Craniocaudal Mammograms Generation using Image-to-Image Translation Techniques

Valentina Piras, Amedeo F. Bonatti, Carmelo De Maria, Paolo Cignoni, and Francesco Banterle *†‡§

January 15, 2026

Abstract

Breast cancer is the leading cause of cancer death in women worldwide, emphasizing the need for prevention and early detection. Mammography screening plays a crucial role in secondary prevention, but large datasets of referred mammograms from hospital databases are hard to access due to privacy concerns, and publicly available datasets are often unreliable and unbalanced. We propose a novel workflow using a statistical generative model based on generative adversarial networks to generate high-resolution synthetic mammograms. Utilizing a unique 2D parametric model of the compressed breast in craniocaudal projection and image-to-image translation techniques, our approach allows full and precise control over breast features and the generation of both normal and tumor cases. Quality assessment was conducted through visual analysis, and statistical analysis using the first five statistical moments. Additionally a questionnaire was administered to 45 medical experts (radiologists and radiology residents). The results showed that the features of the real mammograms were accurately replicated in the synthetic ones, the image statistics overall correspond reasonably well, and the two groups of images were statistically indistinguishable in almost all cases according to the experts. The proposed workflow generates realistic synthetic mammograms with fine-tuned features. Synthetic mammograms are powerful tools that can create new or balance existing datasets, allowing for the training of machine learning and deep learning algorithms. These algorithms can then assist radiologists in tasks like classification and segmentation, improving diagnostic performance. The code and dataset are available at: `https://github.com/cnr-isti-vclab/CC-Mammograms-Generation_GUI`.

Keywords: Breast cancer, breast parametric model, deep learning, mammogram synthesis.

†Valentina Piras is with the Department of Computer Science of the University of Pisa, Pisa, Italy, and with ISTI-CNR, Pisa, Italy (e-mail: valentina.piras@phd.unipi.it).

‡Amedeo F. Bonatti and Carmelo De Maria are with the Department of Information Engineering and Research Center E. Piaggio of the University of Pisa, Pisa, Italy (e-mail: amedeo.bonatti@ing.unipi.it; carmelo.demaria@unipi.it).

§Paolo Cignoni and Francesco Banterle are with ISTI-CNR, Pisa, Italy (e-mail: paolo.cignoni@isti.cnr.it; francesco.banterle@isti.cnr.it).

1 Introduction

Breast cancer is the most commonly diagnosed cancer and the leading cause of cancer-related death in women, with 2.3 million new cases worldwide in 2020 (11.7% of all cancers) [1, 2]. In the US, one in eight women will develop invasive breast cancer, and 25–30% of these cases result in death [3, 4]. Given

the high incidence of breast cancer, prevention and early detection are two critical factors in decreasing the mortality rate of such cancer. Therefore, mammography screening is the most important secondary prevention example [5] that plays a crucial role in breast cancer early detection and diagnosis [6]. With the outbreak of the COVID-19 pandemic, screening mammograms have decreased and, consequently, breast cancer diagnoses [7]. These data show that much research is needed in early detection to find solutions that physicians can couple with mammography screening. Mammograms can be used to train new machine learning (ML) algorithms that can assist radiologists in tasks like classification and segmentation, improving their diagnostic performance. To achieve high-quality results, there is the need to access a large dataset of reliable images. However, hospitals and clinics, which have a large number of available images, typically do not release such images to the public for patient privacy reasons. Furthermore, gaining access to such data necessitates special permissions requiring a long time to be granted, which can take up to a year in some cases.

In this context, we propose the following **main contributions**.

1) We present a novel pipeline within the realm of statistical generative models [8], based on generative adversarial networks (GANs), allowing for the generation of synthetic mammograms with full control for the first time. Using our new 2D parametric breast model, implemented in a graphical user interface (GUI), users can choose breast shape with five parameters, add skin thickening and control its thickness, determine the shape and position of the fibroglandular tissue region, and specify the size and location of benign and malignant solid tumors. Previous statistical generative models based on GANs for mammogram synthesis do not allow any form of user control. This capability is fundamental for creating balanced datasets.

2) We show how to adapt pix2pix [9] for our method, including a super-resolution training to generate high-quality images. This is particularly important, as physicians typically work with high-resolution mammograms.

To perform a quality assessment of the generated mammograms, we conducted a visual analysis, examined the statistics of synthetic and real mammograms through the plot of the first five statistical moments, and administered a questionnaire comparing real and synthetic mammograms to a panel of 45 radiologists and radiology residents. The following sections are organized as: Sec. 1.1 introduces breast anatomy; Sec. 2 describes the main works in the state-of-the-art; Sec. 3 proposes our approach and dataset; Sec. 4 shows our results and discusses them; Sec. 5 concludes our work. In addition, further details are available in the supplementary material.

1.1 Radiological Breast Anatomy

In several countries, breast cancer screening typically uses two-view mammography [10]: the mediolateral oblique (MLO) view and the craniocaudal (CC) view, both taken with breast compression. The MLO view is from the center of the chest outward, while the CC view is from above the breast. Since this article focuses on CC mammographic images, this section reviews the key breast regions visible in these images.

The female mammary gland comprises three components: fibrous tissue, glandular tissue (collectively called fibroglandular tissue; see Fig. 1.E), and adipose tissue (see Fig. 1.D). In Fig. 1, the CC images show normal mammographic anatomy. Adipose tissue appears radiolucent (Fig. 1.D), while fibroglandular tissue is radiopaque (Fig. 1.E). The skin, a thin, uniform radiopaque border ($\sim$ 1 mm thick; Fig. 1.C), contrasts with the subcutaneous fat's radiolucency (Fig. 1.D). The areola (Fig. 1.B) is 3-5 mm thick, with the nipple represented by a central cylindrical opacity (Fig. 1.A). Milk ducts are usually not visible. Fibroglandular tissue, shown in red in Fig. 1, assumes an almost triangular shape, located in the upper CC view with its apex towards the nipple. Its shape and density vary with individuals and hormonal factors. Behind this tissue, adipose tissue outlines the retromammary space, separating the breast from the pectoralis major muscle. The muscle, though radiopaque, is not always visible in the CC view. Blood vessels (Fig. 1.F) are more visible in adipose-rich breasts, appearing as thin, tortuous linear opacities [11].

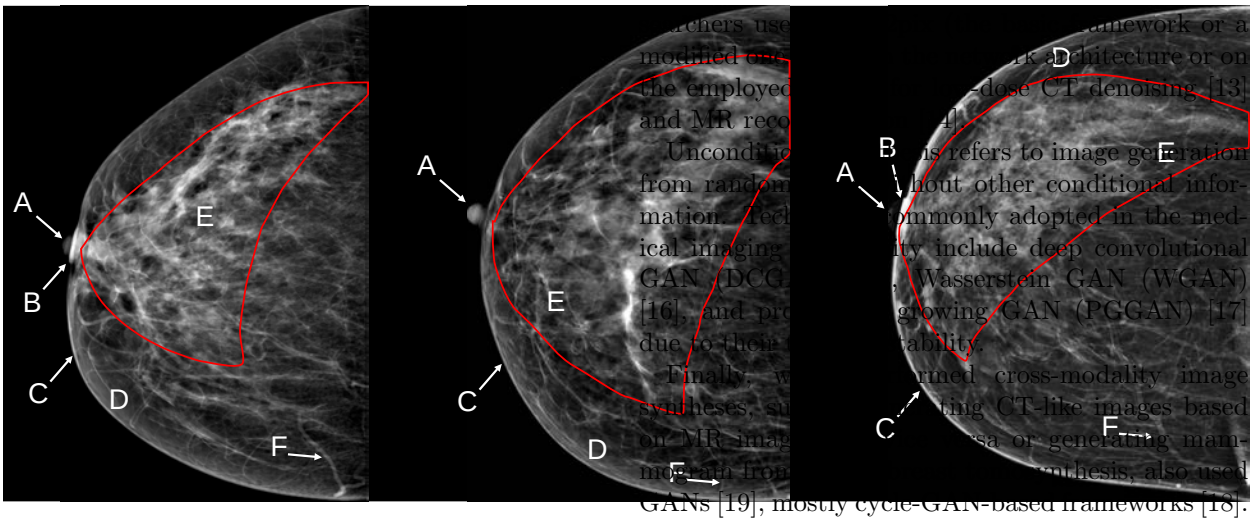


Figure 1: Examples of mammographic CC projections of normal right breasts. A: nipple; B: areola; C: skin; D: subcutaneous fat; E: fibroglandular tissue, limited by the red curve; F: blood vessel.

2 Related work

Techniques for synthetic data generation can be broadly classified into three categories: statistical generative modeling, physics-based modeling, and hybrid physics-informed modeling [8]. Our work falls within the category of statistical generative models, specifically GANs. Nonetheless, we will discuss some breast physical models to provide a more comprehensive overview.

In medical imaging, I2I tools that rely on GANs [12] are widely used for synthetic image generation [13–18], mainly focusing on computed tomography (CT) and magnetic resonance (MR) images. Depending on the approach used, we can identify three main categories of generation exploiting different I2I tools: conditional image synthesis, unconditional image synthesis, and cross-modality image synthesis.

Regarding conditional image synthesis, most of the studies use pix2pix, a conditional GAN (cGAN) framework proposed by Isola *et al.* [9], in which the conditional input is an image related to the desired output; e.g., similar edges or areas. For instance, re-

GANs [19], mostly cycle-GAN-based frameworks [18].

Regarding studies aiming to generate realistic mammographic images, we provide examples for two of the three categories of techniques for synthetic data generation [8]: physical models and statistical generative models.

Starting from physical models, Barufaldi *et al.* [20] proposed the OpenVCT, an open-source software designed to simulate breast anatomy, mammography imaging, and image interpretation for conducting Virtual Clinical Trials (VCTs). This software was used in the study by Tomic *et al.* [21], a growing tumor model for breast cancer analysis. OpenVCT was also used by Barufaldi *et al.* [22], who proposed a multi-scale Perlin noise method to simulate anatomical structures of breast tissue in a breast phantom. With this phantom, digital mammography and tomosynthesis projections were synthesized using computer simulations of a clinical system. Other digital breast phantoms were developed by Graff [23] and Sarno *et al.* [24] while growing breast lesion models were proposed by de Sisternes *et al.* [25] and Sengupta *et al.* [26]. Sizikova *et al.* [27] generated a simulated image dataset for comparative analysis of mammography artificial intelligence (AI). These last two studies both used the publicly available Virtual Imaging Clinincal Trial for Regulatory Evaluation (VICTRE) [28] to obtain synthetic digital mammograms. Li *et al.* [29, 30] introduced a method to mathematically model small and medium-scale fibroglandular and intra-glandular adipose tissues present in the

center part of the breast. Nevertheless, this model does not consider pathological breast variations, it lacks texture deformation and was derived from a limited sample size of only 16 patients. As synthetic data generation using physical modelling typically includes two components, a digital model representing the patient and a digital model of the imaging system, we included the work of Badal *et al.* [31] that described the Monte Carlo (MC) x-ray transport simulation code that was developed to replicate the mammography and digital breast tomosynthesis (DBT) devices in the VICTRE project. Although physical models can produce high-quality, detailed mammograms that are controlled by input parameters, these methods often demand more time and computational resources compared to statistical generative models. Furthermore, physical modeling approaches are limited by the variability within the parameter space of the digital human model and the acquisition system [8].

Moving on to statistical generative models, Korkinof *et al.* [32, 33] proposed a method, MammoGAN, using progressive GANs to synthesize high-resolution mammograms. Although they provided useful attributes, they do not provide information regarding the shape of the breast, the shape of the fibroglandular tissue, and the presence of tumors. Additionally, it is not clear whether the images generated by MammoGAN are limited to healthy breasts or also include breasts with tumors. Shen *et al.* [34] introduced a new method utilizing GANs to create various mass images, followed by contextual infilling to integrate synthetic lesions into real healthy screening mammograms. Although the authors focused on synthesizing malignant masses, the method is constrained by the availability of real normal mammograms. Additionally, users do not have the option to select features of the tumor mass, such as size and location within the breast. Yamazaki *et al.* [35] conducted cross-modality image synthesis, generating CC-view mammograms from MLO views using a complete representation GAN (CR-GAN). Since it utilizes real mammograms as input, this work diverges from the goal of our paper, which is to generate artificial mammograms without being conditioned on real data. The same applies to Lee *et al.* [36], who developed a cGAN to conduct a cross-modality image synthesis. They simulated the contralateral mammogram using a single-sided mammogram (either left or right) as input. Tripathi *et al.* [37] proposed a GAN architecture for segmentation-driven synthetic mammogram generation. However, in this case as well, segmentation is conditioned on real mammograms. Furthermore, the images generated in the article appear to be of low resolution and quality. Shah *et al.* [38] analyzed the reliability of synthetic mammogram images produced by DCGAN and emphasized the pressing need for advancements in the development of synthetic mammogram images, focusing on both quality and authenticity.

The ultimate goal of all these works is to construct large and balanced datasets of synthetic mammograms that can be used for training ML and deep learning (DL) algorithms, such as those proposed by Wu *et al.* [39] and Li *et al.* [40]. These algorithms can assist radiologists in the classification and segmentation of images, ultimately improving their diagnostic performance. Although traditional computer-aided detection (CAD) systems are commonly utilized by radiologists to aid in image interpretation, numerous multicenter studies have demonstrated that these CAD programs do not enhance their diagnostic accuracy [41].

Compared to these existing GAN-based methods, our work offers the following key advantages:

1) Once the networks are trained, our method enables not only the balancing of existing mammography datasets but also the generation of entirely new datasets composed of synthetic mammograms. This eliminates the need for real mammograms as a reference, which is a limitation in the works of Shen *et al.* [34], Yamazaki *et al.* [35], Lee *et al.* [36], and Tripathi *et al.* [37].

2) During the generation of synthetic mammograms, our method allows complete control over the features of the generated image, including breast size, the shape and size of the fibroglandular tissue area, the presence or absence of skin thickening, and the presence or absence of a controllable number of benign and/or malignant tumors, as well as the size and location of these tumors. The inclusion of all these features together has not been demonstrated in any

of the above-mentioned GAN-based methods.

3) Our code is openly available as Open Source, which is not the case for any of the GAN-based methods cited above. This openness is essential for fostering transparency, reproducibility, and collaboration within the scientific community. By making our implementation accessible, we aim to empower researchers and practitioners to adapt and extend our approach for further advancements in the field of synthetic mammogram generation.

This combination of flexibility, independence from real datasets, and commitment to openness significantly enhances the practicality and impact of our method, addressing key gaps in existing GAN-based approaches.

Regarding the parametric modeling of the breast, some parametric breast models exist in the literature, such as those of Chen *et al.* [42] and Gallo *et al.* [43]. However, they do not fit our needs because they are 3D models of a non-compressed breast and are characterized by a large number of parameters. In our model, we wanted to keep the number of parameters small to allow the GUI to be as simple and intuitive as possible for users. Furthermore, the model we aimed to develop is 2D and specific to the compressed breast in a particular mammographic projection.

3 Our Approach

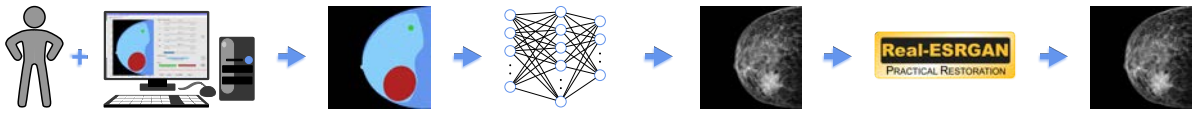
We started by selecting CC view mammograms from a source dataset, creating a 'selected dataset' that included both normal mammograms and mammograms with findings (benign and/or malignant). We processed these images to extract breast profiles, establishing fitting parameters. Using this data, we developed a 2D breast parametric model implemented in a GUI that allows users to control breast shape, insert fibroglandular tissue, solid tumors (benign and/or malignant), and skin thickening (see Fig. 2.a). With our GUI, we segmented the selected dataset into various regions (nipple, fibroglandular tissue region, malignant and benign masses, skin thickening, and the rest of the breast; see Fig. 2(b)), and used these labels to train the pix2pix I2I network [9] for synthetic 256×256 mammogram generation; see Fig. 2(d)). Finally, we fine-tuned Real-ESRGAN [44] to upscale the images to 1024×1024 pixels; see Fig. 2(f). In Fig. 2, the pipeline for generating synthetic mammograms is illustrated.

3.1 Data

For our research, we used the Categorized Digital Database for Low energy and Subtracted Contrast Enhanced Spectral Mammography images (CDD-CESM) [45] publicly available in the Cancer Imaging Archive [46], which includes mammograms with segmentation annotations for abnormal findings and diagnostic labels. As noted in Section 1, screening mammograms include CC and MLO views. Given an equal amount of data in both views in the source dataset, we focused on the more symmetrical CC view to simplify parameter extraction for the 2D breast model. After selecting the CC view as our reference we collaborated with a breast radiologist to filter out unsuitable mammograms. We excluded images with acquisition errors (e.g., skin folds, nipple within the breast profile), non-screening mammograms (e.g., postoperative with scars or metal clips), and those with significant breast distortion or nipple retraction. We also excluded mammograms with malignant microcalcifications because their irregular shapes are difficult to model and require a higher resolution than what our study achieved. Similarly, we discarded mammograms with benign calcifications, as these also necessitate higher resolution despite their regular shapes. Additionally, we excluded cases where cancer affected the entire breast, as these are rare and not representative of typical cancer appearances in screening mammograms. Thus, in addition to normal mammograms, we included those showing skin thickening, benign solid tumors (sharp-edged benign opacities), and malignant solid tumors (see Fig. 4), thereby creating our selected dataset (163 images).

3.2 2D Breast Parametric Model

In this initial phase (see Fig. 2(a)), we developed a 2D parametric model of the compressed breast in the CC view. Our GUI allows users to adjust breast



(a) User + 2D Breast Parametric Model | (b) Label Map | (c) I2I Network (pix2pix) | (d) Low-Resolution Synthetic Mammogram | (e) Super-resolution Network (Real-ESRGAN) | (f) High-Resolution Synthetic Mammogram

Figure 2: Our pipeline for generating synthetic mammograms. First, the user creates labels using our GUI, controlling breast shape, fibroglandular region, nipple position, tumor size and location (both malignant and benign), and skin thickening (a). These labels (b) are then converted into low-resolution mammograms (256×256) (d) using pix2pix (c). Finally, these mammograms are upsampled by Real-ESRGAN (e), obtaining high-resolution ones (1024×1024) (f).

shape parameters, add skin thickening, and insert both benign and malignant solid tumors. The selected dataset included CC mammograms of both right breasts (nipple facing left) and left breasts (nipple facing right). To ensure consistency and avoid redundancy, we mirrored all left breast mammograms so that all images show the nipple facing left. This augmentation created a uniform dataset with consistent nipple orientation. Next, we processed the mammograms to extract the breast profile and fitted a curve to generate the 2D parametric breast model. This involved the following image processing operations on all mammograms within the selected dataset: conversion to grayscale, contrast enhancement, binarization, and closure. First, we converted all CC mammograms to grayscale, as all color channels had identical intensity values, resulting in no information loss. We then applied histogram equalization to enhance contrast, improving the differentiation between foreground and background. Following this, we used a closure operation to completely separate the breast (assigned a value of 1) from the background (assigned a value of 0). Due to varying sizes of CC mammograms in the selected dataset, we standardized them to 512×512 pixels by scaling and padding with background pixels on the left. We then extracted the breast profiles using Canny's edge detector. After extracting the compressed breast profiles in CC projection, we selected the curve that provided the best approximation with the fewest parameters. We used ellipse fitting with a base of five parameters: a, b, ϕ, X_0, Y_0 (see Table 1). The parametric form of the ellipse is given by

$$\begin{cases} x(t) = X_0 + a\cos(t)\cos(-\phi) - b\sin(t)\sin(-\phi), \\ y(t) = Y_0 + a\cos(t)\sin(-\phi) + b\sin(t)\cos(-\phi) \end{cases} \tag{1}$$

In addition, we derived two additional parameters defining the start and end points of the breast profile, which correspond to the range of the parameter t (see Table 1). In total, we extracted seven shape parameters from each of the 163 CC mammograms in the selected dataset.

Table 1: Ellipse Parameters

Parameter	Description	Minimum Value	Maximum Value
a	Horizontal semi-axis	0.27	1.94
b	Vertical semi-axis	0.31	0.72
ϕ	Tilt angle	$-41°$	$44°$
X_0	Center's abscissa	Fixed	
Y_0	Center's ordinate	Fixed	
α_e	End angle	$45°$	$148°$
α_s	Start angle	Merged with α_e	
θ	Nipple positioning angle	$113°$	$233°$

After defining the model shape, we identified key internal features of the breast based on their visual appearance in mammograms. In normal CC views, these include the nipple, fibroglandular tissue, and the rest of the breast (mostly composed of fat tissue with blood vessels). We also included additional findings in the model, such as skin thickening, malignant solid lesions, and benign solid lesions.

3.3 GUI Implementation

In the next phase, we developed a user-friendly GUI to visually represent the parametric 2D model of the compressed breast in the CC view; see Fig. 3. This GUI allows users to adjust breast shape parameters, add fibroglandular tissue, skin thickening, and malignant or benign masses. Designed to be user-friendly, the GUI is suitable for both technical users, like engineers and computer scientists, who need to create and balance synthetic mammogram datasets, and medical professionals, including students, residents, and professors, who use it for education and radiology training.

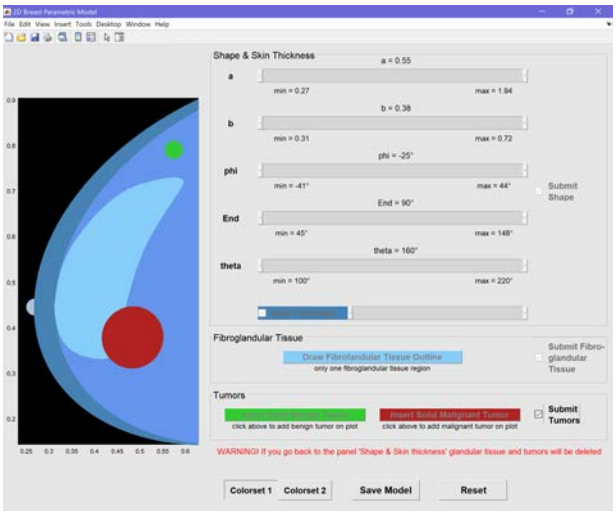


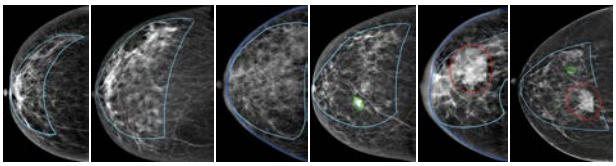
Figure 3: Example of our GUI using color set 1 (see Sec. 4.1). The user follows three panels: (i) selects breast shape parameters and skin thickening, (ii) inserts the fibroglandular tissue region, and (iii) adds malignant or benign tumors. The model can then be saved or reset to default.

To simplify the creation of the 2D parametric model, we minimized the number of breast shape parameters in the GUI. This approach prevents clutter and maintains user engagement by reducing complexity. We combined the start and end points of the breast profile into a single parameter, aligning them vertically, so the model starts at the lower-right and ends at the upper-right corner of the image. We also fixed parameters X_0, Y_0 allowing us to reduce the user-modifiable parameters to four: a, b, ϕ, α_e (see Table 1). Additionally, we introduced a fifth parameter, θ, to account for the nipple's presence and location (see Table 1). For each of these five parameters, the GUI includes a slider that allows users to adjust their values. A checkbox and slider were also added for controlling skin thickening. We enhanced the GUI with a freehand drawing tool for defining the fibroglandular tissue region, which varies greatly in CC mammograms. Additionally, we added buttons to insert and adjust ellipses for benign (green) and malignant (red) tumors. Users can move, resize, or modify the fibroglandular tissue region and the tumors; see Fig. 3. For more details on how to use the GUI, please refer to the user guide 'quick_start' available at the GitHub link[??]. The guide has been provided to assist the user step-by-step in creating the labels with the 2D compressed breast model implemented in our GUI. However, to ensure the push buttons and parameters are intuitive, we added pop-up messages and descriptive text below the buttons for guiding the input of fibroglandular tissue and masses, even without the user guide. Using our GUI (Fig. 3), users can create 2D models of compressed breasts in CC view, adjusting the shape, size, and features like fibroglandular tissue, nipple, fat tissue, skin thickening, and malignant and benign solid lesions. After training pix2pix and real-ESRGAN, these labels will be used to generate synthetic mammograms with corresponding features (see Fig. 2).

3.4 Image-to-Image Translation Tool Training

In this work block, we trained an I2I tool, designed to transform one representation of a scene into another by leveraging sufficient training data, in order to generate synthetic mammograms from the labels created using our GUI. This approach ensures that the generated synthetic images accurately replicate the features defined by the labels.

We opted to utilize a supervised learning tool due to the availability of a dataset comprising real mammograms, for which we generated corresponding labels using our GUI. A qualitative and quantitative



(a) (b) (c) (d) (e) (f)

Figure 4: Examples of segmentations on images from the selected dataset, assisted by a breast radiologist to identify key mammogram features. (a) and (b) show normal breasts with fibroglandular tissue in light blue. (c) highlights skin thickening in blue. (d) presents a benign solid tumor in green. (e) displays both skin thickening in blue and a malignant tumor in red. (f) shows two tumors, one benign in green and one malignant in red.

evaluation of various methods showed that pix2pix [9] is among the most effective I2I tools [47]. Since it is also the most essential and widely used method in medical image generation studies, we chose it for the mammogram generation phase. Pix2pix is a conditional GAN (cGAN), meaning it generates images based on a given input, which in our case are the labels of the mammograms. These networks learn a mapping from observed image x and random noise vector z, to y, $G : f(x, z) \mapsto y$. The generator G, with a U-Net based architecture in pix2pix, is trained to create outputs that are indistinguishable from real images, as determined by an adversarially trained discriminator. The discriminator D, a convolutional PatchGAN classifier in pix2pix, in turn, is trained to be as effective as possible at identifying the generator's fake images. Unlike previous approaches, pix2pix uses the following objective:

$$G^* = \arg\min_G \max_D \left(L_{\text{cGAN}}(G, D) + \lambda L_{L1}(G)\right) \quad (2)$$

with

$$\begin{aligned} L_{\text{cGAN}}(G, D) &= \mathbb{E}_{x,y}\left[\log D(x, y)\right] \\ &\quad + \mathbb{E}_{x,z}\left[\log\left(1 - D(x, G(x, z))\right)\right] \quad (3) \\ L_{L1}(G) &= \mathbb{E}_{x,y,z}\left[\|y - G(x, z)\|_1\right] \end{aligned}$$

To train the pix2pix network we prepared two datasets: one of the breast models (labels) as input and one of the real mammograms as output, both at a resolution of 512×512. Using our GUI, we created a label for each of the 163 CC mammograms in the selected dataset. A medical expert's guidance was crucial in accurately delineating regions of interest, ensuring that the models faithfully represented the characteristics of the corresponding mammograms, and creating high-quality training pairs for pix2pix (Fig. 4). We divided the 163 input-output pairs (labels-real mammograms) into a training set (124 image pairs) and a test set (39 image pairs). We formed the two groups so that the various cases identified in the mammograms were evenly distributed. For the training phase, we kept all the main hyperparameters at their default values (learning rate of 0.0001, batch size of 1, and 1000 epochs). The only adjustments made were to set the augmentation to 8, referring to the number of times the dataset is augmented, and to set the encoding to 1, specifying the encoding method used for the label. The training on a PC running Ubuntu 22.04 with an NVIDIA GTX 3080 video card with 10GB RAM took 17.5 hours, and we used the Adam optimizer [48] with the following parameters: $l_r = 1{\cdot}10^{-4}$ (learning rate), $\beta_1 = 0.5$, and $\beta_2 = 0.999$. After training, the network inference time on the same computer was 3 ms.

3.5 Super-Resolution

After training pix2pix, we obtained a network that generates a synthetic mammogram of size 256×256 pixels from an input label. Since classic pix2pix generates high-quality but low-resolution synthetic mammograms, we upscaled these 256×256 images to 1024×1024 using real-ESRGAN [44], a popular image upscaling network, after applying other upsampling methods as well (more information in Section 4.1). Real-ESRGAN extends the classic ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) [49] to restore real-world low-resolution images by modeling practical degradations through a high-order degradation process, which is more flexible compared to previous methods and attempts to mimic the real degradation generation process. Addi-

tionally, real-ESRGAN incorporates sinc filters during synthesis to simulate common artifacts like ringing and overshoot. The sinc filter kernel can be expressed as:

$$\mathbf{k}(i,j) = \frac{\omega_c}{2\pi\sqrt{i^2+j^2}} J_1\left(\omega_c\sqrt{i^2+j^2}\right), \qquad (4)$$

where (i, j) are the kernel coordinates, ω_c is the cutoff frequency, and J_1 is the first order Bessel function of the first kind. Real-ESRGAN also extends the architecture of the classic ESRGAN, employing a U-Net discriminator with spectral normalization to address a much larger degradation space than ESRGAN. Eventually, trained with synthetic data, real-ESRGAN effectively restores a wide range of real-world images, outperforming previous methods in visual quality and making it highly practical for real-world applications. For our work, we fine-tuned real-ESRGAN with our training dataset using the original resolution images. The fine-tuning process involved patches of size 64×64 for low-resolution inputs and 256×256 for high-resolution targets for 128 epochs. The fine-tuning training time on the same system setup used for the pix2pix training was 10 hours while the inference time is 70 ms. We also conducted experiments with different input patch sizes (i.e., 32×32 and 128×128). A qualitative visual inspection did not reveal noticeable differences, and a quantitative analysis performed using a one-way ANOVA test with a 5% significance level confirmed that there were no statistically significant differences in PSNR, SSIM and DreamSim metrics as the patch size varied (further details are provided in the supplementary material).

Initial inspection showed that images upsampled using real-ESRGAN, post-processed to improve the appearance of breast contours and nipples (enhanced), obtained more than satisfactory results, achieving images at 1024×1024 resolution than other methods; see Fig. 5 and Section 4.1 for more details. Even more quantitative results demonstrate that combining pix2pix and real-ESRGAN with post-processing (enhanced) generates more realistic images than other approaches; see Table 2 and Fig. 6.

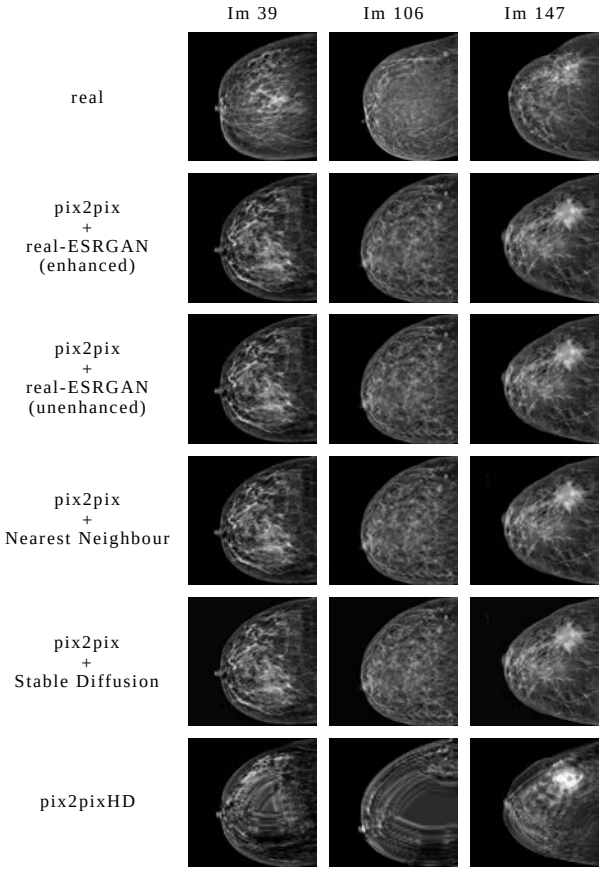


Figure 5: Comparison of real mammograms and synthetic ones generated using pix2pix + real-ESRGAN with post-processing (enhanced), pix2pix + real-ESRGAN without post-processing (unenhanced), pix2pix + nearest neighbour interpolation, pix2pix + Stable Diffusion and pix2pixHD. Images include a normal mammogram (Im 39), one with a benign tumor (Im 106), and one with a malignant tumor (Im 147), all at 1024×1024 resolution.

4 Results and Discussion

4.1 Ablation Study

We performed an ablation study to qualitatively assess the impact of all the choices made during the training phase. Initially, we trained pix2pix for 100

epochs without data augmentation, using a color set where each color had components in all RGB channels (see Fig. 7). The results show little resolution of all structures and artifacts, i.e., a regular grid pattern inside the breast. We then increased the training to 1000 epochs and incorporated data augmentation, resulting in improved resolution of the breast structures and elimination of artifacts.

Despite the improved quality of the results, some structures, like benign masses, were inadequately represented or missing. To address this, we revised the color coding for better differentiation of breast features. We assigned benign lesions to the green channel, malignant lesions to the red channel, and anatomical structures to the blue channel, ensuring an intensity difference of at least 50 between structures (color set 2; see Fig. 7). To determine which method generates the most realistic synthetic mammograms, we calculated PSNR and SSIM indices by comparing real images, scaled to a resolution of 1024×1024, with synthetic images generated using the following methods: pix2pixHD, pix2pix combined with Stable Diffusion [50], pix2pix combined with nearest neighbor interpolation, pix2pix combined with real-ESRGAN without post-processing (unenhanced), and pix2pix combined with real-ESRGAN with post-processing (enhanced) to improve the appearance of breast contours and nipples (see Tab. 2).

Table 2: PSNR, SSIM and DreamSim score (mean ± standard deviation) of pix2pixHD, pix2pix+Stable Diffusion, pix2pix+nearest neighbor, pix2pix+real-ESRGAN (unenhanced), pix2pix+real-ESRGAN (enhanced)

	PSNR ↑	**SSIM ↑**	**DreamSim ↓**
pix2pixHD	17.2201± 1.7234	0.6076± 0.0993	0.2219± 0.0887
pix2pix + Stable Diffusion	17.0992± 1.7324	0.2110± 0.0393	0.1361± 0.0386
pix2pix + Nearest Neighbor	17.2990± 1.8063	0.6220± 0.0938	0.1401± 0.0417
pix2pix + real-ESRGAN (unenhanced)	17.2442± 1.7811	0.6197± 0.0936	0.1348± 0.0391
pix2pix + real-ESRGAN (enhanced)	17.2687± 1.7886	0.6188± 0.0936	0.1327± 0.0399

However, PSNR and SSIM, being low-level metrics, focus on pixel-wise details and fail to capture higher-level structures. To better evaluate the overall structure of the images, rather than just pixel-to-pixel similarity, we employed DreamSim [51], a novel perceptual similarity metric that aligns more closely with human perception, to assess the realism of the synthetic images. DreamSim measures the perceptual distance between real mammograms and synthetic ones, where a higher score indicates greater dissimilarity and a lower score indicates higher similarity.

From the metrics, we observed that the pix2pix+nearest neighbor method achieved the best scores in both PSNR and SSIM, while the pix2pix+real-ESRGAN (enhanced) method achieved the best performance in DreamSim. However, an ANOVA test revealed that, for the PSNR metric, there were no statistically significant differences between the results obtained by any of the methods. For the SSIM metric, a statistically significant difference was found only between the worst-performing method, pix2pix+Stable Diffusion, and all other methods. Similarly, for the DreamSim metric, a statistically significant difference was observed only between the worst-performing method, pix2pixHD, and all other methods.

Among the methods with the best and second-best values, pix2pix+nearest neighbor, pix2pix+real-ESRGAN (unenhanced), and pix2pix+real-ESRGAN (enhanced), no statistically significant differences were observed. Given the absence of statistically significant differences among these methods, we chose to proceed with the method that achieved the best performance in the DreamSim metric: pix2pix+real-ESRGAN (enhanced). This decision is based on the fact that DreamSim is specifically designed to simulate human perception of images, making it a more reliable metric for assessing the realism of synthetic mammograms. While PSNR and SSIM focus on pixel-wise accuracy, DreamSim captures higher-level perceptual qualities that are more aligned with the visual evaluation of synthetic images. As realism is a critical factor in the context of this study, DreamSim provides a more meaningful measure for selecting the most appropriate method.

This decision is further supported by visual analy-

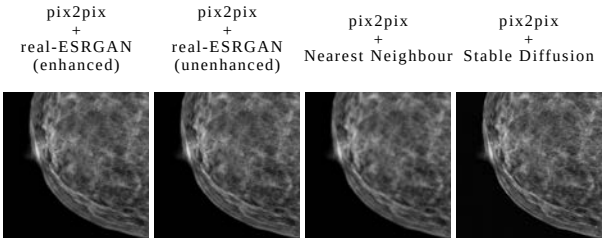


Figure 6: Comparison of synthetic mammograms generated by different methods: pix2pix+Stable Diffusion, pix2pix+nearest neighbor, pix2pix+real-ESRGAN (unenhanced), and pix2pix+real-ESRGAN (enhanced). Pix2pix+Stable Diffusion shows unrealistic needle-like patterns, while pix2pix+nearest neighbor appears pixelated. Pix2pix+real-ESRGAN (unenhanced) exhibits pixelation in the nipple area. Pix2pix+real-ESRGAN (enhanced) delivers the most realistic results.

sis of the generated images, which highlights qualitative differences between the methods. In particular, as shown in Fig. 6, the pix2pix+Stable Diffusion method exhibits an unrealistic needle-like pattern within the breast region. Although the pix2pix+nearest neighbor method generates images at the same resolution as the other methods, it appears to be of lower resolution due to the visibly pixelated structure. The pix2pix+real-ESRGAN (unenhanced) method shows evident pixelation in the nipple area, whereas the pix2pix+real-ESRGAN (enhanced) method produces the most realistic synthetic images.

4.2 Quality Assessment

To assess the quality of the generated images, we rely on three methods: qualitative visual analysis, statistical analysis by plotting the first five statistical moments, and evaluation of perceived realism by domain experts through a questionnaire.

We first assessed qualitatively the synthetic mammograms generated using pix2pix + real-ESRGAN with post-processing (enhanced), through a visual examination performed by us based on our personal perception; see Fig. 7. We observed that in the generated images, the nipple was accurately reproduced in the correct position. Additionally, in our opinion, the shape and position of the fibroglandular tissue region were generated correctly, and its texture appeared very realistic. The results related to skin thickening were also of good quality, according to our perception: the shape and area of the thickened area were generated correctly, as well as the uniform texture of this region. Regarding sharp-edged opacities, i.e., benign solid tumors, better results were obtained in cases where the finding was located outside the glandular tissue. Finally, the results related to malignant solid lesions also appeared good to us: we observed that the network learned the main characteristic of these lesions, namely the spiculated margins. The network may generate realistic images in which multiple overlapping tumors are rendered with lower fidelity than single tumors. This is because they are underrepresented in the training set. We attempted to address this using a random oversampling strategy targeting these cases, but this approach did not lead to a significant improvement (details in supplementary material). A meaningful enhancement in generation quality for such complex patterns would likely require a larger number of representative training examples.

As a second quality assessment, we plotted the first five statistical moments (mean, variance, skewness, kurtosis, and hyperskewness) to evaluate the similarity between the low-level pixel distributions of real and synthetic images, as done by Korkinof *et al.* [33] and Sizikova *et al.* [27]; see Fig 8. We observed that the image statistics overall correspond reasonably well. Additional evaluation metrics can be taken into account; however, they are primarily useful for comparing different synthesis methods (see Table 2) and not for assessing human perception and the perceived realism by domain experts. To support this, consider how little the values in Table 2 differ and how different the images in Fig. 5 are.

As a third quality assessment, we conducted a more robust evaluation of the visual realism of the generated images by performing a user study with two groups of experts: biomedical engineers with master's degrees and medical professionals (radiology residents and radiologists) with varying levels of exper-

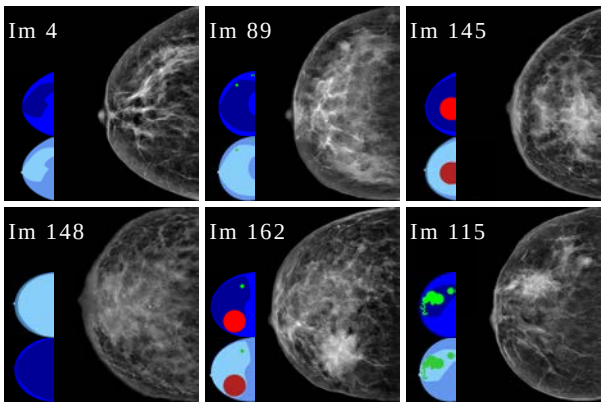


Figure 7: Labels with two color sets and their corresponding mammograms generated using pix2pix + real-ESRGAN with post-processing (enhanced). The top labels use color set 2, and the bottom labels use color set 1. Im 4 is normal; Im 89 has two benign tumors; Im 145 shows a malignant tumor with skin thickening; Im 148 has skin thickening only; Im 162 includes both benign and malignant tumors; and Im 155 shows multiple overlapping benign tumors.

tise. Although the engineers are not medical specialists, we decided to involve them in the assessment, because they are familiar with both real and synthetic biomedical images.

We administered a questionnaire to both groups with 10 pairs of images: one real mammogram and its corresponding synthetic mammogram. For each image pair, users were asked to identify the real mammogram among the two options. We observed that the 9 biomedical engineers had difficulty recognizing real mammograms among the pairs proposed in the questionnaire. The average number of real mammograms identified was 2.89 out of 10 with a standard deviation of 1.88; the mode value was 2. From these initial results, we understood that the synthetic images achieved a high level of realism. Therefore, we could also administer the questionnaire to a group of medical experts. This group consisted of 42 radiology residents and 3 radiologists. As in the previous case, we provided 10 pairs of images (one real and one synthetic) and asked the participants to choose the real image from each pair (all images employed in this user study are available in the supplemental material). The average number of real mammograms identified was 3.04 out of 10 with a standard deviation of 2.07; the mode value was 2. Considering a significance level of 0.1%, the difference between the two groups in votes to have statistical differences was set at 23 votes [52]. From this, we can conclude that there is a statistical difference between the two groups (those who voted for the real image and those who voted for our synthetic image) only in the case of the first, fourth, and eighth pairs of images; see Table 3. In all other cases, the real and the generated images are statistically indistinguishable.

Table 3: The results of our user study. **GT** is a real mammogram, and **our** is a synthetic mammograms generated by our method. Any image within the same circle cannot be considered perceptually different.

	ζ	X^2 (sign. 0.5%)	**Ranking** **1st**	**2nd**
Pair 01	0.293	13.889 ✓	**OUR(35)**	**GT(10)**
Pair 02	0.018	1.800 ×	**OUR(27)**	**GT(18)**
Pair 03	0.200	9.800 ✓	**OUR(33)**	**GT(12)**
Pair 04	0.345	16.200 ✓	**OUR(36)**	**GT(9)**
Pair 05	0.123	6.422 ✓	**OUR(31)**	**GT(14)**
Pair 06	0.160	8.022 ✓	**OUR(32)**	**GT(13)**
Pair 07	0.160	8.022 ✓	**OUR(32)**	**GT(13)**
Pair 08	0.402	18.689 ✓	**OUR(37)**	**GT(8)**
Pair 09	0.091	5.000 ✓	**OUR(30)**	**GT(15)**
Pair 10	-0.010	0.556 ×	**OUR(20)**	**GT(25)**

In addition, we analyzed the agreement among voters by calculating the coefficient of agreement, ζ, [52] and determining if the agreement ($\zeta = 1$) or disagreement ($\zeta = -1$) among voters was statistically significant using the chi-square, X^2, test with a significance level of 0.5%. From Table 3, we can observe that there is agreement among voters in most cases, and this agreement is statistically significant. In the case of the second and tenth pairs of images, there is low agreement and low disagreement, respectively, which, however, are not statistically significant. To complement this analysis, we also computed Fleiss' kappa [53] over the entire set of responses (10 im-

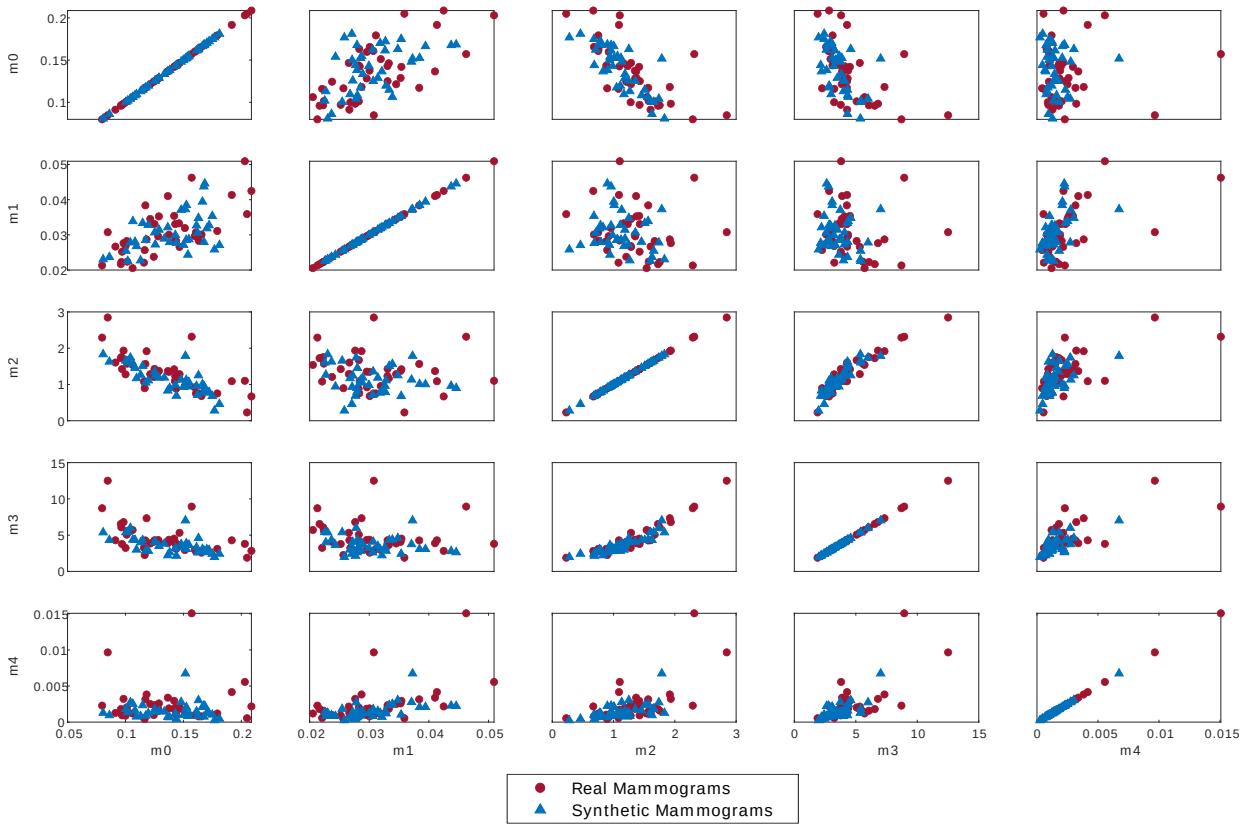


Figure 8: First five statistical moments for real and synthetic mammograms of the test set (39 images). m0: mean, m1: variance, m2: skewness, m3: kurtosis, m4: hyperskewness.

age pairs evaluated by 45 participants), obtaining a value of $\kappa = 0.03$, which falls in the "slight agreement" range according to Landis and Koch's interpretation scale [54]. Such low inter-rater agreement further supports the notion that the synthetic images are highly realistic and often indistinguishable from real ones, making consistent classification by human observers particularly challenging.

5 Conclusions

We have proposed a workflow that includes a statistical generative model based on GANs for the generation of high-resolution synthetic mammograms (1024×1024) with precise feature control and the ability to generate both normal and tumor cases. We achieved this result thanks to **two fundamental steps**.

1) The definition of a unique 2D parametric model of the compressed breast in the CC view that we have implemented in our GUI

2) The adaptation of the pix2pix [9] and the super-resolution [44] frameworks for our method to generate high-quality images.

Starting with a publicly available mammogram dataset [45, 46] we selected images that suited our research, creating our selected dataset. Using these mammograms, we fitted breast profiles with ellipses and implemented the 2D parametric model of the

compressed breast in CC view in a GUI. We then trained the pix2pix I2I and real-ESRGAN super-resolution networks, enabling the generation of high-quality synthetic mammograms with a 1024×1024 resolution and fully controllable features through our interface. The limited size of the dataset (163 images) may affect the model's generalization. Its expansion, especially with more cases of overlapping tumors, will be crucial to improve both performance and representation of complex findings. We believe the only feasible way to achieve this is through collaboration with a hospital. Anonymization of screening images must be carried out by hospital staff to ensure patient privacy, followed by expert supervision from a breast radiologist for accurate segmentation. Such a partnership is key to building a larger, reliable dataset for enhancing synthetic mammogram generation. Another improvement involves incorporating MLO views, as multi-view synthesis would further enhance realism and clinical relevance.

Quality assessment was conducted through visual analysis, statistical analysis using the first five statistical moments, and a questionnaire administered to 45 medical experts. The results showed that the features of the real mammograms were accurately replicated in the synthetic ones, the image statistics overall correspond reasonably well, and the two groups of images were statistically indistinguishable in almost all cases according to the experts. Our approach allow to generate new synthetic datasets tailored to researchers' needs and balances existing datasets by providing innovative control over image features. These datasets can be used both for educational purposes to train future radiologists and for research purposes to train new ML and DL algorithms. These algorithms can then assist radiologists in tasks like classification and segmentation, improving diagnostic performance. To explore this potential, we conducted a preliminary assessment using synthetic mammograms to train a U-Net segmentation network. The goal was to evaluate the feasibility of employing synthetic data in downstream tasks. Given the early and exploratory nature of this evaluation, it was not included in the main manuscript but is reported in the Supplementary Material (Section VII). These initial results support the idea that synthetic images can be used effectively to train deep learning models, laying the groundwork for more extensive investigations in future work.

Acknowledgment

The authors thank Dr. Damiana Francesca (Versilia Hospital) for her support in selecting and segmenting mammograms. They also acknowledge Dr. Lorenzo Faggioni (radiologist at Pisa University Hospital) and Dr. Salvatore Claudio Fanni (Ph.D. student at the University of Pisa) for distributing the questionnaire among radiologists and trainees. A.F. Bonatti and C. De Maria acknowledge the support of the CrossLab "Cloud Computing, Big Data & Cybersecurity" and FoReLab, Department of Information Engineering, University of Pisa.

References

[1] V. Terrasse, "Latest global cancer data: Cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020.."

[2] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, Feb., 2021. DOI: 10.3322/caac.21660.

[3] R. T. Greenlee *et al.*, "Cancer statistics, 2001," *CA Cancer J. Clin.*, vol. 51, no. 1, pp. 15–36, Jan./Feb., 2001. DOI: 10.3322/canjclin.51.1.15.

[4] J. R. Harris *et al.*, "Breast cancer," *N. Engl. J. Med.*, vol. 327, no. 7, pp. 473–480, 1992. DOI: 10.1056/NEJM199208133270706.

[5] A. Kolak *et al.*, "Primary and secondary prevention of breast cancer," *Ann. Agric. Environ. Med.*, vol. 24, no. 4, pp. 549–553, dec., 2017. DOI: 10.26444/aaem/75943.

[6] A. L. Siu and U. P. S. T. Force, "Screening for breast cancer: US preventive services task force recommendation statement," *Ann. Intern. Med.*, vol. 164, no. 4, pp. 279–296, Jan., 2016.

[7] T. Li *et al.*, "A systematic review of the impact of the COVID-19 pandemic on breast cancer screening and diagnosis," *Breast*, vol. 67, pp. 78–88, Jan., 2023. DOI: 10.1016/j.breast.2023.01.001.

[8] E. Sizikova *et al.*, "Synthetic data in radiological imaging: current state and future outlook," *BJR Artif. Intell.*, vol. 1, no. 1, p. ubae007, May, 2024. DOI: 10.1093/bjrai/ubae007.

[9] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE CVPR*, pp. 1125–1134, 2017.

[10] W. Ren *et al.*, "Global guidelines for breast cancer screening: A systematic review," *Breast*, vol. 64, pp. 85–99, Aug., 2022. DOI: 10.1016/j.breast.2022.04.003.

[11] D. De Benedetto *et al.*, "Radiological anatomy of the breast," *Ital. J. Anat Embryol.*, vol. 121, no. 1, pp. 20–36, 2016.

[12] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. NIPS*, vol. 27, 2014.

[13] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1348–1357, April, 2018. DOI: 10.1109/TMI.2018.2827462.

[14] T. M. Quan *et al.*, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1488–1497, March, 2018. DOI: 10.1109/TMI.2018.2820120.

[15] A. Madani *et al.*, "Chest x-ray generation and data augmentation for cardiovascular abnormality classification," in *Proc. SPIE Med. Imaging*, vol. 10574, pp. 415–420, 2018.

[16] C. Han *et al.*, "Gan-based synthetic brain MR image generation," in *Proc. IEEE 15th ISBI*, pp. 734–738, 2018.

[17] C. Baur *et al.*, "Generating highly realistic images of skin lesions with GANs," in *Proc. CARE CLIP OR 2.0 ISIC*, vol. 11041, pp. 260–267, 2018.

[18] C.-B. Jin *et al.*, "Deep CT to MR synthesis using paired and unpaired data," *Sens.*, vol. 19, no. 10, p. 2361, May, 2019. DOI: 10.3390/s19102361.

[19] G. Jiang *et al.*, "Synthesis of mammogram from digital breast tomosynthesis using deep convolutional neural network with gradient guided cGANs," *IEEE Trans. Med. Imaging*, vol. 40, no. 8, pp. 2080–2091, April, 2021. DOI: 10.1109/TMI.2021.3071544.

[20] B. Barufaldi *et al.*, "OpenVCT: a GPU-accelerated virtual clinical trial pipeline for mammography and digital breast tomosynthesis," in *Proc. Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573, p. 1057358, 2018.

[21] H. Tomic *et al.*, "Development and evaluation of a method for tumor growth simulation in virtual clinical trials of breast cancer screening," *J. of Med. Imaging*, vol. 9, no. 3, p. 033503, June, 2022. DOI: 10.1117/1.JMI.9.3.033503.

[22] B. Barufaldi *et al.*, "Computational breast anatomy simulation using multi-scale Perlin noise," *IEEE Trans. Med. Imaging*, vol. 40, no. 12, pp. 3436–3445, June, 2021. DOI: 10.1109/TMI.2021.3087958.

[23] C. G. Graff, "A new, open-source, multi-modality digital breast phantom," in *Proc. Medical Imaging 2016: Physics of Medical Imaging*, pp. 72–81, 2016.

[24] A. Sarno *et al.*, "Dataset of patient- derived digital breast phantoms for in silico studies in breast computed tomography, digital breast tomosynthesis, and digital mammography," *Med. Phys.*, vol. 48, pp. 2682–2693, May, 2021. DOI: 10.1002/mp.14826.

[25] L. de Sisternes *et al.*, "A computational model to generate simulated three-dimensional breast

masses," *Med. Phys.*, vol. 42, no. 2, pp. 1098–1118, Feb., 2015. DOI: 10.1118/1.4905232.

[26] A. Sengupta *et al.*, "Computational model of tumor growth for in silico trials," in *Proc. Medical Imaging 2021: Physics of Medical Imaging*, p. 115954S, 2021.

[27] E. Sizikova *et al.*, "Knowledge-based in silico models and dataset for the comparative evaluation of mammography ai for a range of breast characteristics, lesion conspicuities and doses," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, pp. 37401–37412, 2023.

[28] A. Badano *et al.*, "Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial," *JAMA Netw. Open*, vol. 1, no. 7, pp. e185474–e185474, Nov., 2018. DOI: 10.1001/jamanetworkopen.2018.5474.

[29] Z. Li *et al.*, "A novel 3D stochastic solid breast texture model for X-ray breast imaging," in *Proc. 13th IWDM*, vol. 9699, pp. 660–667, 2016.

[30] Z. Li *et al.*, "A 3D mathematical breast texture model with parameters automatically inferred from clinical breast CT images," *IEEE Trans. Med. Imaging*, vol. 42, no. 4, pp. 1107–1120, Nov., 2022. DOI: 10.1109/TMI.2022.3224223.

[31] A. Badal *et al.*, "Mammography and breast tomosynthesis simulator for virtual clinical trials," *Comput. Phys. Commun.*, vol. 261, p. 107779, 2021. DOI: 10.1016/j.cpc.2020.107779.

[32] D. Korkinof *et al.*, "High-resolution mammogram synthesis using progressive generative adversarial networks," *arXiv:1807.03401*, 2018.

[33] D. Korkinof *et al.*, "MammoGAN: High-resolution synthesis of realistic mammograms," 2019.

[34] T. Shen *et al.*, "Mass image synthesis in mammogram with contextual information based on GANs," *Comput. Methods Programs Biomed.*, vol. 202, p. 106019, April, 2021. DOI: 10.1016/j.cmpb.2021.106019.

[35] A. Yamazaki and T. Ishida, "Two-view mammogram synthesis from single-view data using generative adversarial networks," *Appl. Sci.*, vol. 12, no. 23, p. 12206, Nov., 2022. DOI: 10.3390/app122312206.

[36] J. Lee and R. M. Nishikawa, "Identifying women with mammographically-occult breast cancer leveraging GAN-simulated mammograms," *IEEE Trans. Med. Imaging*, vol. 41, no. 1, pp. 225–236, Aug., 2021. DOI: 10.1109/TMI.2021.3108949.

[37] R. P. Tripathi *et al.*, "Enhancing breast cancer diagnosis through segmentation-driven generative adversarial networks for synthetic mammogram generation," in *Proc. ICTACS*, pp. 1078–1082, 2023.

[38] D. Shah *et al.*, "Enhancing the quality and authenticity of synthetic mammogram images for improved breast cancer detection," *IEEE Access*, vol. 12, pp. 12189–12198, Jan., 2024. DOI: 10.1109/ACCESS.2024.3354826.

[39] N. Wu *et al.*, "Deep neural networks improve radiologists' performance in breast cancer screening," *IEEE Trans. Med. Imaging*, vol. 39, no. 4, pp. 1184–1194, Oct., 2019. DOI: 10.1109/TMI.2019.2945514.

[40] H. Li *et al.*, "Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography," *IEEE Trans. Med. Imaging*, vol. 41, no. 1, pp. 3–13, Aug., 2021. DOI: 10.1109/TMI.2021.3102622.

[41] C. D. Lehman *et al.*, "Diagnostic accuracy of digital screening mammography with and without computer-aided detection," *JAMA Intern. Med.*, vol. 175, no. 11, pp. 1828–1837, Nov., 2015. DOI: 10.1001/jamainternmed.2015.5231.

[42] D. T. Chen *et al.*, "Modeling for plastic and reconstructive breast surgery," in *Proc. MICCAI*, vol. 1935, pp. 1040–1050, 2000.

[43] G. Gallo *et al.*, "Parametric representation of human breast shapes," in *Proc. IEEE MEMEA*, pp. 26–30, 2009.

[44] X. Wang *et al.*, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF ICCV*, pp. 1905–1914, 2021.

[45] R. Khaled *et al.*, "A categorized digital database for low energy and subtracted contrast enhanced spectral mammography images," 2021.

[46] K. Clark *et al.*, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imaging*, vol. 26, pp. 1045–1057, July, 2013. DOI: 10.1007/s10278-013-9622-7.

[47] Y. Pang *et al.*, "Image-to-image translation: Methods and applications," *IEEE Trans. Multimedia*, vol. 24, pp. 3859–3881, Sept., 2021. DOI: 10.1109/TMM.2021.3109419.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.

[49] X. Wang *et al.*, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proc. ECCV workshops*, pp. 0–0, 2018.

[50] R. Rombach *et al.*, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF CVPR*, pp. 10684–10695, June 2022.

[51] S. Fu *et al.*, "Dreamsim: Learning new dimensions of human visual similarity using synthetic data," *arXiv:2306.09344*, 2023.

[52] H. A. David, *The Method of Paired Comparisons, 2nd ed.* Oxford University Press, 1988.

[53] J. L. Fleiss, "Measuring nominal scale agreement among many raters.," *Psychol. Bull.*, vol. 76, no. 5, p. 378, 1971.

[54] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.