES.** The sequential steps from image acquisition to the creation of the final DES image. (1) Initially, the raw LE image is processed to generate a processed LE image. (2) This processed LE image is then combined with the raw HE image through dual-energy subtraction, resulting in a hybrid DES image. Abbreviations: CEM, contrast-enhanced mammography; DES, dual-energy subtraction; HE, high-energy; LE, low-energy.

When interpreting a CEM study, radiologists are presented only with processed LE and DES images. The raw LE and HE images are not directly viewed or interpreted, as they are used solely for obtaining the processed LE and DES images. The interpretation is performed with both the processed LE and DES images viewed for comparison, as presented in Figure 9 <sup>26</sup>.



**Figure 9: Representative normal CEM study with standard 8 views showing the LE and DES.** Adapted from *Contrast enhanced mammography: focus on frequently encountered benign and malignant diagnoses*. Mindy L. Yang et al. (2023). Abbreviations: CC, craniocaudal; DES, dual-energy subtraction; LE, low-energy; MLO, mediolateral oblique.

Compared to mammography, CEM further improves contrast by using an iodinated contrast agent, which highlights areas of increased vascularity, such as tumors, offering better differentiation from normal breast tissue. This technique is especially beneficial for detecting lesions in dense breasts, where traditional mammography may struggle to provide sufficient contrast <sup>17</sup>.

Furthermore, CEM has emerged as a highly effective imaging technique, offering high sensitivity and specificity that rivals MRI, while being more cost-effective and accessible <sup>27,28</sup>.

Despite its advantages, CEM does have some limitations, including increased radiation dose compared to standard mammography, and a low risk of allergic reactions related to the iodine-based contrast agents.<sup>27,29</sup>

Studies have shown that the radiation dose for CEM is about 30% higher than that of standard mammography, attributed to the dual-energy approach used in this technique <sup>30</sup>. Such an increment was close to the dose increase reported for DBT, which is already used in both screening and diagnostic settings. Despite the increased dose, the diagnostic benefits of CEM, especially in dense breast tissue, often outweigh the risks, making the higher radiation dose a manageable concern in clinical practice.

In addition to the radiation dose, the administration of iodinated contrast agents in CEM introduces potential risks, which can be classified into immediate and delayed reactions.

Immediate reactions are the ones that occur within the first hour of contrast administration and are usually anaphylactoid or IgE dependent. These reactions may be mild and include pruritus, urticaria, and flushing, but they can also be life threatening and include cardiovascular shock, cardiac arrhythmia, and even cardiac arrest <sup>31</sup>.

Delayed reactions, however, can take between one hour and seven days after the administration of the contrast. These reactions are generally T-cell mediated and the manifestations are mainly cutaneous including exanthematous reactions and erythema multiforme. Less common delayed reactions include severe conditions such as Stevens-Johnson syndrome and toxic epidermal necrolysis.

Contrast-induced nephropathy (CIN) is another dangerous complication of the use of iodinated contrast media, especially among patients with prior kidney disease or other risk factors such as diabetes and advanced age <sup>31</sup>. Cancer patients are more prone to CIN because of the more-frequent requirement for contrast-enhanced imaging studies and exposure to nephrotoxic chemotherapy drugs <sup>32</sup>.

Another limitation of CEM is the uptake of contrast medium by normal fibroglandular breast tissue, known as background parenchymal enhancement (BPE). This also occurs in breast MRI but can be mitigated by timing the exam to coincide with the menstrual and proliferative phases of the cycle <sup>33</sup>.

Nonetheless, CEM is generally well-tolerated by patients, often better than breast MRI, and is considered a safe procedure when proper precautions are taken <sup>34</sup>. These precautions include assessing renal function, confirming the absence of known allergies, and obtaining informed consent before administering the contrast medium.

In summary, while CEM does involve higher radiation exposure and the use of contrast agents that carry some risk of adverse reactions, these risks are generally outweighed by its significant benefits in BC detection. **Reducing or eliminating these risks without compromising diagnostic accuracy could lead to even broader clinical adoption of this powerful imaging technique.**

## 1.4. Breast Imaging Reporting and Data System

The Breast Imaging Reporting and Data System (BI-RADS), developed by the American College of Radiology (ACR), is a standardized system designed to enhance the consistency of breast imaging reports and improve communication among healthcare providers <sup>35</sup>. BI-RADS includes a lexicon of descriptors, numerical assessment categories (0-6), and breast composition categories (A-D).

The breast composition categories include <sup>35-37</sup>:

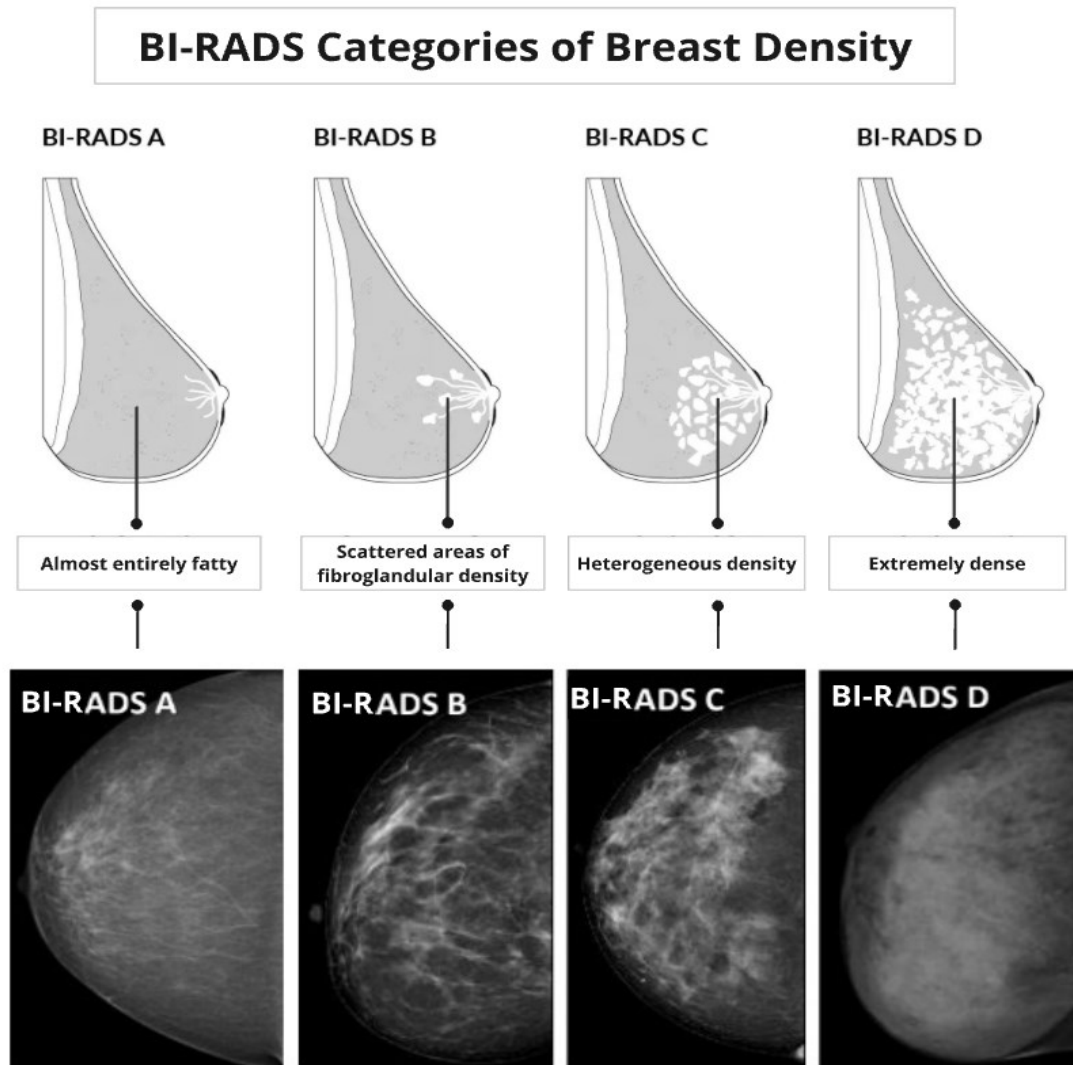
1. Category A - Almost entirely fatty breast
2. Category B - Scattered areas of fibroglandular density
3. Category C - Heterogeneously dense breast
4. Category D - Extremely dense breast

These categories, often referred to as BI-RADS density categories, are used as a qualitative measurement for tissue density and can be applied across mammography, US, and MRI, making it a versatile tool in breast imaging.

Breast density is a key factor in both BC risk and the effectiveness of mammographic screening. Women with dense breasts, characterized by higher proportions of fibroglandular tissue, have a 4-6 times greater risk of developing breast cancer and face reduced mammographic sensitivity, which drops from 87% in fatty breasts to 62.9% in extremely dense breasts, and a decrease in specificity from 96.9% to 89.1% <sup>19,20</sup>. The dense tissue can obscure tumors, complicating cancer detection and leading to higher rates of interval cancers, especially in women with BI-RADS category D density <sup>38</sup>.

Figure 10 demonstrates the BI-RADS density categories as an illustration paired with a corresponding mammography image.





**Figure 10: BI-RADS density categories both illustrated and in mammography.** Abbreviations: BI-RADS, Breast Imaging Reporting and Data System. Source: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System (2013).

In recent years, software tools have been developed to assess the density of the breast in a more reproducible manner using quantitative measures instead of qualitative ones<sup>39</sup>. These methods include area-based measures, which analyze specific areas of the image, and volumetric measures, which assess x-ray attenuation properties to quantify breast tissue composition providing volumetric breast density as a percentage<sup>40</sup>.

## 1.5. Generative AI in Medical Imaging

Medical imaging is a cornerstone of modern healthcare, facilitating early diagnosis and the development of efficient treatment plans. Commonly used imaging modalities, including X-ray, computed tomography (CT), MRI, and US, are indispensable for diagnosing and monitoring conditions such as cancer, cardiovascular diseases, and neurological disorders. These technologies have revolutionized clinical practice by enabling accurate and high-resolution imaging, which has improved diagnosis, treatment planning, and outcome assessment.

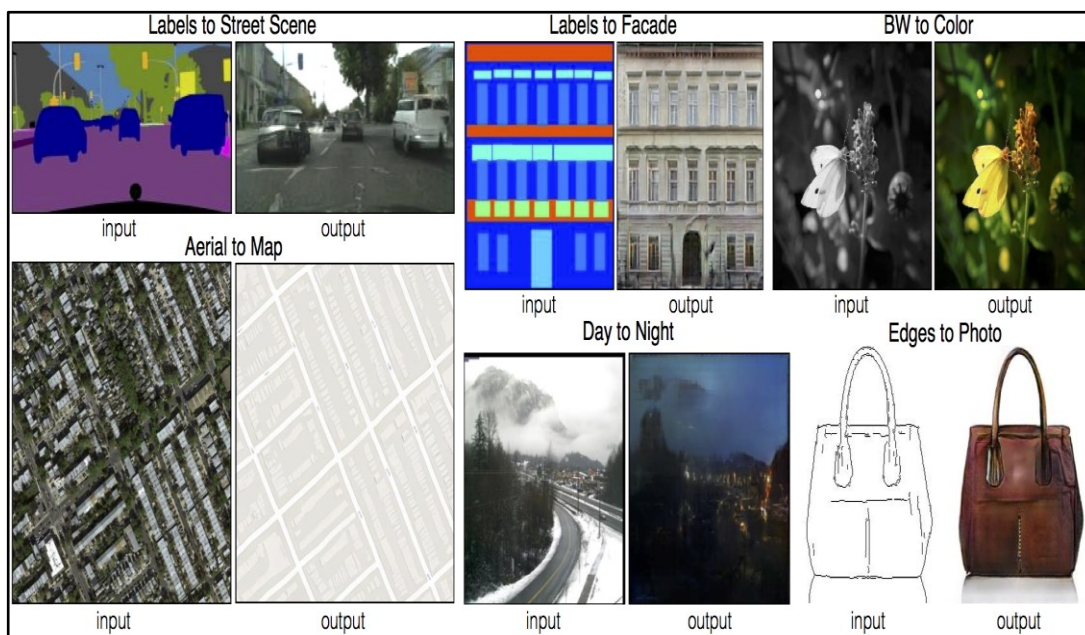
In recent years, generative artificial intelligence (AI) has further enhanced medical imaging by facilitating labor-intensive tasks <sup>41</sup>. Models such as generative adversarial networks (GANs), diffusion models, and variational autoencoders (VAEs) are employed to synthesize realistic medical images by assigning higher probabilities to more plausible images. These models have several applications, such as data augmentation, image synthesis, and image-to-image translation <sup>42,43</sup>.

In data augmentation, generative AI models create synthetic medical images to increase dataset size. This approach helps improve the performance of machine learning models while mitigating issues such as patient privacy concerns, high imaging costs, rarity of certain conditions, and improving model generalization <sup>44,45</sup>.

Another application for generative AI is simulating disease progression and predicting patient outcomes, enhancing clinical decision-making. One such application is the generation of counterfactual images, which allow clinicians to visualize alternative scenarios, such as the effect of a tumor being surgically removed, thereby aiding in treatment planning <sup>46</sup>.

Another application is in image completion, reconstructing missing or corrupted sections of scans, such as incomplete MRIs, enhancing the overall quality and usability of these images <sup>47</sup>. Furthermore, generative models can aid with image normalization, adjusting images for consistency, making them easier to interpret across different cases <sup>48</sup>. Additionally, super-resolution techniques allow these models to produce higher-resolution versions of existing medical images, which can be particularly valuable for detailed tissue analysis and diagnosis <sup>45</sup>.

A promising application within this field is image-to-image translation, where AI models learn the relationships between features in an image from the source domain (e.g., style, structure, or content) and how to associate them with corresponding features in the target domain. Examples for image translation from a source domain to a target domain are demonstrated in Figure 11.



**Figure 11: image translation from source domains to target domains.** Abbreviations: BW, black and white. Source: Image-to-Image Translation with Conditional Adversarial Networks P. Isola et al. (2018)

This technique is particularly useful when certain imaging modalities are unavailable, such as in cases where MRI machines are inaccessible, or when lower-cost alternatives like CT scans are preferred. Recent works focusing on cross-modality image synthesis, such as translating between MRI and CT images, demonstrated that AI models can leverage the unique information provided by different modalities, making the diagnostic process more efficient and cost-effective <sup>44,49,50</sup>.

In mammography, several generative models were developed in recent years. The MAM-E pipeline utilizes diffusion models for mammographic image synthesis, generating specific lesions by text prompts<sup>51</sup>. MammoGANesis, a GAN based framework, can synthesize mammographic images <sup>52</sup>. Moreover, ROI-mammoGAN focuses on synthesizing regions of interest in mammographic images, addressing the challenge of data scarcity <sup>53</sup>.

A particularly interesting application of image-to-image translation in recent years is to reduce patient risk, for instance, by avoiding or reducing the use of contrast agents in imaging.

## 1.6. Contrast generation in breast imaging

Contrast media has recently become a central interest in generative AI application due to its ability to significantly improve the diagnostic accuracy on one hand, and possibly induce side effects on the other. That is why, in recent years, generative AI models have emerged as a powerful tool creating "virtual" or "augmented" contrast media.

Virtual contrast refers to the creation of enhanced images from non-contrast scans using AI models, while augmented contrast amplifies the effects of low-dose contrast administration. This approach aims to reduce the reliance on traditional contrast agents, thereby mitigating their associated risks <sup>54,55</sup>

For example, in brain MRI, AI-generated virtual contrasts have been employed to address the potential risks of gadolinium-based contrast agents, particularly for patients with renal failure <sup>56</sup>. Similarly, in abdominal CT, recent models have shown potential in producing synthetic contrast-enhanced images, improving lesion detection in non-contrast CT scans <sup>57</sup>.

Generative tasks are not without challenges, as they might lead to the introduction of unrealistic information <sup>58</sup> and require geometrical alignment between the non-contrast and contrast-enhanced images <sup>57</sup>. Despite the challenges, these models offer great possibilities if successfully implemented.

In the context of breast imaging, contrast generating models have focused so far on MRI.

GANs have shown promise in synthesizing contrast-enhanced MRI images from non-contrast scans. For instance, works from Kim et al. and Osuala et al. have shown GAN developed to synthesize contrast-enhanced MRI images

from pre-contrast T1-weighted images. The resulting images enhance tumor visibility without the need for invasive contrast administration <sup>59,60</sup>.

Despite its diagnostic potential, generative models in CEM are relatively rare due to the technical challenges involved. These models face high-resolution images, making the data processing complex and resource intensive. Additionally, CEM is a newer technique with a more limited amount of available data compared to other imaging modalities, which further complicates the development of robust generative models. Furthermore, there is an elevated inter-manufacturer variability.

## 1.7. Generative adversarial networks and pix2pix

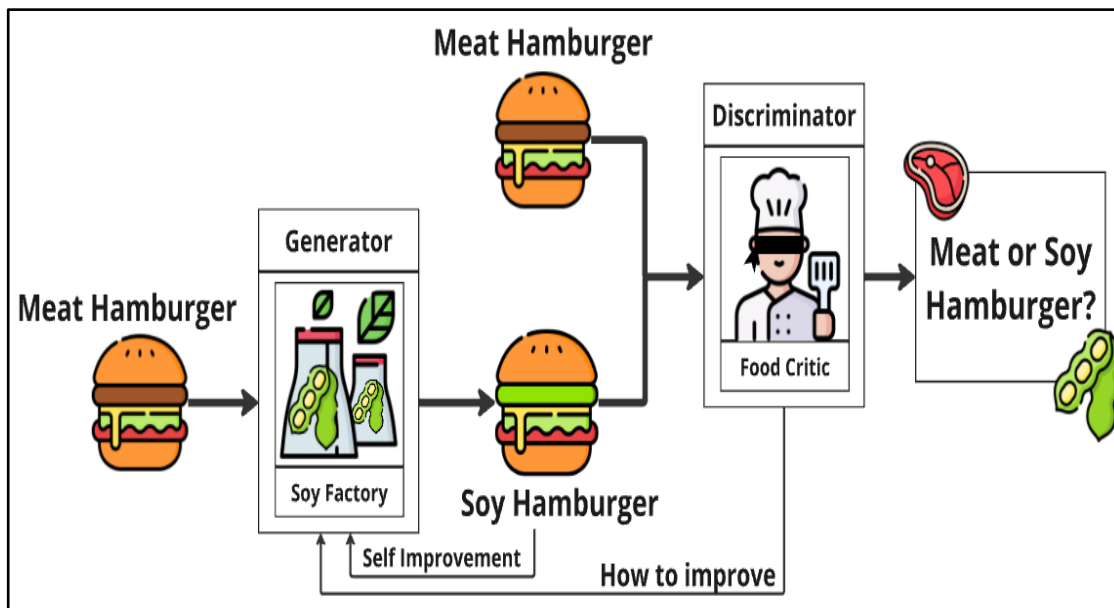
GANs and the pix2pix framework are significant advancements in the field of generative models, particularly for image synthesis and translation tasks.

Introduced by Goodfellow et al. (2014), GANs use a competitive framework involving two neural networks: **a generator and a discriminator**. The generator creates synthetic data, while the discriminator differentiates between real and synthetic samples. These two adversaries train at alternating times, refining both networks through a process of adversarial training <sup>61</sup>.

The core idea of GANs lies in this adversarial relationship, where the generator aims to produce data that is indistinguishable from real samples, and the discriminator's task is to identify which samples are real and which are fake. The generator initially starts by producing random noise as data and gradually learns to generate more realistic samples over time.

Through this alternating iterative process, illustrated in Figure 12, the generator improves its ability to create increasingly realistic samples (represented as creating soy hamburgers from meat hamburgers), while the discriminator

becomes better at detecting generated samples (represented as distinguishing between soy and meat hamburgers). This unique approach has proven to be highly effective in tasks that require generating new data that mirrors the characteristics of the original dataset, making GANs widely adopted in various domains <sup>61</sup>.



**Figure 12: The adversarial process of GANs.** This figure illustrates the core adversarial process of GANs as “Meat vs Soy Hamburger”, where the generator is represented as a “soy factory” attempting to produce soy hamburgers that resemble real meat hamburgers. The discriminator, depicted as a “blindfolded food critic” aims to distinguish between real meat hamburgers and fake soy hamburgers. Through this iterative process, the generator improves its ability to create more convincing soy hamburgers, while the discriminator sharpens its skill in identifying which hamburgers are real. This adversarial relationship is key to generating realistic data in GAN models. Abbreviations: GAN, generative adversarial network.

GANs have improved how we approach the generation of realistic images, as the adversarial training leads to sharper details compared to VAEs which tend to have more blurry images <sup>62–65</sup>.

GANs have certain limitations. One of which is that the generator's ability to produce discrete data, such as a word or a character, is restricted. Another limitation is that the generator produces only a limited variety of outputs, even when more diverse outputs are expected, confining GANs to tasks where generating a narrow range of outputs is acceptable <sup>65</sup>.

Building on the GAN architecture, Isola et al. (2016) introduced the pix2pix framework for image-to-image translation <sup>66</sup>. Unlike the original GAN framework, where the input is typically random noise, pix2pix takes an image from one domain (e.g., a grayscale image) and translates it to a target domain (e.g., a colored image). The discriminator in pix2pix distinguishes between the generated output and the ground truth image, further refining the quality of the generated images. To achieve this task, pix2pix framework requires a "paired" dataset, containing both the origin and the target images. Several modifications, such as feature-matching loss and style transfer techniques, have been introduced to enhance the performance of pix2pix for more complex tasks <sup>67</sup>.

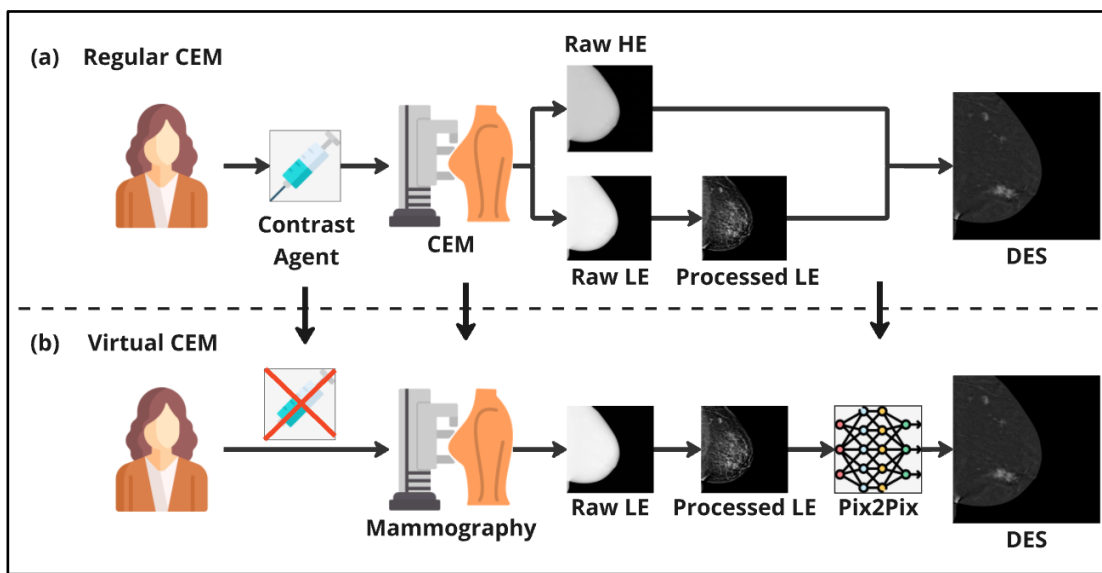
Hence, **effectively building a generative model for CEM requires the gathering of a large enough image sample** from the same manufacturer and model, with verified ground-truth annotation. **Once such a dataset is prepared, pix2pix would be the architecture to begin with**, as it is a well-established model for image-to-image translation with demonstrated strong performance across similar tasks across the various imaging techniques <sup>68</sup>.



## 2. THESIS PURPOSE

Our goal is to harness recent advancement in generative AI to create the dual-energy subtracted images of contrast enhanced mammography from low-energy images without the need for contrast agents, as illustrated in Figure 13.

We argue that by training generative AI models on a large dataset of these images, we can generate realistic and accurate images that could offer the same diagnostic performance as contrast-enhanced mammography without the risks involved in radiation and contrast agents.



**Figure 13 - Generating DES images using GANs.** (a) Regular CEM, involving the injection of a contrast agent, capturing of HE and LE images and for the creation of a DES image. (b) Virtual CEM, eliminating the need for a contrast agent, captures only LE images (just as a regular mammography) and uses pix2pix to generate the DES images for them. Abbreviations: CEM, contrast-enhanced mammography; DES, dual-energy subtraction; HE, high-energy; LE, low-energy.

### **3. MATERIALS AND METHODS**

The study protocol was approved by the appropriate institutional ethics committee. Given its retrospective nature, the study was exempted from the requirement for individual informed consent. All patient images were anonymized, and the study was conducted in compliance with relevant ethical standards.

#### **3.1. Study population**

This study is a retrospective observational analysis conducted at the Veneto Institute of Oncology (IOV – IRCCS), Padua, Italy. The study includes two major groups of patients: (1) Intermediate and high-risk screening group, and (2) consultation group.

The intermediate and high-risk screening protocol group includes women undergoing annual surveillance, with intermediate-risk individuals receiving mammography and US, while high-risk individuals also undergo MRI. 7014 images were acquired from a total of 488 patients in this group between April 2019 and October 2022.

The consultation group includes patients who arrived at the IOV - IRCCS for second opinions, suspected cases, or biopsies. This cohort comprises 2184 images acquired from 148 patients between March 2019 and January 2024.

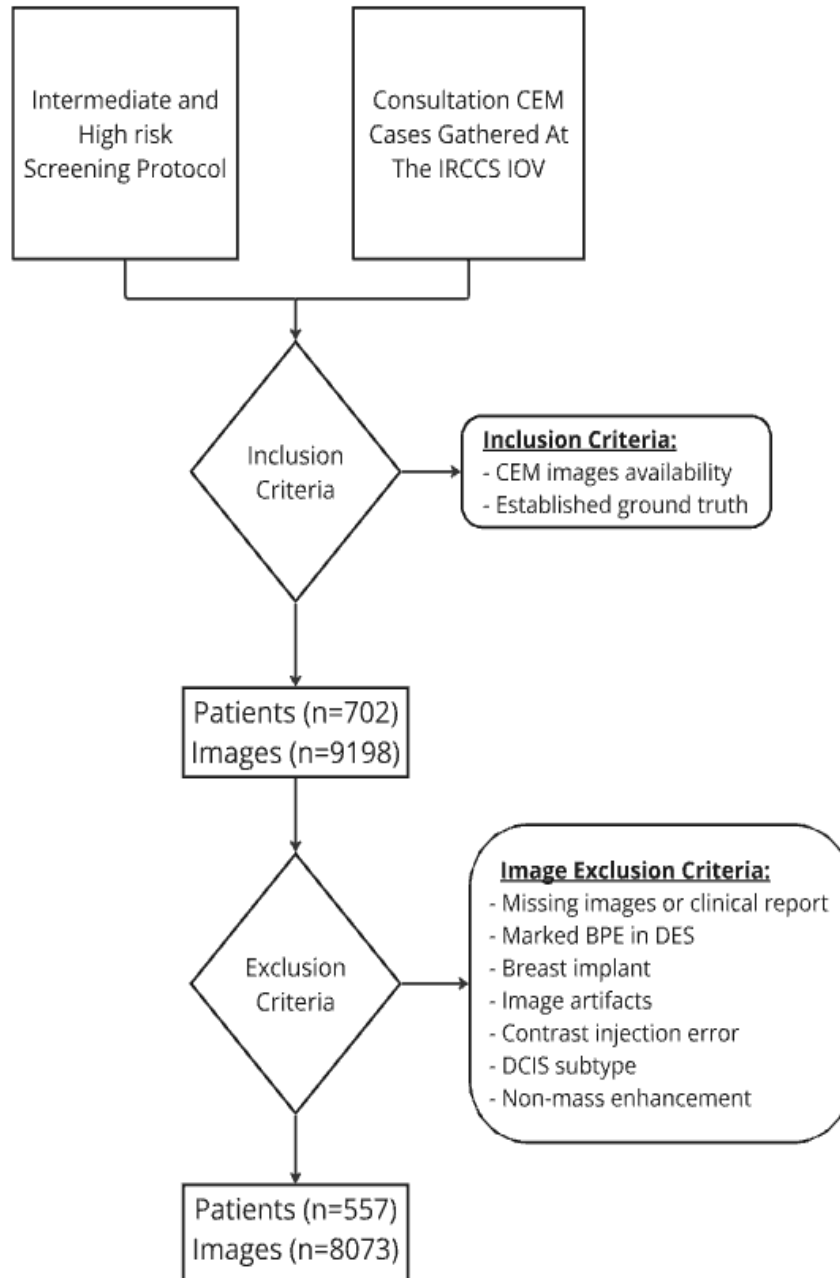
Inclusion Criteria:

- Availability of CEM images (LE Raw, LE Processed, DES)
- Ground truth established by either biopsy results for masses or at least 1 year of follow-up for negative cases.

Exclusion Criteria:

- Missing images or clinical report
- Marked BPE in DES
- Presence of breast implants
- Presence of Image artifacts
- Contrast injection error
- DCIS subtype (due to lack of contrast uptake in CEM)
- Non-mass enhancement

The final population for this study consists of 557 women and 8073 breast images, as described in figure 14.



**Figure 14: Study inclusion and exclusion criteria.** Flowchart of the datasets included in the analysis, detailing patient and image selection criteria. Ground truth established by biopsy results for masses or at least 1 year of follow-up for negative cases. Abbreviations: BPE, Background Parenchymal Enhancement; CEM, Contrast-Enhanced Mammography; DES, Digital Energy Subtraction; IOV-IRCCS, Istituto Oncologico Veneto - Istituto di Ricovero e Cura a Carattere Scientifico.

### **3.1.2. Image acquisition**

All images were acquired and interpreted by three experienced breast radiologists at the IOV. For each breast, the CC, MLO, and ML views were acquired. For each view, four images: raw LE, raw HE, processed LE, and DES, were stored in the picture archiving and communication system (PACS).

### **3.1.3. Image preprocessing**

Based on the ground truth, each image was further classified based on the presence or absence of contrast uptaking masses and labeled as "With Mass" or "Without Mass" images.

All the images were extracted from their original 12-bit or 13-bit DICOM file format, converted and saved as 8-bit grayscale images in TIFF format. The grayscale 8-bit image is represented by a single array, defined as a channel, with pixel values ranging from 0 to 255 reflecting the intensities of the grayscale spectrum.

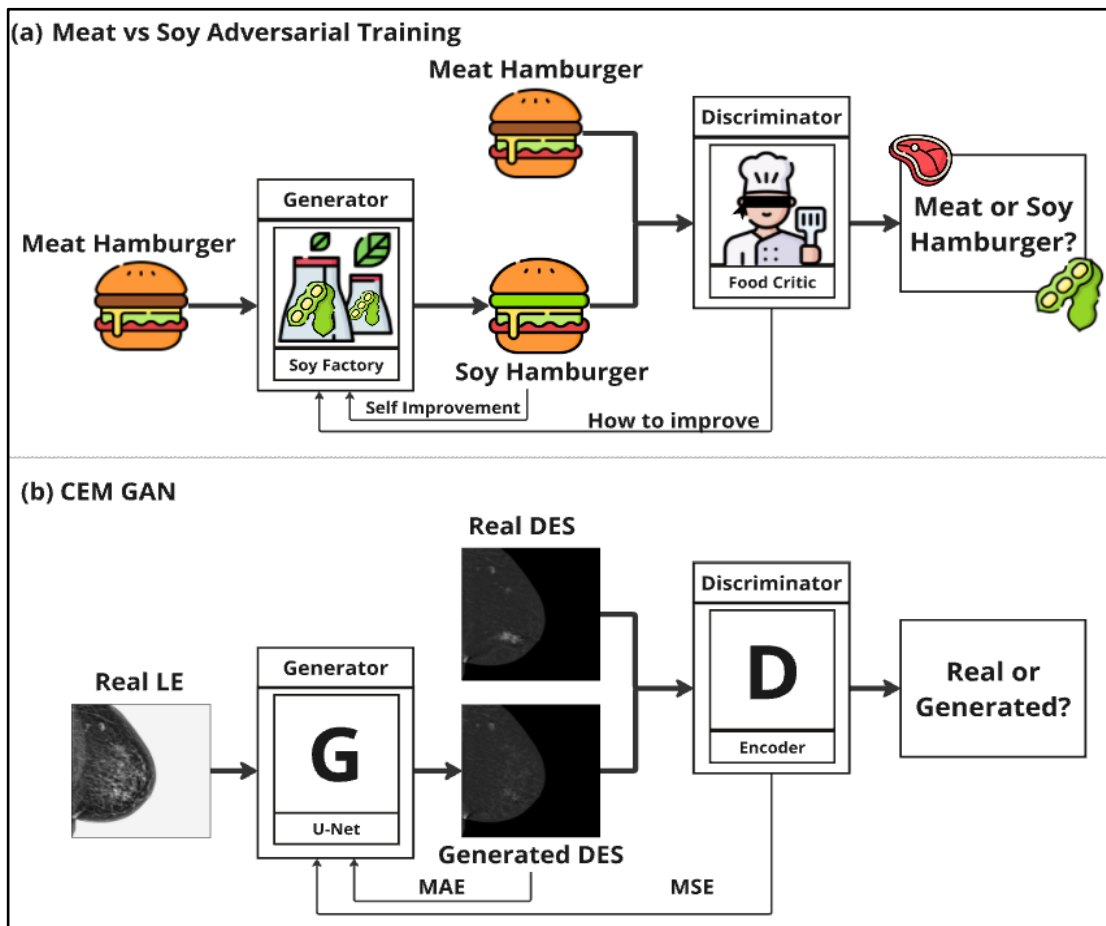
## **3.2. Train and validation distribution**

To estimate the model's performance, the study population (n=557 patients) was divided into a training group of 519 patients (93.2%) and a validation group of 38 patients (6.8%) with similar clinical characteristics. This resulted with 7593 images in the training group and 480 images in the validation group.

The validation group further stratified based on BI-RADS density categories derived from the LE images. As almost 50% of the validation subset was classified as BI-RADS category d, the images were divided into adipose (categories a-c) and dense breast (category d) subgroups. The number of patients in each subgroup is represented in table II of the results section.

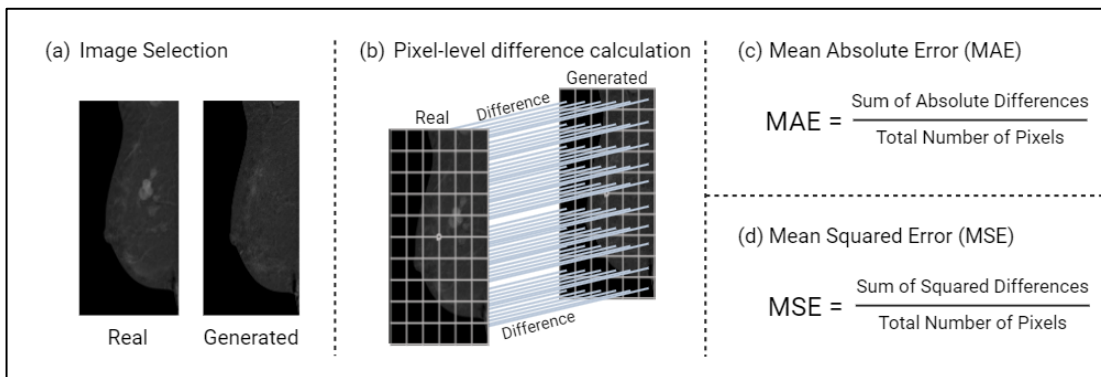
### 3.3. GAN model

We developed an architecture based on the pix2pix model, which extends the principles of GANs specifically for translating one image domain to another. The generator (G) was implemented with a U-Net architecture with skip connections to better retain essential spatial information. The implementation is illustrated in Figure 15.



**Figure 15: GAN architecture - both illustrated and realistic.** (a) Illustration representing the concept of the generator and discriminator in a GAN. (b) Scheme describing the components of the CEM GAN as well as the MAE and MSE losses involved in its training. Abbreviations: CEM, contrast-enhanced mammography; DES, dual-energy subtraction; GAN, generative adversarial network; LE, low-energy; MAE, Mean Absolute Error; MSE, Mean Squared Error.

To optimize the performance of our model, we incorporated two measures aimed at evaluating the similarity between generated images and the original ones: Mean Absolute Error (MAE) loss and Mean Squared Error (MSE) loss during training, both loss calculations are represented in Figure 16. MAE measures the average absolute differences between the predicted images and the target images. Attempting to minimize the MAE, the model aims to produce images that are structurally similar to the target. MSE calculates the average squared differences between the predicted and target images, penalizing larger errors more heavily.



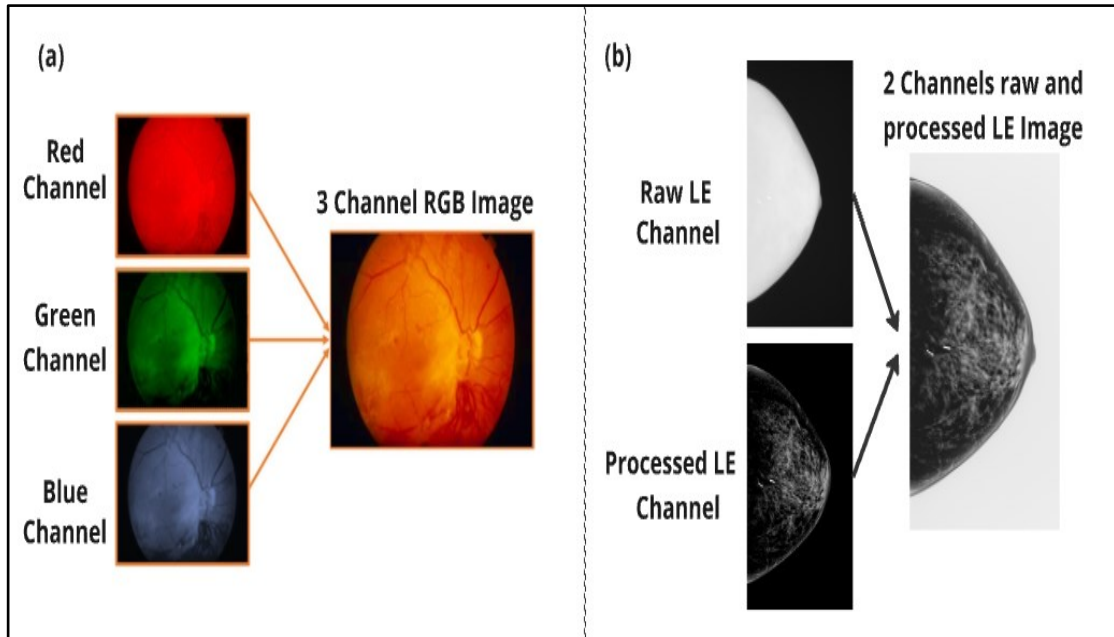
**Figure 16: Loss calculation.** (a) Real and generated paired images are selected; (b) Pixel-level differences are calculated between each pixel of the two images; (c) MAE is calculated as the sum of the absolute of each pixel difference divided by the total number of pixels; (d) MSE is calculated as the sum of the squared values of each pixel difference divided by the total number of pixels. Abbreviations: MAE, mean absolute error; MSE, mean squared error.

### 3.3.1 Stacked channel combinations

The raw LE images undergo an image processing procedure that results in processed images containing similar, though not identical, information.

Each image can be described as consisting of "channels", described in Figure 17, which represent the positions where information is stored. Grayscale images typically have a single channel, while colored images typically have stacked three (red, green, and blue), but both can represent the same object.

To explore how variations in the information affect model performance, multiple channel combinations from the original LE images were created and tested.



**Figure 17: Channels in images.** (a) Red, green, and blue channels are combined to create one three channel color RGB image. Source: *A primer on deep learning and convolutional neural networks for clinicians*. Lloret et al. (2021). (b) Raw LE image and processed LE image are combined to create a two-channel raw and processed LE image. Abbreviations: LE, low energy; RGB, red, green blue.



Three primary channel combinations were created for each image to be used as training combinations for the model:

1. **Raw Only:** A single-channel image containing only the raw LE image.
2. **Processed Only:** A single-channel image derived from the processed LE image.
3. **Raw and Processed:** A stacked two-channel image where the first channel contains the raw LE image and the second channel contains the processed LE image.

The three channel combinations are summarized in Figure 18.

### 3.3.2. Oversampling

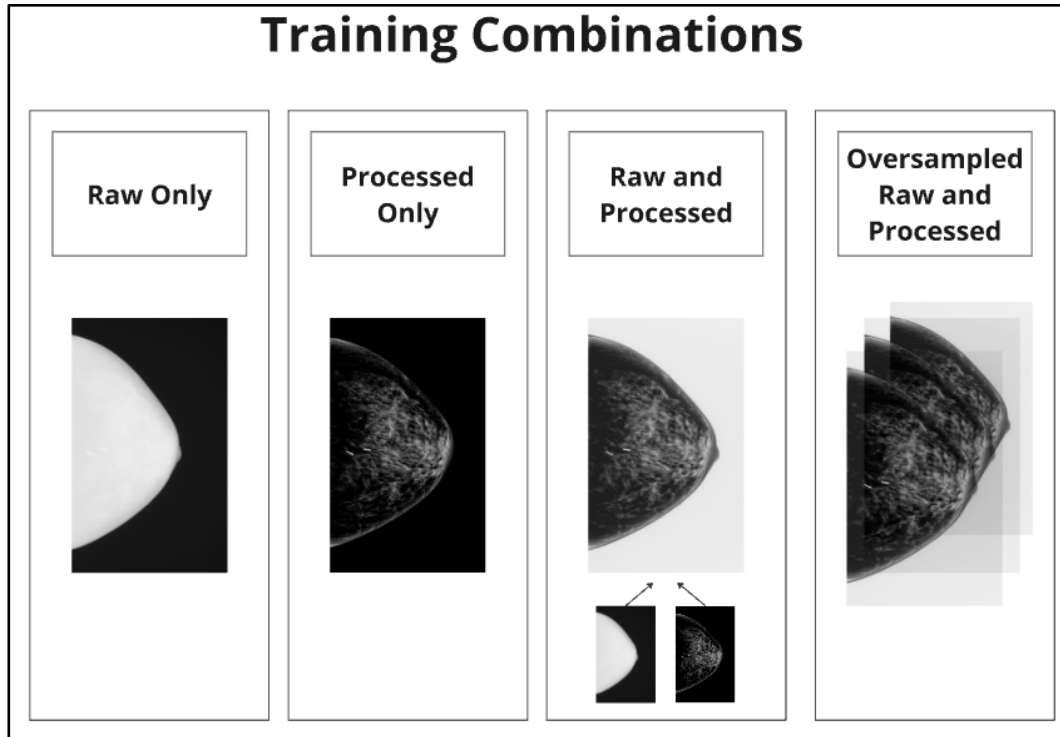
The training group had fewer “With Mass” images compared to “Without Mass” images, while the validation group had the same amount of each category. To test the impact of this imbalance, another training combination was formed by employing oversampling techniques to increase the amount of “With Mass” images. This new training combination was named “**Oversampled Raw and Processed**”.

### 3.3.3. Model training

For each training combination, a separate model was created and trained exclusively on that specific combination. The training process was conducted over 300 epochs for each model (the entire list of hyperparameters can be found in statistical analysis). Images were randomly flipped horizontally during training to expose the model to a broader range of variations within the dataset. The MAE and MSE values were continuously monitored and recorded for each epoch.

To allow faster training during exploration, the model also incorporated a preprocessing step that converted the images from their original 2048x2394 pixels resolution to 512x512 pixels <sup>69</sup>.

All the training combinations and the image types they contain are described in Figure 18.



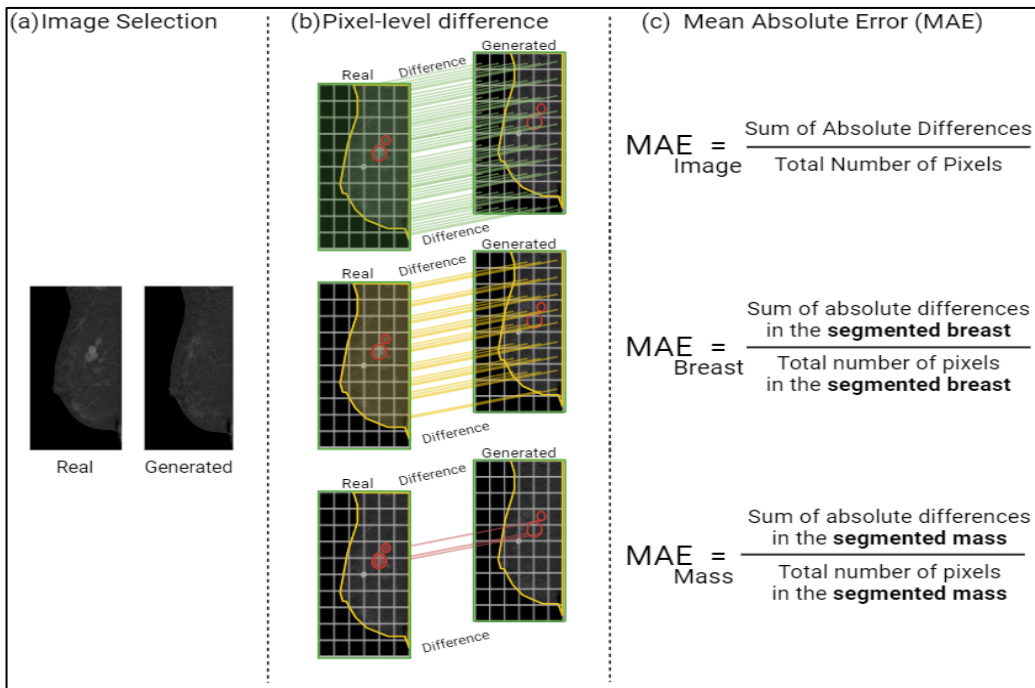
**Figure 18: Training combinations.** *Raw Only*, including only raw LE images; *Processed Only*, including only processed LE images; *Raw and Processed*: stacked two-channel image where the first channel contains the raw LE image and the second channel contains the processed LE image. *Oversampled Raw and Processed*: similar channels to “Raw and Processed”, with increased representation of “With Mass” images using oversampling techniques.

### 3.4. Model performance metrics

The primary task for each model was to generate DES images based on the corresponding LE image combination that the model was trained on. The evaluation of model performance was centered on MAE, calculated at three levels:

1. **MAE<sub>Image</sub>** - the MAE between the whole generated image and the whole real DES images.
2. **MAE<sub>Breast</sub>** - calculated as MAE between generated and real DES images, but **only for the breast area**.
3. **MAE<sub>Mass</sub>** - the MAE between the generated and real DES images in the “With Mass” category, **focused only on the mass area**.

The calculation of these performance metrics is illustrated in Figure 19.



**Figure 19: The calculation of MAE<sub>Image</sub>, MAE<sub>Breast</sub>, and MAE<sub>Mass</sub>.** (a) Image selection; (b) pixel-level difference at three levels: whole image (green), breast (yellow), and mass (red); (c) MAE calculations based. Abbreviations: MAE, mean absolute error.

To calculate the metrics for the breast area, breast segmentation was performed by applying a Gaussian Blur to the raw LE images, acquiring a segmentation mask of the breast and pectoral muscle area in the image.

For the evaluation of mass-specific performance, the "**Mass Only**" category was established, focusing exclusively on the mass regions within the "With Mass" validation images. Mass segmentation was performed by two readers.

### 3.5. Statistical analysis

The MAE of the 300<sup>th</sup> epoch, representing the final state of the model and presented as  $MAE_{Image}$ , was used to estimate the overall performance of each model on the respective channel combinations. As  $MAE_{Image}$  tends to be smaller than the other MAE-based metrics,  $MAE_{Image}$  values were multiplied by 100 and are presented as  $100 * MAE_{Image}$ .

$MAE_{Breast}$  was calculated for every segmented breast image in the "Without Mass" and "With Mass" categories.  $MAE_{Mass}$  was also calculated for each segmented mass in the "Mass Only" category.

Subsequently, averages were computed for each metric in each category, both as the overall average as well as stratified by breast density into adipose and dense breasts. Lower MAE values indicated a higher degree of similarity between the generated and original images.

The generated images of the best performing model in each analysis were presented and divided based on the variability in performance. The mean performance was termed as "Average" while the first and third quartiles were termed as "Above Average" and "Below Average" correspondingly.

The models were trained using the hyperparameters listed in Table I. For the “raw only” and “processed only” models, the input number of channels (input\_nc) was set to 1, while for the other two models, it was set to 2. The preprocess flag was defined as "resize\_and\_crop". Only the model trained on resampled raw and processed images had “resample\_dataset” set to true, while for the rest it was set to false.

Hyperparameter	Value	Hyperparameter	Value	Hyperparameter	Value
batch size	10	beta1	0.5	ndf	64
crop size	512	crop threshold	0.2	netG	"unet_256"
dataset mode	"aligned"	direction	"AtoB"	norm	"batch"
display freq	100	dropout rate	0.5	output nc	1
gan mode	"lsgan"	init gain	0.02	print_freq	100
init type	"normal"	lambda L1	100	netD	"basic"
load size	512	lr	0.0002	ngf	64
lr decay iters	50	lr policy	"linear"	num threads	2
model	"pix2pix"	n epochs	100	pool size	0
n epochs decay	200	n layers D	3	resample ratio	1

**Table I: Model training hyperparameters**

All statistical analyses were conducted using Python 3.10.14. The training was carried out on a single Nvidia V100 graphic processing unit (GPU) available through Google Cloud Platform (GCP) and required approximately 18 hours per model.

## 4. RESULTS

### 4.1. Study population characteristics

Table II presents the demographic and clinical characteristics of the study population, with a total of 557 patients divided into a train group of 519 patients (93.2%) and a validation group of 38 patients (6.8%). The median patient age was similar across both groups, with a median age of 50.0 years (IQR, 45.0-57.5) in the train group and 50.5 years (IQR, 46.0-60.75) in the validation group.

The percentage of patients in each menopausal state was similar between the groups, including premenopausal patients (Train = 45.9%, Validation = 44.7%), perimenopausal (Train = 10.2%, Validation = 10.6%) and postmenopausal (Train = 43.9%, Validation = 44.7%).

The median breast volume was 561.4 cm<sup>3</sup> (IQR = 339.2-912.8) in the train group and 557.0 cm<sup>3</sup> (IQR = 341.3-812.6) in the validation group. BI-RADS density categories were similarly distributed, with category D being the most common (Train=46.0%, Validation = 44.7%).

8073 images were analyzed, with 7593 images (94.1%) in the train group and 480 images (5.9%) in the validation group. The ratio between LE and DES images (LE:DES ratio) ratio was consistently 2:1 across both groups. Images with masses were more prevalent in the validation group (50.0%) compared to the train group (26.1%), while images without masses were more common in the train group (73.1%) than in the validation group (50.0%).

Characteristic	Train	Validation	Total
Total N° of Patients	519 (93.2%)	38 (6.8%)	557 (100%)
Median Patient Age (y)	50.0 (45.0-57.5)**	50.5 (46.0-60.75)**	50.0 (45.0-58.0)**
Menopausal State			
Premenopausal	238 (45.9%)	17 (44.7%)	255 (45.8%)
Perimenopausal	53 (10.2%)	4 (10.6%)	57 (10.2%)
Postmenopausal	228 (43.9%)	17 (44.7%)	245 (43.9%)
Median Breast Volume (cm <sup>3</sup> )	561.4 (339.2-912.8)**	557.0 (341.3-812.6)**	561.3 (341.2-912.6)**
BI-RADS Density			
A	33* (6.4%)	2 (5.3%)	35 (6.3%)
B	81* (15.6%)	7 (18.4%)	88 (15.8%)
C	166* (32.0%)	12 (31.6%)	181 (32.6%)
D	239* (46.0%)	17 (44.7%)	253 (45.5%)
Total N° of Images	7593 (94.1%)	480 (5.9%)	8073 (100.0%)
LE:DES ratio	2:1	2:1	2:1
With Mass	1980 (26.1%)	240 (50.0%)	2220 (27.5%)
Without Mass	5613 (73.1%)	240 (50.0%)	5854 (72.5%)

**Table II: Study Population.** Characteristics of the patient and image datasets included in the analysis. \*Derived from the projected percentages. \*\* Values represented as interquartile ranges. Abbreviations: BI-RADS, Breast Imaging-Reporting and Data System; LE, Low-Energy; DES, Digital Energy Subtraction; N°, number.

## 4.2. Whole-image performance comparison across training and validation

Table III presents the performance of the models on the entire image, denoted as  $\mathbf{MAE}_{\text{Image}}$ , demonstrating the difference in performance between the train and validation groups across different training combinations.

The models trained on raw images alone showed the lowest  $\mathbf{MAE}_{\text{Image}}$  in the validation group ( $100 * \mathbf{MAE}_{\text{Image}} = 1.01$ ) as well as the least difference between the train and validation groups ( $100 * \mathbf{MAE}_{\text{Image}}$ : Train=0.97, Validation=1.01; Difference=0.04).

The model trained on both raw and processed images without oversampling had the lowest MAE in the training group ( $100 * \mathbf{MAE}_{\text{Image}} = 0.81$ ) but the highest difference in  $\mathbf{MAE}_{\text{Image}}$  between train and validation among all models ( $100 * \mathbf{MAE}_{\text{Image}}$ : Train=0.81, Validation=1.17; Difference=0.36).

LE Images Combination	Group	100 * MAE <sub>Image</sub>	Difference
Only Raw	Train (n=2531)	0.97	0.04
	Validation (n=160)	1.01	
Only Processed	Train (n=2531)	1.38	0.10
	Validation (n=160)	1.48	
Raw and Processed	Train (n=5062)	0.81	0.36
	Validation (n=320)	1.17	
Oversampled Raw and Processed	Train (n=7372)	1.37	0.05
	Validation (n=320)	1.42	

**Table III: Whole image performance difference between train and validation.** The larger the MAE, the worse the performance. The oversampling was performed only on the train group aimed at having an equal amount of images with and without masses. LE, Low Energy; MAE, Mean Absolute Error.





### 4.3. Visual comparison of model outputs

Figures 20 and 21 provide a detailed comparison of the model outputs for the training and validation image groups, respectively. Each figure is structured to highlight the model performance across different channel combinations, with a focus on images both with and without masses.

Figure 20 presents the results for the training images. In the subfigure illustrating training images without masses (top half), all models demonstrated high similarity, generating images that closely resemble the desired DES outputs. However, in the subfigure showcasing images with masses (bottom half), the model trained solely on processed images did not accurately generate the tumor seen in the desired DES images. The other models produced outputs that more closely align with the DES images.

Training Images				
	LE Images Combination	LE Image Input	DES Image Desired Output	Generated Image
Images Without Mass	Raw			
	Processed			
	Raw and Processed			
	Oversampled Raw and Processed			
Images With Mass	Raw			
	Processed			
	Raw and Processed			
	Oversampled Raw and Processed			

**Figure 20: Visual comparison of model outputs on training images.** Generated images marked in red. A comparison of model outputs across four LE image combinations based on the training images. The figure is divided into two subfigures: the upper presents images without masses, and the lower shows images with masses. For each condition, the subfigure includes the LE images combination, the LE image input, the DES-transformed image, and the generated image. Abbreviations: LE, Low Energy; DES, Dual Energy Subtraction.

Figure 21 depicts the results for the validation images, following the same structural layout as Figure 20. The top subfigure includes images without masses. Similarly to the results of the training images, all models successfully generated outputs similar to the desired DES images, albeit with minor variations in the depiction of breast parenchyma. The lower subfigure, which focuses on images with masses, reveals once again similarity to the results of the training images, where only the models trained on processed images alone failed to generate images that correctly depicted the masses present in the desired DES images. Notably, the model trained on the resampled raw and processed images produced the closest match to the desired output.

These figures collectively demonstrate that the models generated more similar outputs in images without masses compared to those with masses. Furthermore, across both the training and validation groups, the model trained exclusively on processed images, produced outputs that were the least similar to the desired DES images.

Validation Images				
Images Without Mass	LE Images Combination	LE Image Input	DES Image Desired Output	Generated Image
	Raw			
	Processed			
	Raw and Processed			
	Oversampled Raw and Processed			
Images With Mass	LE Images Combination	LE Image Input	DES Image Desired Output	Generated Image
	Raw			
	Processed			
	Raw and Processed			
	Oversampled Raw and Processed			

**Figure 21: Visual comparison of model outputs on validation images.** Generated images marked in red. A comparison of model outputs across four LE image combinations based on the validation images. The figure is divided into two subfigures: the upper presents images without masses, and the lower shows images with masses. For each condition, the subfigure includes the LE images combination, the LE image input, the DES-transformed image, and the generated image. Abbreviations: LE, Low Energy; DES, Dual Energy Subtraction.

#### 4.4. Model performance on the segmented breast

Table IV summarizes the model performance on the segmented breast images from the validation group, termed  $MAE_{Breast}$ . The evaluation was conducted across four models, each trained on a different image combination. The validation group was further stratified based on two descriptors: the presence of a mass ("With Mass" or "Without Mass") and breast density (Adipose or Dense breast).

The model was evaluated using  $MAE_{Breast}$ , calculated based on the segmented breast instead of the whole image. This metric represents an error measurement, thus values which are closer to zero indicate better performance.

In the "Without Mass" category, the model trained on both raw and processed images exhibited the best performance, with the lowest  $MAE_{Breast}$  in every subgroup ( $MAE_{Breast}$ : Adipose Breast = 6.0, Dense Breast = 6.4, Overall = 6.2). This model performed particularly well in the adipose breast subgroup, where it achieved a  $MAE_{Breast}$  of 6.0. The model trained on the resampled raw and processed images was the second best performing one.

The model trained only on raw images had the lowest performance in the dense breast subgroup ( $MAE_{Breast} = 7.1$ ), which was the lowest performance overall in the "Without Mass" category ( $MAE_{Breast} = 6.8$ ). The model trained only on processed images had the lowest performance in the adipose breast subgroup ( $MAE_{Breast} = 6.6$ ).

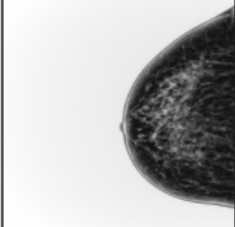
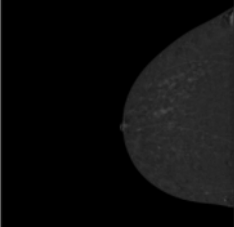
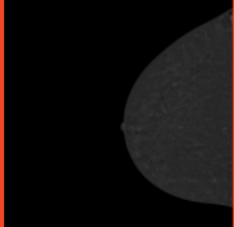
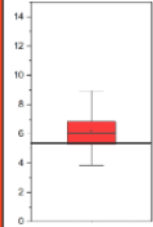
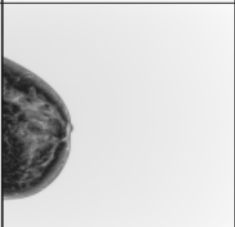
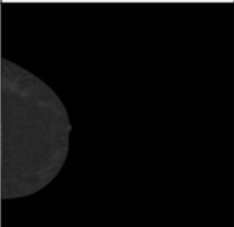
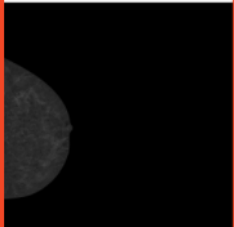
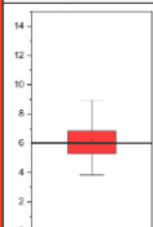
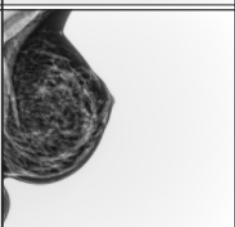
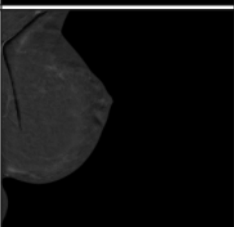
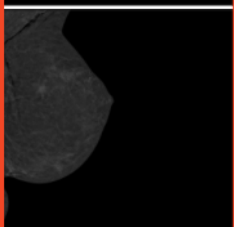
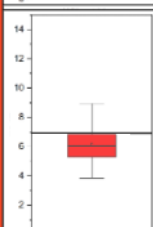
In the "With Mass" category, the model trained on raw and processed images was also the best performing, achieving the lowest  $MAE_{Breast}$  in all subgroups ( $MAE_{Breast}$ : Adipose breast = 7.3; Dense breast = 9.1, Overall = 8.2). The best performance within this category was in the adipose subgroup ( $MAE_{Breast} = 7.3$ ).

The model trained only on processed images had the lowest performance in the dense breast subgroup ( $MAE_{Breast} = 9.9$ ), which was the lowest performance in the “With Mass” category. This model also resulted in the lowest overall performance in this category ( $MAE_{Breast} = 8.9$ ).

For each model and in each of the subgroups, the  $MAE_{Breast}$  in the “With Mass” category was higher than in the “Without Mass” category.

Breast Mean Absolute Error ( $MAE_{Breast}$ )				
Breast Segmentation	LE images combination	Adipose Breast	Dense Breast	Overall
Without Mass	Only Raw	6.5	7.1	6.8
	Only Processed	6.6	6.7	6.7
	<b>Raw and Processed</b>	<b>6.0</b>	<b>6.4</b>	<b>6.2</b>
	Oversampled Raw and Processed	6.3	6.6	6.4
With Mass	Only Raw	7.5	9.6	8.6
	Only Processed	7.9	9.9	8.9
	<b>Raw and Processed</b>	<b>7.3</b>	<b>9.1</b>	<b>8.2</b>
	Oversampled Raw and Processed	7.7	9.2	8.5

**Table IV: Performance of models across different segmentation categories and image subgroups.** The table summarizes the average error per pixel for models evaluated on segmented breast images from the validation group. Models were assessed across four training combinations and further stratified into adipose breast and dense breast subgroups. The performance is grouped into “Without Mass” and “With Mass” categories. The performance was evaluated based on  $MAE_{Breast}$ . Closer to zero values indicate better performance. Best performance in each subgroup as well as overall is marked in bold. Abbreviations: LE, Low Energy; MAE: Mean Absolute Error.

The Best Performing Model				
LE images combination:		Raw and Processed		
Images Without Mass				
Performance Quality	LE Image Input	DES Image Desired Output	Generated Image	MAE Breast
Above Average				
Average				
Below Average				

**Figure 22: Images generated by the best performing model in the “Without Mass” category stratified by performance quality.** Comparison of images generated by the model trained on raw and processed images based on their  $MAE_{Breast}$ . Generated images are marked in red. The relative position of the  $MAE_{Breast}$  of the image is demonstrated by the boxplot. Abbreviations: LE, low energy; MAE, mean absolute error.

The best-performing model in the “Without Mass” category (the model trained on raw and processed images) for breast segmentations achieved a overall  $MAE_{Breast}$  of 6.2. The images it generated exhibit minimal variation in performance ( $MAE_{Breast}$ : Above Average = 5.3, Below Average = 6.8) across this category, as depicted in Figure 22.



In the “With Mass” category, the same model has an overall  $MAE_{Breast}$  of 8.2. The images generated by it display great variation, as illustrated in Figure 23. In the below average images ( $MAE_{Breast} = 8.9$ ), the model failed to generate all the expected masses. At average performance ( $MAE_{Breast} = 8.2$ ), the tumors are faintly visible, while above-average images ( $MAE_{Breast} = 6.1$ ) show them clearly.

The Best Performing Model				
LE images combination:		Raw and Processed		
Images With Mass				
Performance Quality	LE Image Input	DES Image Desired Output	Generated Image	$MAE_{Breast}$
Above Average				
Average				
Below Average				

**Figure 23: Images generated by the best performing model in the “With Mass” category stratified by performance quality.** Comparison of images generated by the model trained on raw and processed images based on their  $MAE_{Breast}$ . Generated images are marked in red. The relative position of the  $MAE_{Breast}$  of the image is demonstrated by the boxplot. Abbreviations: LE, low energy; MAE, mean absolute error.

## 4.5. Model performance in the “Mass Only” category

The performance of the models was also evaluated on mass-specific segmentation, focusing on the same density subgroups. The evaluation was based on  $MAE_{Mass}$ , as summarized in Table V.  $MAE_{Mass}$  is an error measurement, thus the lower it is the better the performance.

The model trained on the oversampled raw and processed images demonstrated the best performance in each subgroup as well as overall ( $MAE_{Mass}$ : Adipose Breast = 39.5, Dense Breast = 35.1, Overall = 37.3). Notably, the lowest  $MAE_{Mass}$  was observed in the dense breast subgroup.

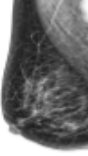
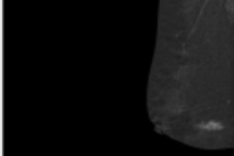
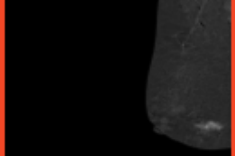
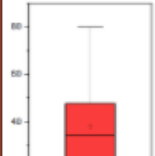


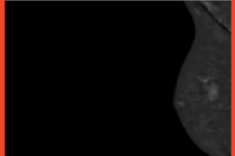
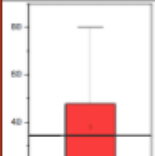
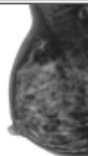
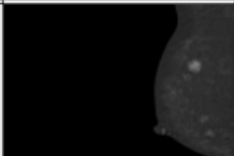

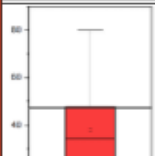
The model trained only on processed images had the highest  $MAE_{Mass}$  in the adipose breast subgroup and overall ( $MAE_{Mass}$ : Adipose Breast = 41.5, Overall = 38.4). The model trained on raw images alone had the highest  $MAE_{Mass}$  in the dense breast group ( $MAE_{Mass}$  = 36.3).

Mass Only	Mass-Only Mean Absolute Error ( $MAE_{Mass}$ )			
	LE images combination	Adipose Breast	Dense Breast	Overall
	Only Raw	39.6	36.3	38.0
	Only Processed	41.5	35.3	38.4
	Raw and Processed	40.9	35.7	38.3
	Oversampled Raw and Processed	39.5	35.1	37.3

**Table V: Model performance in the “Mass Only” category.** Performance of the models trained on the different image combinations across the four subgroups. The performance was evaluated based on  $MAE_{Mass}$ . Closer to zero values indicate better performance. The best performance in each subgroup is marked in bold. The best overall training combination is marked in bold as well. Abbreviations: LE, Low Energy; MAE, mean absolute error.

Overall, for each model and in each of the subgroups, the performance in both metrics in the “Mass Only” category was worse compared to the performance in the “Without Mass” and “With Mass” categories.

The best performing model in the “Mass Only” category (the model trained on resampled raw and processed images) achieved an overall  $MAE_{Breast}$  of 37.3. Figure 24 depicts the variability in its performance. Images with average or below-average performance ( $MAE_{Breast} = 37.3\text{--}46.0$ ) show the expected tumor, but only partially. In contrast, images with above-average performance ( $MAE_{Breast} = 21$ ) clearly show the expected tumors found in the DES images.

The Best Performing Model				
LE images combination:		Resampled Raw and Processed		
Mass Only				
Performance Quality	LE Image Input	DES Image Desired Output	Generated Image	$MAE_{Mass}$
Above Average				
Average				
Below Average				

**Figure 24: Images generated images by the best performing model in the “Mass Only” category stratified by performance quality.** Comparison of images generated by the model trained on resampled raw and processed images based on their  $MAE_{Mass}$ . Generated images are marked in red. The relative position of the  $MAE_{Mass}$  of the image is demonstrated by the boxplot. Abbreviations: LE, low energy; MAE, mean absolute error.

## 5. DISCUSSION

In this study, we provided key insights into the development and evaluation of GAN models designed to generate DES images from LE CEM images. We created and trained several models, each trained on different combinations of raw and processed LE images, which were validated on an unseen validation group of images.

The consistency between training and validation groups confirms that the models maintained their generative capacity on unseen data. This consistency is important for future applications, as models will need to perform reliably across a broad range of patient images.

Subsequently, we tested each model's performance across various categories and subgroups. The primary categories included whole image generation, segmented breast image generation (with or without mass), and mass-specific image generation. The subgroups were defined as adipose (BI-RADS category a-c) and dense (BI-RADS category d) breasts. Each model was rigorously evaluated to determine its effectiveness in generating clinically relevant features within these categories and subgroups.

One of the most significant findings was the superior performance of models trained on both raw and processed LE images. The combined-input approach allowed the model to leverage detailed information from the raw images, which often contain nuanced features that are not retained in the processed images, as reflected by the lower MAE.

The model trained with resampling techniques, designed to increase the representation of mass-containing images, demonstrated the best performance in mass-specific MAE ( $MAE_{\text{Mass}}$ ). This finding aligns with the principle that targeted resampling can improve model sensitivity to less frequent but clinically

significant features, such as masses. The enhanced performance in detecting and accurately replicating masses is a promising result for potential clinical applications, where precise mass characterization is essential for early breast cancer detection.

The presence of dense fibroglandular tissue complicates the diagnosis, as it can obscure tumors, making interpretation more challenging in dense breasts. Thus, adipose breasts with a higher fat content are generally associated with simpler mammographic interpretations<sup>70</sup>. Interestingly, in the “Mass Only” category, the model performed better on the dense breast than on the adipose one. The difference can be attributed to the average mass size, which is 2047 pixels in the adipose breast group compared to 2941 pixels in the dense breast group. As the MAE is inversely proportional to the total number of pixels, larger masses may lead to a lower relative MAE.

Building on that notion, the difference in the total number of pixels might explain the lower performance in the “Mass Only” category compared to the segmented breast categories (“Without Mass” and “With Mass”). The segmented mass area, 2000-3000 pixels, is about 20-30 times smaller than the segmented breast area, 65,000 pixels on average. As a result, the effect of the number of pixels on the MAE is even greater.

Despite these promising results, several limitations were identified in the study. While the current models demonstrated strong performance, there are additional factors that could be explored to improve the results. These include hyperparameters and clinical parameters such as breast density.

Another limitation is the size of the dataset, which was sufficient to demonstrate the model's potential, but a larger dataset together with the use of K-fold cross-validation would be necessary to reduce the risk for overfitting.

To address these limitations and further enhance the model's performance, future work should focus on integrating breast density into the training process. This approach could help optimize the model's ability to distinguish between subtle differences that are critical in clinical imaging.

Additionally, expanding the dataset and implementing K-fold cross-validation will be essential steps in reducing overfitting, validating the model's robustness and ensuring that it performs consistently across diverse patient populations.

Another limitation may be due to the reduced image resolution (512x512) used in this analysis, as the identified strengths and weaknesses might not be translatable to the original image resolution (2048x2394). As demonstrated by Sabottke and Spieler, the performance of models for radiological imaging exhibit varying performance changes relative to increased image resolution <sup>71</sup>. Thus, the best performing models should be retrained with higher resolution to verify that their quality is translatable.

Finally, before these models can be considered for clinical deployment, it is imperative to conduct clinical reader studies. These studies would involve radiologists assessing the quality and diagnostic utility of the generated DES images, thereby providing direct feedback on the clinical validity of the model. Such validation is a critical step in transitioning from research setting to practical, real-world applications.

In summary, generating DES images from LE CEM images using GANs holds great promise in assisting BC diagnosis. The generated images based on both raw and processed images show high resemblance to the original desired ones, demonstrating the potential of this method for reducing patient risks while maintaining diagnostic accuracy.

## 6. CONCLUSION

This study introduces a novel approach to generating dual-energy subtracted images from low-energy contrast-enhanced mammography images by leveraging deep learning techniques. The models developed in this study were rigorously tested across multiple categories and subgroups, demonstrating the added value of incorporating both raw and processed images.

The findings of this study suggest that, following successful clinical validation, such models could play a pivotal role in assisting breast cancer diagnosis. Moreover, the use of deep learning-generated dual-energy subtracted images could potentially offer the same diagnostic performance of contrast-enhanced mammography without the need for additional imaging procedures, thereby minimizing patient risks associated with radiation exposure and contrast media.

## 7. REFERENCES

1. Surveillance, Epidemiology, and End Results (SEER) Program. U.S. Population Data 1969-2022. Published online March 2024. Accessed May 19, 2024. <https://seer.cancer.gov/statfacts/html/breast.html>
2. Admoun C, Mayrovitz HN. The Etiology of Breast Cancer. In: Department of Medical Education, Dr. Kiran C. Patel College of Allopathic Medicine, Nova Southeastern University, FL, USA, Mayrovitz HN, eds. *Breast Cancer*. Exon Publications; 2022:21-30. doi:10.36255/exon-publications-breast-cancer-etiology
3. Li CI, Uribe DJ, Daling JR. Clinical characteristics of different histologic types of breast cancer. *Br J Cancer*. 2005;93(9):1046-1052. doi:10.1038/sj.bjc.6602787
4. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. 2024;74(1):12-49. doi:10.3322/caac.21820
5. Van Der Groep P, Bouter A, Van Der Zanden R, et al. Distinction between hereditary and sporadic breast cancer on the basis of clinicopathological data. *J Clin Pathol*. 2006;59(6):611-617. doi:10.1136/jcp.2005.032151
6. Poletes C, Amanirad B, Santiago AT, et al. The incidence of brain metastases in breast cancer according to molecular subtype and stage: a 10-year single institution analysis. *J Neurooncol*. Published online May 13, 2024. doi:10.1007/s11060-024-04707-1
7. Burguin A, Diorio C, Durocher F. Breast Cancer Treatments: Updates and New Challenges. *J Pers Med*. 2021;11(8):808. doi:10.3390/jpm11080808
8. Koo MM, Von Wagner C, Abel GA, McPhail S, Rubin GP, Lyratzopoulos G. Typical and atypical presenting symptoms of breast cancer and their associations with diagnostic intervals: Evidence from a national audit of cancer diagnosis. *Cancer Epidemiol*. 2017;48:140-146. doi:10.1016/j.canep.2017.04.010
9. Jallah JK, Dweh TJ, Anjankar A, Palma O. A Review of the Advancements in Targeted Therapies for Breast Cancer. *Cureus*. Published online October 28, 2023. doi:10.7759/cureus.47847
10. Henderson JT, Webber EM, Weyrich MS, Miller M, Melnikow J. Screening for Breast Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. Published online April 30, 2024. doi:10.1001/jama.2023.25844
11. Farrell K, Bennett DL, Schwartz TL. Screening for Breast Cancer: What You Need to Know. *Mo Med*. 2020;117(2):133-135.
12. Shah T, Guraya S. Breast cancer screening programs: Review of merits, demerits, and recent recommendations practiced across the world. *J Microsc Ultrastruct*. 2017;5(2):59. doi:10.1016/j.jmau.2016.10.002



13. Duffy SW, Tabár L, Yen AMF, et al. Beneficial Effect of Consecutive Screening Mammography Examinations on Mortality from Breast Cancer: A Prospective Study. *Radiology*. 2021;299(3):541-547. doi:10.1148/radiol.2021203935
14. Witten DM. Mammography: Advantages and Limitations. *Postgrad Med*. 1964;36(3):242-246. doi:10.1080/00325481.1964.11695254
15. Heywang-Köbrunner SH, Hacker A, Sedlacek S. Advantages and Disadvantages of Mammography Screening. *Breast Care*. 2011;6(3):2-2. doi:10.1159/000329005
16. Olsen AH, Schwartz W, Vejborg IMM. [Mammography screening--advantages and disadvantages]. *Ugeskr Laeger*. 2007;169(36):2977-2979.
17. Fico N, Grezia GD, Cuccurullo V, et al. Breast Imaging Physics in Mammography (Part II). *Diagnostics*. 2023;13(23):3582. doi:10.3390/diagnostics13233582
18. Fico N, Di Grezia G, Cuccurullo V, et al. Breast Imaging Physics in Mammography (Part I). *Diagnostics*. 2023;13(20):3227. doi:10.3390/diagnostics13203227
19. Boyd NF, Guo H, Martin LJ, et al. Mammographic Density and the Risk and Detection of Breast Cancer. *N Engl J Med*. 2007;356(3):227-236. doi:10.1056/NEJMoa062790
20. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and Combined Effects of Age, Breast Density, and Hormone Replacement Therapy Use on the Accuracy of Screening Mammography. *Ann Intern Med*. 2003;138(3):168. doi:10.7326/0003-4819-138-3-200302040-00008
21. Al Mousa DS, Ryan EA, Mello-Thoms C, Brennan PC. What effect does mammographic breast density have on lesion detection in digital mammography? *Clin Radiol*. 2014;69(4):333-341. doi:10.1016/j.crad.2013.11.014
22. Bhimani C, Matta D, Roth RG, et al. Contrast-enhanced Spectral Mammography. *Acad Radiol*. 2017;24(1):84-88. doi:10.1016/j.acra.2016.08.019
23. Izumi Y, Ohara M, Fujii K, Yokoya A, Ogawa M. X-ray photoemission and absorption spectroscopy of a hypervalent iodine compound, 2-iodosobenzoic acid. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater At*. 2024;547:165211. doi:10.1016/j.nimb.2023.165211
24. Lobbes MBI, Smidt ML, Houwers J, Tjan-Heijnen VC, Wildberger JE. Contrast enhanced mammography: techniques, current results, and potential indications. *Clin Radiol*. 2013;68(9):935-944. doi:10.1016/j.crad.2013.04.009
25. Dromain C, Thibault F, Diekmann F, et al. Dual-energy contrast-enhanced digital mammography: initial clinical results of a multireader, multicase study. *Breast Cancer Res BCR*. 2012;14(3):R94. doi:10.1186/bcr3210
26. Yang ML, Bhimani C, Roth R, Germaine P. Contrast enhanced

- mammography: focus on frequently encountered benign and malignant diagnoses. *Cancer Imaging*. 2023;23(1):10. doi:10.1186/s40644-023-00526-1
27. Coffey K, Jochelson MS. Contrast-enhanced mammography in breast cancer screening. *Eur J Radiol*. 2022;156:110513. doi:10.1016/j.ejrad.2022.110513
  28. Lalji UC, Houben IPL, Prevos R, et al. Contrast-enhanced spectral mammography in recalls from the Dutch breast cancer screening program: validation of results in a large multireader, multicase study. *Eur Radiol*. 2016;26(12):4371-4379. doi:10.1007/s00330-016-4336-0
  29. Houben IPL, Van De Voorde P, Jeukens CRLPN, et al. Contrast-enhanced spectral mammography as work-up tool in patients recalled from breast cancer screening has low risks and might hold clinical benefits. *Eur J Radiol*. 2017;94:31-37. doi:10.1016/j.ejrad.2017.07.004
  30. Gennaro G, Cozzi A, Schiaffino S, Sardanelli F, Caumo F. Radiation Dose of Contrast-Enhanced Mammography: A Two-Center Prospective Comparison. *Cancers*. 2022;14(7):1774. doi:10.3390/cancers14071774
  31. Bottinor W, Polkampally P, Jovin I. Adverse Reactions to Iodinated Contrast Media. *Int J Angiol*. 2013;22(03):149-154. doi:10.1055/s-0033-1348885
  32. Heiken JP. Contrast safety in the cancer patient: preventing contrast-induced nephropathy. *Cancer Imaging*. 2008;8(Special Issue A):S124-S127. doi:10.1102/1470-7330.2008.9018
  33. Kamitani T, Yabuuchi H, Kanemaki Y, et al. Effects of menstrual cycle on background parenchymal enhancement and detectability of breast cancer on dynamic contrast-enhanced breast MRI: A multicenter study of an Asian population. *Eur J Radiol*. 2019;110:130-135. doi:10.1016/j.ejrad.2018.11.025
  34. Rousset-Jablonski C, Gompel A. Screening for familial cancer risk: Focus on breast cancer. *Maturitas*. 2017;105:69-77. doi:10.1016/j.maturitas.2017.08.004
  35. Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. BI-RADS® fifth edition: A summary of changes. *Diagn Interv Imaging*. 2017;98(3):179-190. doi:10.1016/j.diii.2017.01.001
  36. D'Orsi C, Sickles E, Mendelsohn EB. *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*.; 2013.
  37. Christopher L. Vaughan. New developments in medical imaging to detect breast cancer. *CME*. 2012;30:12-15.
  38. Posso M, Louro J, Sánchez M, et al. Mammographic breast density: How it affects performance indicators in screening programmes? *Eur J Radiol*. 2019;110:81-87. doi:10.1016/j.ejrad.2018.11.012
  39. Holland K, Van Zelst J, Den Heeten GJ, et al. Consistency of breast density categories in serial screening mammograms: A comparison

- between automated and human assessment. *The Breast*. 2016;29:49-54. doi:10.1016/j.breast.2016.06.020
40. Highnam R, Jeffreys M, McCormack V, Warren R, Davey Smith G, Brady M. Comparing measurements of breast density. *Phys Med Biol*. 2007;52(19):5881-5895. doi:10.1088/0031-9155/52/19/010
  41. Shad R, Cunningham JP, Ashley EA, Langlotz CP, Hiesinger W. Designing clinically translatable artificial intelligence systems for high-dimensional medical imaging. *Nat Mach Intell*. 2021;3(11):929-935. doi:10.1038/s42256-021-00399-8
  42. Alamir M, Alghamdi M. The Role of Generative Adversarial Network in Medical Image Analysis: An In-depth Survey. *ACM Comput Surv*. 2023;55(5):1-36. doi:10.1145/3527849
  43. Van Der Zant T, Kouw M, Schomaker L. Generative Artificial Intelligence. In: Müller VC, ed. *Philosophy and Theory of Artificial Intelligence*. Vol 5. Studies in Applied Philosophy, Epistemology and Rational Ethics. Springer Berlin Heidelberg; 2013:107-120. doi:10.1007/978-3-642-31674-6\_8
  44. Yu B, Zhou L, Wang L, Shi Y, Fripp J, Bourgeat P. Ea-GANs: Edge-Aware Generative Adversarial Networks for Cross-Modality MR Image Synthesis. *IEEE Trans Med Imaging*. 2019;38(7):1750-1762. doi:10.1109/TMI.2019.2895894
  45. Gupta R, Sharma A, Kumar A. Super-Resolution using GANs for Medical Imaging. *Procedia Comput Sci*. 2020;173:28-35. doi:10.1016/j.procs.2020.06.005
  46. Ghorbani A, Natarajan V, Coz D, Liu Y. *DermGAN: Synthetic Generation of Clinical Skin Images with Pathology.*; 2020:155-170.
  47. Dikici E, Bigelow M, White RD, Erdal BS, Prevedello LM. Constrained generative adversarial network ensembles for sharable synthetic medical images. *J Med Imaging*. 2021;8(02). doi:10.1117/1.JMI.8.2.024004
  48. Zhang Q, Wang H, Lu H, Won D, Yoon SW. Medical Image Synthesis with Generative Adversarial Networks for Tissue Recognition. In: *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE; 2018:199-207. doi:10.1109/ICHI.2018.00030
  49. Lyu Q, Wang G. Conversion Between CT and MRI Images Using Diffusion and Score-Matching Models. Published online 2022. doi:10.48550/ARXIV.2209.12104
  50. Cahan N, Klang E, Aviram G, et al. X-ray2CTPA: Generating 3D CTPA scans from 2D X-ray conditioning. Published online 2024. doi:10.48550/ARXIV.2406.16109
  51. Montoya-del-Angel R, Sam-Millan K, Vilanova JC, Martí R. MAM-E: Mammographic Synthetic Image Generation with Diffusion Models. *Sensors*. 2024;24(7):2076. doi:10.3390/s24072076
  52. Zakka C, Saheb G, Najem E, Berjawi G. MammoGANesis: Controlled Generation of High-Resolution Mammograms for Radiology Education.

- Published online 2020. doi:10.48550/ARXIV.2010.05177
53. Oyelade ON, Ezugwu AE, Almutairi MS, Saha AK, Abualigah L, Chiroma H. A generative adversarial network for synthetization of regions of interest based on digital mammograms. *Sci Rep.* 2022;12(1):6166. doi:10.1038/s41598-022-09929-9
  54. Romano A, Rossi-Espagnet MC, Pasquini L, et al. Cerebral Venous Thrombosis: A Challenging Diagnosis; A New Nonenhanced Computed Tomography Standardized Semi-Quantitative Method. *Tomography.* 2021;8(1):1-9. doi:10.3390/tomography8010001
  55. Henrik S. Thomsen. ESUR Guidelines on Contrast Agents v.10.0. Published online March 2018. Accessed September 5, 2024. [https://www.esur.org/wp-content/uploads/2022/03/ESUR-Guidelines-10\\_0-Final-Version.pdf](https://www.esur.org/wp-content/uploads/2022/03/ESUR-Guidelines-10_0-Final-Version.pdf)
  56. Pasquini L, Napolitano A, Pignatelli M, et al. Synthetic Post-Contrast Imaging through Artificial Intelligence: Clinical Applications of Virtual and Augmented Contrast Media. *Pharmaceutics.* 2022;14(11):2378. doi:10.3390/pharmaceutics14112378
  57. Han S, Kim JM, Park J, et al. Clinical feasibility of deep learning based synthetic contrast enhanced abdominal CT in patients undergoing non enhanced CT scans. *Sci Rep.* 2024;14(1):17635. doi:10.1038/s41598-024-68705-z
  58. Armanious K, Jiang C, Fischer M, et al. MedGAN: Medical image translation using GANs. *Comput Med Imaging Graph.* 2020;79:101684. doi:10.1016/j.compmedimag.2019.101684
  59. Kim E, Cho HH, Kwon J, Oh YT, Ko ES, Park H. Tumor-Attentive Segmentation-Guided GAN for Synthesizing Breast Contrast-Enhanced MRI Without Contrast Agents. *IEEE J Transl Eng Health Med.* 2023;11:32-43. doi:10.1109/JTEHM.2022.3221918
  60. Osuala RO, Joshi S, Tsirikoglou A, et al. Pre- to post-contrast breast MRI synthesis for enhanced tumour segmentation. In: Colliot O, Mitra J, eds. *Medical Imaging 2024: Image Processing.* SPIE; 2024:31. doi:10.1117/12.3006961
  61. Zhao J, Mathieu M, LeCun Y. Energy-based Generative Adversarial Network. Published online 2016. doi:10.48550/ARXIV.1609.03126
  62. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. Published online 2014. doi:10.48550/ARXIV.1406.2661
  63. Kingma DP, Welling M. Auto-Encoding Variational Bayes. Published online 2013. doi:10.48550/ARXIV.1312.6114
  64. Rezende DJ, Mohamed S, Wierstra D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. Published online 2014. doi:10.48550/ARXIV.1401.4082
  65. Goodfellow I. NIPS 2016 Tutorial: Generative Adversarial Networks. Published online 2017. doi:10.48550/ARXIV.1701.00160

66. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. Published online 2016. doi:10.48550/ARXIV.1611.07004
67. Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. Published online 2017. doi:10.48550/ARXIV.1711.11585
68. Gong M, Chen S, Chen Q, Zeng Y, Zhang Y. Generative Adversarial Networks in Medical Image Processing. *Curr Pharm Des*. 2021;27(15):1856-1868. doi:10.2174/1381612826666201125110710
69. Ju RY, Wong K, Chiang JS. Efficient GANs for Document Image Binarization Based on DWT and Normalization. Published online July 4, 2024. Accessed September 12, 2024. <http://arxiv.org/abs/2407.04231>
70. Soguel L, Durocher F, Tchernof A, Diorio C. Adiposity, breast density, and breast cancer risk: epidemiological and biological considerations. *Eur J Cancer Prev*. 2017;26(6):511-520. doi:10.1097/CEJ.0000000000000310
71. Sabottke CF, Spieler BM. The Effect of Image Resolution on Deep Learning in Radiography. *Radiol Artif Intell*. 2020;2(1):e190015. doi:10.1148/ryai.2019190015

## 8. RINGRAZIAMENTI

I would like to express my deepest gratitude to my wife, who has been my partner in every step of this journey. Your unwavering support and encouragement mean the world to me.

To my family, thank you for always being there, no matter the circumstances. Your love and understanding have carried me through even the toughest moments.

A special thank you to Gisella, for giving me the opportunity to work on this amazing project. Your guidance and trust have been invaluable.

*"We are the champions, my friends*

*And we'll keep on fighting till the end*

*We are the champions*

*We are the champions*

*No time for losers*

*'Cause we are the champions of the World"*