

Article

Revealing Occult Malignancies in Mammograms Through GAN-Driven Breast Density Transformation

Dionysios Anyfantis ¹, Athanasios Koutras ^{1,*}, George Apostolopoulos ² and Ioanna Christoyianni ²

¹ SIPPRE Group, Electrical and Computer Engineering, University of Peloponnese, 221 00 Patras, Greece; d.anyfantis@go.uop.gr

² Electrical and Computer Engineering, University of Patras, 265 04 Patras, Greece; gapost@upatras.gr (G.A.); ioachrist@upatras.gr (I.C.)

* Correspondence: koutras@uop.gr

Abstract: Breast cancer remains one of the primary causes of cancer-related deaths among women globally. Early detection via mammography is essential for improving prognosis and survival rates. However, mammogram diagnostic accuracy is severely hindered by dense breast tissue, which can obstruct potential malignancies, complicating early detection. To tackle this pressing issue, this study introduces an innovative approach that leverages Generative Adversarial Networks (GANs), specifically CycleGAN and GANHopper, to transform breast density in mammograms. The aim is to diminish the masking effect of dense tissue, thus enhancing the visibility of underlying malignancies. The method uses unsupervised image-to-image translation to gradually alter breast density (from high (ACR-D) to low (ACR-A)) in mammographic images, detecting obscured lesions while preserving original diagnostic features. We applied this approach to multiple mammographic datasets, demonstrating its effectiveness in diverse contexts. Experimental results exhibit substantial improvements in detecting potential malignancies concealed by dense breast tissue. The method significantly improved precision, recall, and F1-score metrics across all datasets, revealing previously obscured malignancies and image quality assessments confirmed the diagnostic relevance of transformed images. The study introduces a novel mammogram analysis method using advanced machine-learning techniques, enhancing diagnostic accuracy in dense breasts and potentially improving early breast cancer detection and patient outcomes.



Citation: Anyfantis, D.; Koutras, A.; Apostolopoulos, G.; Christoyianni, I. Revealing Occult Malignancies in Mammograms Through GAN-Driven Breast Density Transformation. *Electronics* **2024**, *13*, 4826. <https://doi.org/10.3390/electronics13234826>

Keywords: ACR BI-RADS; breast cancer; breast mammographic density; computing methodologies; CycleGAN/Multi-Hop GAN/gGANHopper; deep learning; generative adversarial networks; image transformations; mammography; masking effect

Academic Editors: Irena Galić,
Marija Habijan, Daniel Vasić and
Mirela Kundid Vasić

Received: 1 November 2024

Revised: 1 December 2024

Accepted: 4 December 2024

Published: 6 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial intelligence, particularly deep-learning approaches, has attracted attention due to its excellent performance in medical imaging [1]. These methods span from feature engineering in conjunction with supervised classifiers to learn low-level pixel representations up to the more recent trend of feature learning with convolutional neural networks (CNNs) [2]. The introduction of generative adversarial networks (GANs) [3] is one of the most intriguing breakthroughs in this field. GANs, with their ability to replicate data distributions and synthesize images at previously unheard-of degrees of realism, have paved the way for new techniques to bridge the gap between supervised learning and synthetic image production. The enhancements mentioned above are primarily due to the following characteristics: (1) GANs maximize the probability density over the data generating distribution by utilizing density ratio estimation [4] in an indirect manner of supervision; and (2) GANs can discover the high dimensional latent distribution of data, which has resulted in significant performance gains in the extraction of visual features.

Generative Adversarial Networks (GANs) are increasingly used in the literature to address various medical problems, such as data augmentation and synthesis. GAN-based

models should be able to extract common characteristics from both domains and reveal their underlying relationship for picture transformation between two domains; CycleGAN is a conventional GAN model that uses this idea. The concept of an adversarial loss, which requires the generated images to be, in theory, identical to real images, is crucial to CycleGAN's effectiveness.

In this work, we exploit an unsupervised image-to-image translator based on CycleGAN [5] and GANHopper [6] that is constrained to transform images gradually between two domains related to breast density characterization. Breast density can only be identified by mammography, an essential characteristic influencing mammography sensitivity [7–11]. Due to its impact on the accurate detection of cancer in mammograms, it would be preferable if a method could be proposed to alleviate the masking (the loss of lesion conspicuity due to the superposition of the X-Ray attenuation of typical fibroglandular structures in the breast). The sensitivity of detecting malignancies is reduced in dense breasts, with reports ranging from 80% in non-dense breasts to 30% in dense breasts [11,12]. Women with dense breasts are more likely to undergo additional diagnostic procedures, such as conventional magnetic resonance imaging (MRI), due to these known diagnostic challenges. Although MRI has significantly higher sensitivity than mammography, its limited specificity can result in unnecessary biopsies.

In cases of high MBD (Mammographic Breast Density), a mammogram's ability to detect cancers is about the same as flipping a coin. A study that was published in the American Journal of Roentgenology [13] demonstrated how MBD dramatically lowers the sensitivity of breast cancer diagnosis consisting of a substantial group of women who underwent automated volumetric analysis mammography screening. Mammography's sensitivity to detect cancerous regions declines linearly with increasing age. Mammography's sensitivity drops to less than 50% when volumetric breast density rises above 25%. In addition to increasing the risk of breast cancer, elevated MBD significantly impairs the capacity of mammography, the most popular method of breast screening, to identify cancers. High MBD is the most common cause of interval malignancies in screening programs.

We propose an approach to "transform" breast density to the lower characterized level, thus enhancing diagnostic accuracy. Treatment and early detection could prevent much of the global cancer burden. For example, early detection can allow a tumor to be treated before it acquires critical combinations of genetic alterations. Instead of transforming directly from a highly dense breast image (ACR-D) to an entirely fatty breast image (ACR-A), we gradually change the image using a sequence of CycleGANs leading to the desired target domain of ACR-A. The transformation of breast density regions ACR-D to ACR-A reveals previously undetected findings in conjunction with classical Deep learning methods. This process significantly improves recognition accuracy while reducing the number of undetected ROIs due to their dense breast tissue surrounding. A similar approach was investigated in [14], in which a CycleGAN-based model was used to transform images between ACR-A and ACR-B, C, and D classes. Still, this research is more detailed, and the performance is examined in more well-established datasets. The following are this work's key contributions:

- Creating a novel method for detecting hidden suspicious abnormalities in ROIs that are partially or entirely masked by surrounding tissue, taking into account the local breast density characterized by ACR levels.
- The removal of the masking effect caused by surrounding tissue on the examined ROI by converting the ROI's ACR level to the ACR-A category. This procedure uses a modified GANHopper/CycleGAN topology that cycles through the four ACR levels.
- Improving overall abnormal region detection performance with a CNN network architecture based on a VGG16 [15] network and extended tests on the most well-known datasets in the field.

2. Literature Review

The recent advances in Generative Adversarial Networks (GANs) [16] in computer vision and medical imaging provide enhanced cancer detection and analysis capabilities. Most works published in the literature propose exploiting GAN-based architectures to synthesize full-size images or specific areas (ROIs) to overcome data scarcity. In this section, we summarize the latest GAN-based approaches in breast imaging applications focused on those related to applications in digital mammography.

In a study [17], the authors trained a GAN model to generate synthetic patches with benign and malignant tumors. The synthetically generated patches had apparent artifacts that did not match the used dataset distribution. One possible reason reported was the phenomenon of artifact injection by the GAN model structure, i.e., checkerboard artifacts and black spots, that impact the quality of the final synthesized image and, of course, are not encountered in real images.

A UNet-based GAN model for data augmentation that can realistically synthesize and remove lesions on mammograms was proposed in [18]. The central concept presented is based on the fact that lesions can develop approximately uniformly across breast tissue, and thus, given the context of surrounding tissue, the proposed model can realistically generate and remove. The UNet-based architecture can create images of 256×256 pixels using self-attention and semi-supervised learning components. The authors report that the proposed GAN model significantly improved malignancy classification performance. This work focused on developing a GAN model capable of generating multiclass high-resolution images for digital mammograms covering all categories of breast cancer abnormalities.

The first study that reported the synthesis of high-resolution full-field digital mammograms (FFDM) images in the literature proposed the use of a Progressively Growing GAN (PGGAN) [19]. This model was trained using more than 400,000 FFDMs and was demonstrated to generate mammograms up to a 1280×1024 -pixel resolution. Overall, in this study, the synthetic images were shown to be indistinguishable from original mammograms. However, it was unclear and was not further investigated whether the synthetic FFDMs can be used for relevant clinical applications.

In a recent work [20], the researchers tried to improve mass detection performance in high-density breasts using synthetic high-density full-field digital mammograms (FFDM) as data augmentation during breast mass detection model training. Five cycle-consistent GAN (CycleGAN) models were trained on each dataset and breast view (CC and MLO view).

Another study [21] reported GAN artifacts' role in synthesizing feasible-looking breast images. GANs can suffer from multiple artifacts that may impact the final synthesized image's quality and the performance of detecting malignancies. They explored the types of GAN artifacts in mammogram simulations and their effect on mammographically occult (MO) cancer detection.

An attempt to deploy a synthesis image model to produce CC-view mammograms from MLO views was reported in [22]. The techniques the authors applied reduced the model's training time, whereas the feature matching loss considerably improved the quality of the synthesized images. The proposed method succeeded in synthesizing CC views similar to the real images for some (but not all) cases.

The researchers in [23] generated realistic standard mammographic images using the state-of-the-art generation network StyleGAN2. They developed an unsupervised anomaly detection method to detect breast cancer without collecting or annotating cancer datasets.

The authors of the work [24] used a CycleGAN trained on image patches from recalled lesions and normal breast tissue to create a Lesion Remover (LR). The Lesion Removal (LR) adopted an approach of transferring a lesion's style to normal breast tissue using a model based on CycleGAN. Their hypothesis was to highlight the lesion that was already present by merging an image of a lesion that had been excised with its original image. Utilizing color fusion, the highlighted lesion was overlaid onto the original mammogram. The resulting color-fused image highlights the lesion because the Lesion Remover removed the lesion and left the surrounding tissue intact.

Last, the study [25] focused on synthesizing ROI-based digital mammograms by deploying a GAN model named ROI-mammoGAN. Their approach involves designing a GAN model to represent abnormalities in digital mammograms. Attention was given to architectural distortion, asymmetry, mass, and microcalcification abnormalities so that training distinctively learns the features of each abnormality and generates sufficient images for each category.

3. Methodology

This section will cover the methods and philosophy behind the ROI mammogram transformations across the four ACR BI-RADS categories. According to the American College of Radiology Breast Imaging Reporting and Data System [26] (ACR BI-RADS, Rev. 2015) there are four breast density categories: ACR-A (“essentially all fatty”), ACR-B (“sporadic fibroglandular density regions”), ACR-C (“unevenly dense breasts, which may obscure small masses”), and ACR-D (“excessively dense breasts, which lowers the sensitivity of mammography”). The foundation of our approach is developing a process that uses a patch from inside the breast density area and gradually transforms it into the equivalent ACR-A category patch. The process must also handle the boundary case of converting a patch of any ACR category to itself (especially interested in the ACR-A patches case).

Let’s consider X and Y our source and target image domains (the source and target domains are ACR-A and ACR-D breast density space accordingly). Our goal is to progressively learn a transformation function that, given an image $x \in X$ belonging to the ACR-A domain, outputs another image $y' \in Y$ belonging to the ACR-D domain such that y' is perceived to be the counterpart of the image x in the domain Y . The same must be achieved with the reverse transformation from $y \in Y$ to $x' \in X$ (a unique characteristic of CycleGAN). It is anticipated that any appropriate translation function will be intricate and challenging to learn when there are significant differences between the source and target domains, as in our case. The search area for translation functions might be extensive without any paired images; there are more degrees of freedom to consider with significant image changes, so a more limited and controlled search is preferable. Unpaired training data ensured the model’s applicability to our scenario, where image pairs are unavailable because the same breast region of the same patient could not belong to both the high (ACR-D) and low (ACR-A) breast density domains.

The method is based on the CycleGAN [5] and GANHopper [6] (a CycleGAN-based approach with gradual transformations), which we will be referred to from now on as gGANHopper (a.k.a guarded GANHopper). Since the breast density of the tissue according to the ACR BI-RADS is expressed as a proportion of fibroglandular parenchyma in relation to the fatty breast tissue, we are modeling the traverse from one ACR category to its neighbor ACR category as a hop in a breast-dense mix variation manner. This means that each hop is responsible for transforming the breast density in such a manner that, for example, in the used 6-hop (5-node) architecture (Figure 1), the first hop from the right produces images that are 90% ACR-A and 10% ACR-D like, the second hop outputs images that are 70% ACR-A and 30% ACR-D like, the third hop is constrained to produce only ACR-C/ACR-B patches. The last two hops produce patches in a cascading manner (30% ACR-A, 70% ACR-D, and 10% ACR-A, 90% ACR-D) toward the desired output of 100% ACR-D.

Each hop is considered to produce an image of the defined ACR category and progressively prepares the image toward the target domain by altering the breast density. The intermediate image of each hop is declared to have a specific mixture ratio of the input and target image, as explained. The mix ratio chosen for each hop maintains the cumulative ACR between 0% and 50% for the right part of the middle node and 50% to 100% for the left. Repeated consecutive hops produce hybrid images as byproducts of the translation process; although these images are of no immediate use, they provide a visual validation of the ACR transformations. The reverse direction path works similarly by feeding the network from the left side, with patches of the ACR-D category. By making the network create intermediate images that resemble weighted hybrids of the input/output images

of the two domains, we can guide the transformation process and safeguard the overall process by utilizing the functionality of the middle node. The central point of the architecture is a CycleGAN-based model that transforms patches of ACR-B to ACR-C and vice versa. It is trained separately, in an out-of-bound parallel manner, with the gGANHopper network. At specific timestamps during the training process of the gGANHopper network, the updated middle node CycleGAN is reintroduced in the network, and the training of the gGANHopper resumes with the updated response of the central node.

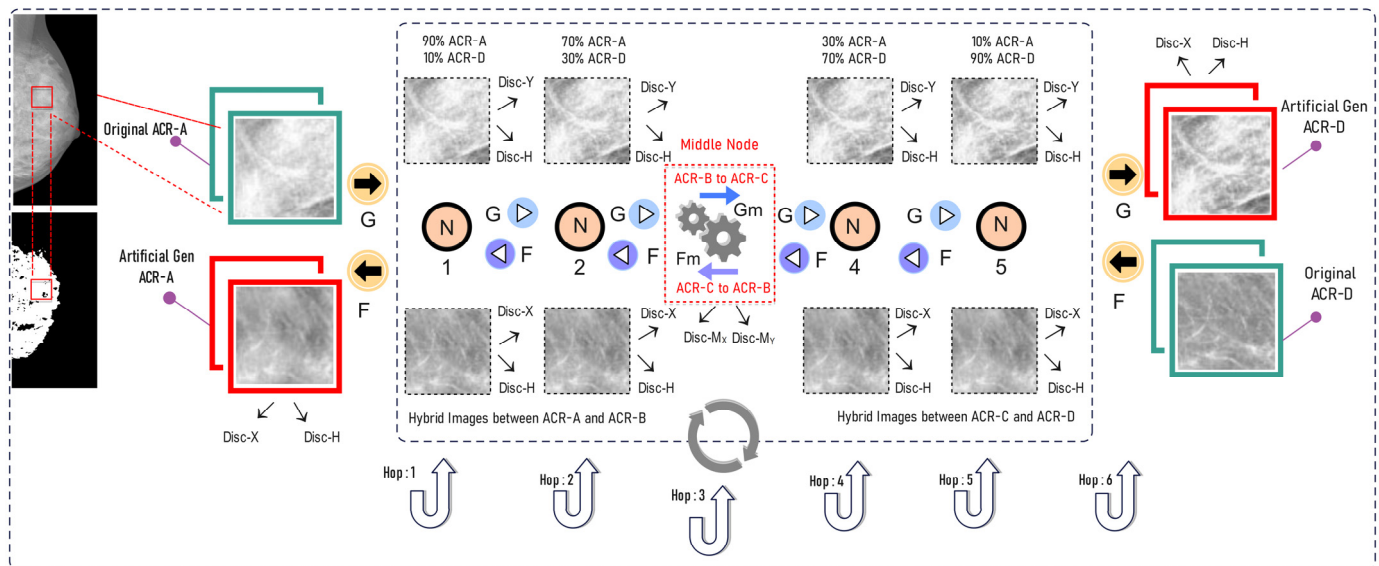


Figure 1. Network architecture overview. Image patch transformation of ACR-A to ACR-D and vice versa.

Each hop is considered to produce an image of the defined ACR category and progressively prepares the image toward the target domain by altering the breast density. The intermediate image of each hop is declared to have a specific mixture ratio of the input and target image, as explained. The mix ratio chosen for each hop maintains the cumulative ACR between 0% and 50% for the right part of the middle node and 50% to 100% for the left. Repeated consecutive hops produce hybrid images as byproducts of the translation process; although these images are of no immediate use, they provide a visual validation of the ACR transformations. The reverse direction path works similarly by feeding the network from the left side, with patches of the ACR-D category. By making the network create intermediate images that resemble weighted hybrids of the input/output images of the two domains, we can guide the transformation process and safeguard the overall process by utilizing the functionality of the middle node. The central point of the architecture is a CycleGAN-based model that transforms patches of ACR-B to ACR-C and vice versa. It is trained separately, in an out-of-bound parallel manner, with the gGANHopper network. At specific timestamps during the training process of the gGANHopper network, the updated middle node CycleGAN is reintroduced in the network, and the training of the gGANHopper resumes with the updated response of the central node.

To analyze the proposed gGANHopper model further, we will start with the middle node, which is based on CycleGAN. A typical architecture of CycleGAN is depicted in Figure 2. The CycleGAN is a generative adversarial network or GAN technique involving image-to-image translation without paired examples. An unsupervised training set of images from both the source and target domains is used to train the model. In our case, a pool of unpaired patch images containing only ROI patches from the breast density mask belonging to ACR-B and ACR-C categories is constructed. However, since patches from these categories are not balanced and exhibit scarcity, data augmentation techniques can be exploited to balance the original input data. In this way, we can balance the class ratio

of ACR-B and ACR-C samples of the training set. In total, 6 million image patches per category ACR-B and ACR-C are constructed and used for model generalization.

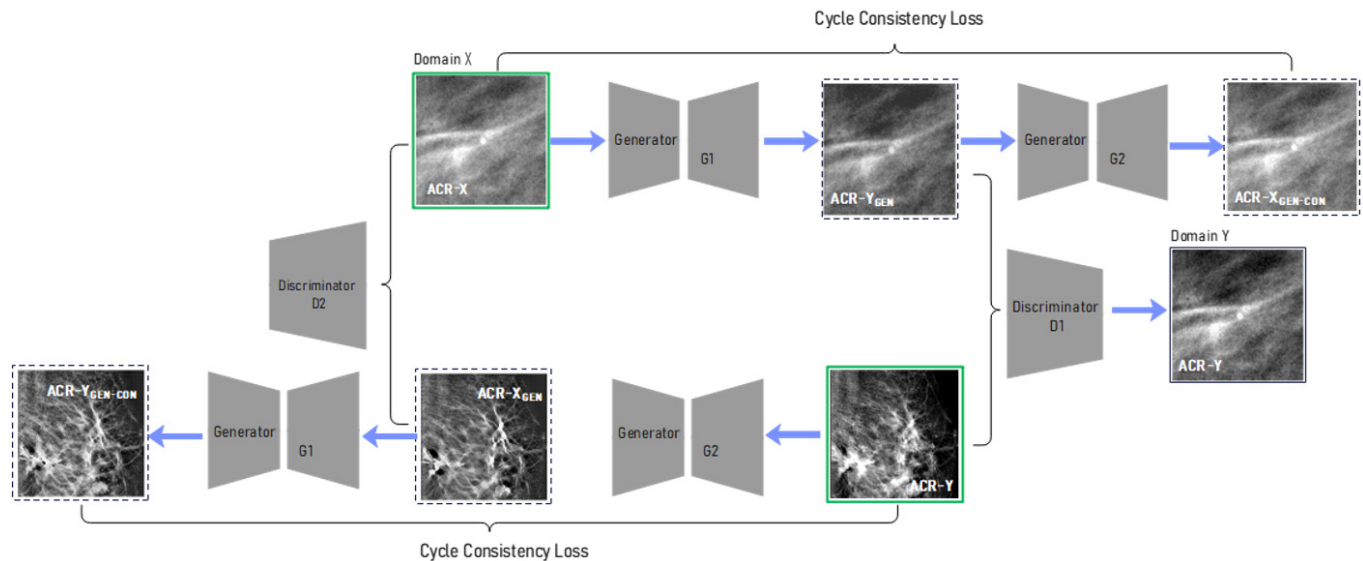


Figure 2. A typical CycleGAN Framework for image translation. Here, ACR-X and ACR-Y are real original patches (denoted with green line border) from the ACR-A and ACR-C domains, respectively, ACR-X_{GEN} are artificial patches produced by Generator G1, while ACR-X_{GEN-CON} are double-passed patches from the two Generators G1, G2 consecutively.

The GAN architecture comprises two models: a generator and a discriminator model. The CycleGAN model interlaces two generator-discriminator pairs for unpaired image-to-image translation. Images from the first domain are fed into one generator, which produces images from the second domain; similarly, images from the second domain are fed into the other generator, which produces images from the first domain in a similar manner. Subsequently, the generator models are modified following the discriminator models' evaluation of the generated images' plausibility. The backpropagating loss, also known as adversarial loss or GAN loss, is used to update the discriminators when it fails to discriminate between translated and real images.

By including an extra loss to quantify the difference between the second generator's output and the original image, as well as the other way around, the CycleGAN promotes cycle consistency. The training of the CycleGAN is running in parallel with the training of the gGANHopper network in a separate dedicated workstation, and at each checkpoint of the CycleGAN model, we are interested only in the two involved generators since the role of the middle node in the gGANHopper architecture is constrained only in the following task: given an image patch of ACR-B, or ACR-C domain produce the counterpart image patch and forward it for further processing to the rest topology of gGANHopper.

The overall gGANHopper model consists of four CycleGAN generators, denoted by G , F , G_m , F_m , and five discriminators, four of which are CycleGAN's original adversarial discriminators Disc-Y, Disc-X, Disc-Mx, Disc-My. The fifth discriminator is the hybrid discriminator Disc-H. The hybrid term assesses the degree to which an image belongs to one of the two domains (ACR-A or ACR-D). For instance, we demand that the first hop, $G_1(x)$, be judged as belonging 10% to the domain Y (ACR-D) and 90% to the domain X (ACR-A). Hence, the network capacity does not exceed that of a double-based CycleGAN network.

According to the creators of the GANHopper architecture, a hop is defined as using either G or F to warp an image towards the domain Y or X, respectively. A complete

translation is achieved by performing six hops using the same generator, G and F , except for the middle hop, where G_m and F_m are responsible for the transformations. For instance,

$$\mathbf{y}' = G \left(G \left(G_m \left(G(G(G(\mathbf{x}))) \right) \right) \right) \quad (1)$$

where $\mathbf{x} \in X$ (ACR-A) and $\mathbf{y}' \in Y$ (ACR-D). Similarly,

$$\mathbf{x}' = F \left(F \left(F_m \left(F(F(F(\mathbf{y}))) \right) \right) \right) \quad (2)$$

where $\mathbf{x}' \in X$ and $\mathbf{y} \in Y$.

Given an image $\mathbf{x}$, the translation hops are defined via the following recurrence relations, with $h \in [1, 6]$:

$$\begin{aligned} G_h(\mathbf{x}) &= G(G_{h-1}(\mathbf{x})), G_0(\mathbf{x}) = \mathbf{x}, h \neq 3 \\ F_h(\mathbf{x}) &= F(F_{h-1}(\mathbf{x})), F_0(\mathbf{x}) = \mathbf{x}, h \neq 3 \\ G_h(\mathbf{x}) &= G_m(G_{h-1}(\mathbf{x})), h = 3 \\ F_h(\mathbf{x}) &= F_m(F_{h-1}(\mathbf{x})), h = 3 \end{aligned} \quad (3)$$

The implementation for the individuals CycleGAN follows that of [27] with a network input of 256×256 . Algorithm 1 shows the training procedure of the gGANHopper, i.e., for each image to be translated, we perform a single hop, update the weights of the generator and discriminator networks, perform the next hop, pass through the middle hop, etc. A revised model for the intermediate node CycleGAN after evaluation using Fréchet Inception Distance (FID) [28] and Inception Score (IS) [29] is performed, and the process resumes from the previous best gGANHopper model in terms of total Generator/Discriminator Loss.

Algorithm 1. Training the of the gGANHopper Model

```

1: Initialization  $G, F, D_X, D_Y, D_H$ 
2: Load ( $G_m, F_m$ )
3:   for each epoch do:
4:      $\mathbf{x}, \mathbf{y} \leftarrow \text{Random\_Image\_Batch}()$ ;
5:      $\mathbf{x}_{\text{real}}, \mathbf{y}_{\text{real}} \leftarrow \mathbf{x}, \mathbf{y}$ ;
6:     for  $n = 1$  to  $h$  do:
7:       Generator_Update ( $G, \mathbf{x}, n$ );
8:       Generator_Update ( $F, \mathbf{y}, n$ );
9:       if  $n = 3$  do:
10:         $\mathbf{x}, \mathbf{y} \leftarrow G_m(\mathbf{x}), F_m(\mathbf{y})$ ;
11:        Discriminator_Update ( $D_X, \mathbf{y}, \mathbf{x}_{\text{real}}, n$ );
12:        Discriminator_Update ( $D_Y, \mathbf{x}, \mathbf{y}_{\text{real}}, n$ );
13:        Classifier_Update( $D_H, \mathbf{x}_{\text{real}}, \mathbf{y}_{\text{real}}$ );
14:      end
15:      if cycleGAN_MiddleHop_Update do:
16:         $G_m, F_m \leftarrow \text{Load} (G_{m\_upd}, F_{m\_upd})$ 
17:      end
18:    end
19:  end

```

Figure 3 depicts an excerpt of the training process (the x -axis corresponds to training steps in epochs), showing gGANHopper Generator/Discriminator Losses and the perturbation in the loss metrics after the update of the middle node.

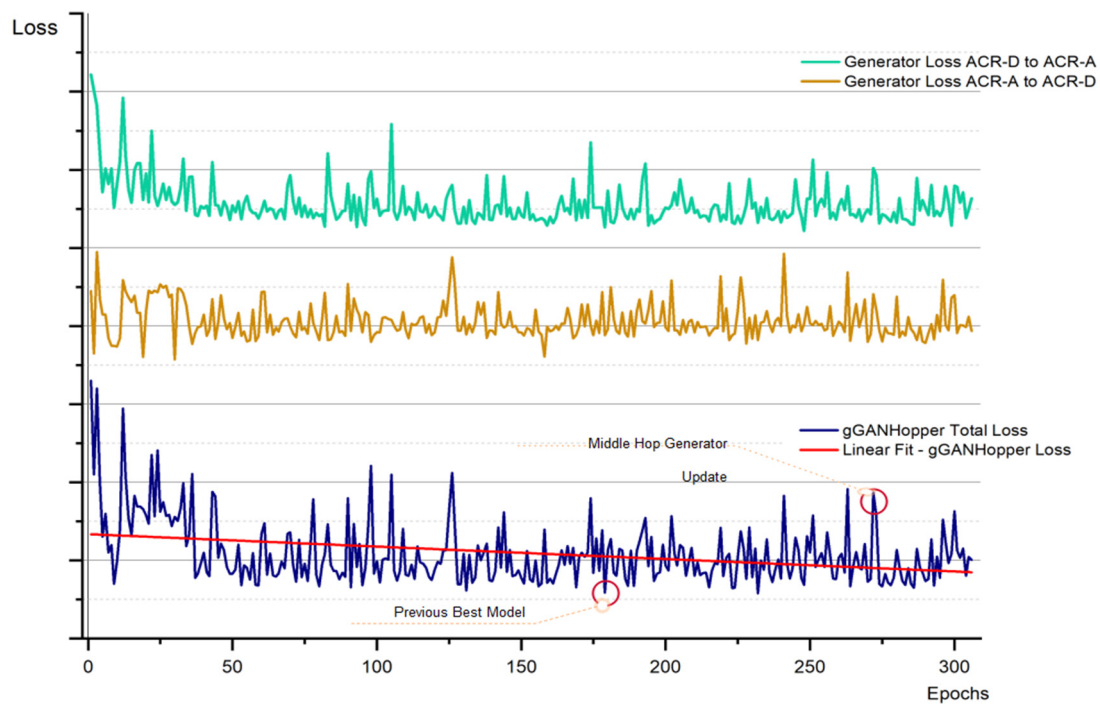


Figure 3. Generator/Discriminator and Total Loss during gGANHopper model training. Insertion of the updated middle node CycleGAN model and continuation of training.

A sample from the middle node is presented in Figure 4. An input patch of ACR-B, or ACR-C, results in an image from the opposite domain ACR-C or ACR-B. The dedicated role of the middle node is to guard the overall ACR breast density transformation process.

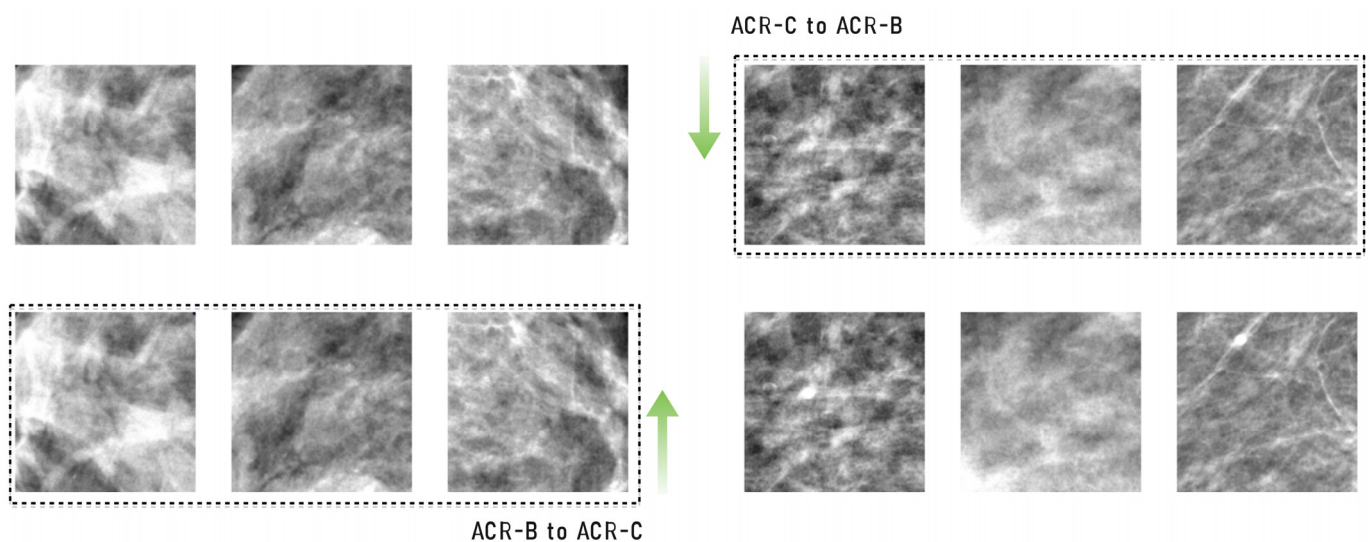


Figure 4. ACR-C to ACR-B translation via middle node. The dotted areas display patches generated via middle mode CycleGAN when fed with appropriate patches from ACR-B or ACR-C. These patches are propagated to the next hops towards the ACR-A or ACR-D domains.

Figure 5 depicts an end-to-end illustration of the gGANHopper network, a sample input of ROI inside the breast density mask of type ACR-D passing the nodes of the network (ACR-D (left) to ACR-A (right)).

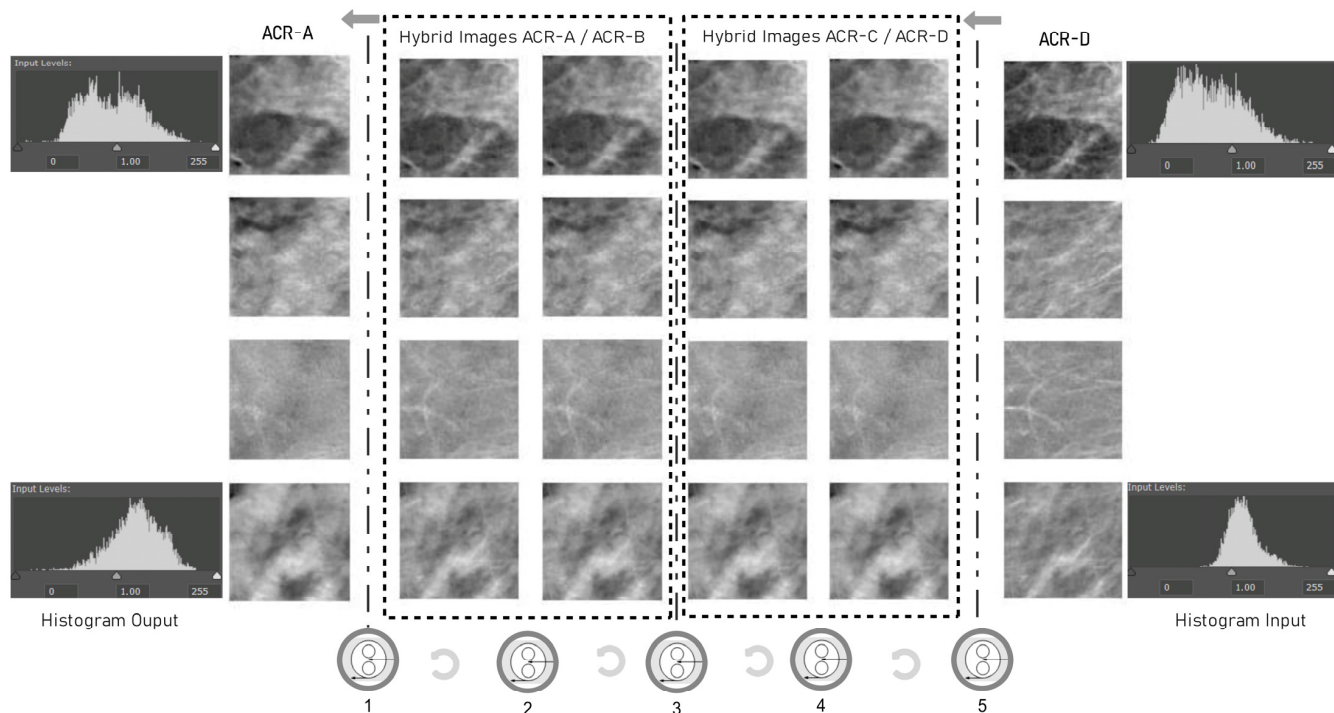


Figure 5. ACR-D to ACR-A translation. Histogram Visualization of input/output images on input/output.

The image transformations applied (especially for the direction ACR-D to ACR-A) aim to reduce the masking effect caused by thick tissue and reveal malignancies from surrounding tissue.

Given an ROI patch of a specific ACR category, i.e., ACR-B, a question that arises is which direction the transformations will follow, in other words, whether we are going to transform it via three hops to ACR-A (traversing the gGANHopper network to the left) or via three hops to ACR-D (traversing the gGANHopper to the right) and six hops to left, to obtain the desired output type of ACR-A. In all experiments, we took the longer distance approach in terms of hops since we wanted to give the model the maximum degrees of freedom to apply the learned transformations on the input patch and take advantage of the expertise of the middle node. The inheritable cycle consistency in each hop enforces forward and backward consistency of the transformations.

4. Experimentation Environment

All experiments were performed on an NVIDIA RTX 3090 24 GB, GDDR6X (using a batch size of 6 for the gGANHopper and a batch size of one for the middle node CycleGAN). The Deep Learning models were all implemented in Python 3.8, in Ubuntu 20.04, with TensorFlow 2.8 and Keras 2.8 API. The unpaired sets of ACR-A and ACR-D contained about 8 million images for each category. In contrast, the odd set used for training the middle CycleGAN (ACR-B and ACR-C) had about 6 million images per category. The datasets were augmented via the albumentations library (<https://albumentations.ai/> (accessed on 1 September 2024)), where various image transform operations were performed (rotation, flipping, scaling, zooming, intensity changes, etc.). The network takes 256×256 images as input and outputs images of the exact resolution. An entire epoch of the modified gGANHopper model takes about one month, while the training of the middle node CycleGAN was performed in parallel, in an out-of-bound manner. The central node CycleGAN was left for 70 epochs to be trained in a separate workstation set up with a similar configuration, while the whole gGANHopper was trained for ten epochs. Whenever the middle node CycleGAN has a better model to offer to the gGANHopper after FID and IS evaluation,

the previous best gGANHopper model (in terms of total Generator/Discriminator loss) is loaded, and the training continues (Figure 3).

5. Datasets

For our study, we analyzed a set of different mammography databases that provide information regarding breast density with ACR characterization, boundaries, and the type of malignancy. To evaluate our approach, we used various types of mammograms, ranging from fully digital to scanned X-ray film. The datasets and their statistics are presented briefly in the section below.

5.1. The MIAS Dataset (Film Dataset)

The Mammographic Image Analysis Society (MIAS) [30] dataset consists of 322 film mammograms (106 fatty and 216 dense images).

Annotations are given in a separate file containing the background tissue type, the class and the severity of the abnormality, the x and y coordinates of the center of the irregularities, and the approximate radius of a circle enclosing the abnormal region in pixels. This dataset has no annotation regarding ACR breast tissue level; the creators have used the old ACR characterization of Fatty/Fatty-Glandular/Dense-Glandular. In the older BI-RADS edition, the assignment was based on the four ACR (American College of Radiology) categories based on the fibroglandular tissue density in the breast. However, in the 2013 edition, the standard was changed to introduce categories ACR-A, ACR-B, ACR-C, and ACR-D, which consider the likelihood that a mass may be obscured by fibroglandular tissue. We used a VGG16-based model for this dataset to categorize the ROIs according to the corresponding BI-RADS ACR breast density categories.

5.2. The CBIS-DDSM Dataset (Film Dataset)

The Curated Breast Imaging Subset of DDSM (CBIS-DDSM) is an updated and standardized version of the Digital Database for Screening Mammography (DDSM) [31,32] and is provided by the University of South Florida. It contains film mammograms digitized using four different types of digitizers. The database contains approximately 2500 studies. Each study includes two images (views) of each breast, as well as some associated patient information (age at time of study, ACR breast density rating, subtlety rating for abnormalities, ACR keyword description of abnormalities) and image information (scanner used for the digitization, scanner spatial resolution, etc.). The ACR keyword description of the database was matched to the ACR BI-RADS categorization.

5.3. The INbreast Dataset (Digital)

The INbreast dataset [33] was acquired from the Breast Centre in CHSJ, Porto. The image sizes are 3328×4084 or 2560×3328 pixels in DICOM format. The DB contains findings of six types: asymmetries, calcifications, distortion, masses, multiple findings, and normal. The annotations about ROIs are in XML format. Information is also included regarding the patient's age during image acquisition, family history, ACR breast density annotation, and BI-RADS classification. It has a total of 115 cases (410 images).

5.4. The SuReMaPP Dataset (Digital)

The SuReMaPP [34], published recently, consists of 343 mammograms hand-labeled by expert radiologists to identify suspicious regions such as abnormalities (benign and malignant) and calcifications. SuReMaPP contains mammograms with ACR keyword descriptions corresponding to the ACR BI-RADS specification.

5.5. The CDD-CESM Dataset (Digital)

The Contrast-Enhanced Spectral Mammography (CESM) Dataset [35,36] is a dataset of CESM images from women with suspicious breast lesions gathered from the Radiology Department of the National Cancer Institute, Cairo University, Egypt. The images are all

high resolution with an average of 2355×1315 pixels. Institutional review board approval and patient informed consent to carry out and publish data were obtained from 326 female patients aged from 18 to 90 years. The dataset contains 2006 images with CC and MLO views (Each view comprises two exposures, one with low energy and one with high energy. Low and high-energy images are then recombined and subtracted using appropriate image processing algorithms to suppress the background breast parenchyma).

5.6. The VinDR Dataset (Digital)

The VinDr-Mammo [37] dataset is a large-scale dataset of full-field digital mammograms comprising 5000 four-view examinations accompanied by breast-level assessments and findings annotations. VinDr-Mammo contains approximately 20,000 scans of full-field digital mammograms that also categorize breast-level BI-RADS assessment with suspicious or possible benign findings that require follow-up examination and ACR breast tissue/finding level annotations.

In Table 1 and Figure 6, we summarize the details of the datasets used in our experiments. The breast density pie chart denotes the total number of mammograms per ACR category. The results section will discuss and interpret this composition.

Table 1. Individual Datasets and annotation details.

	MIAS	CBIS-DDSM	INbreast	SuReMaPP	CDD-CESM	VinDr
Unique Cases	161	1566	115	145	326	5000
Orientation						
RMLO/RCC	161	1756	206	163	503	10,289
LMLO/LCC	161	1682	204	180	500	10,197
Findings						
Malignant	54	1936	1457	57	869	988
Benign	69	2628	269	73	738	18,232
Normal	207	1797	67	100	340	5606
Breast Density						
BI-RADS						
ACR-A	109	889	136	244	207	50
ACR-B	104	2406	146	100	604	954
ACR-C		1903	99		1005	7646
ACR-D	117	1161	28	0	131	1350

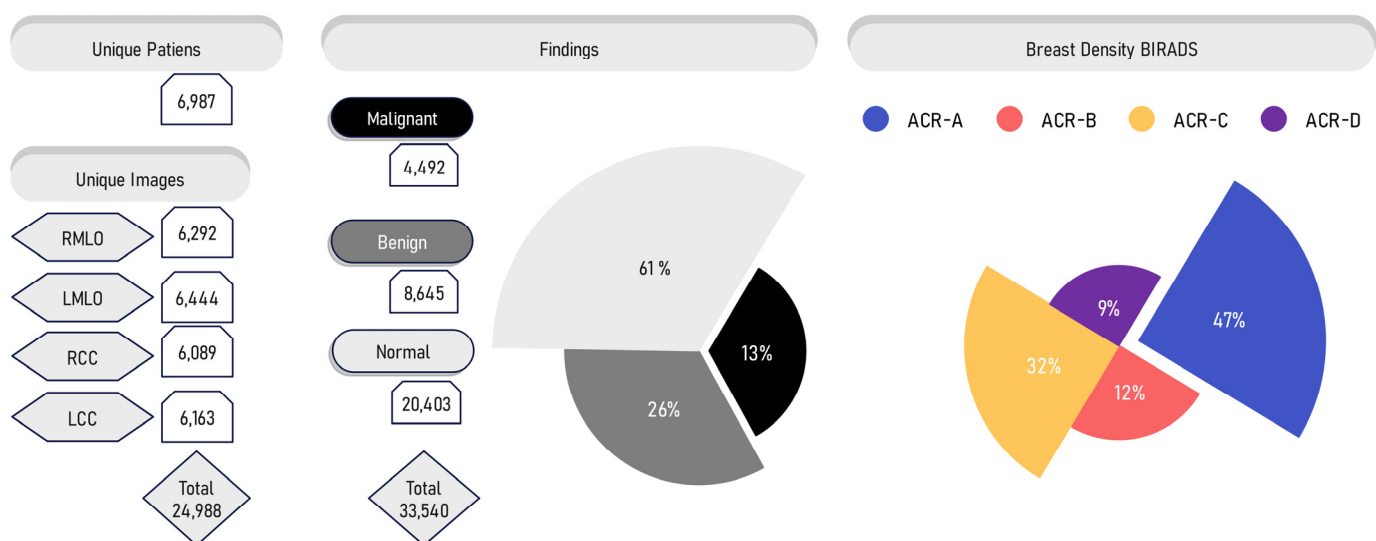


Figure 6. Total Statistics for all mammographic datasets used.

6. System Overview

The complete system from end to end is illustrated in Figure 7.

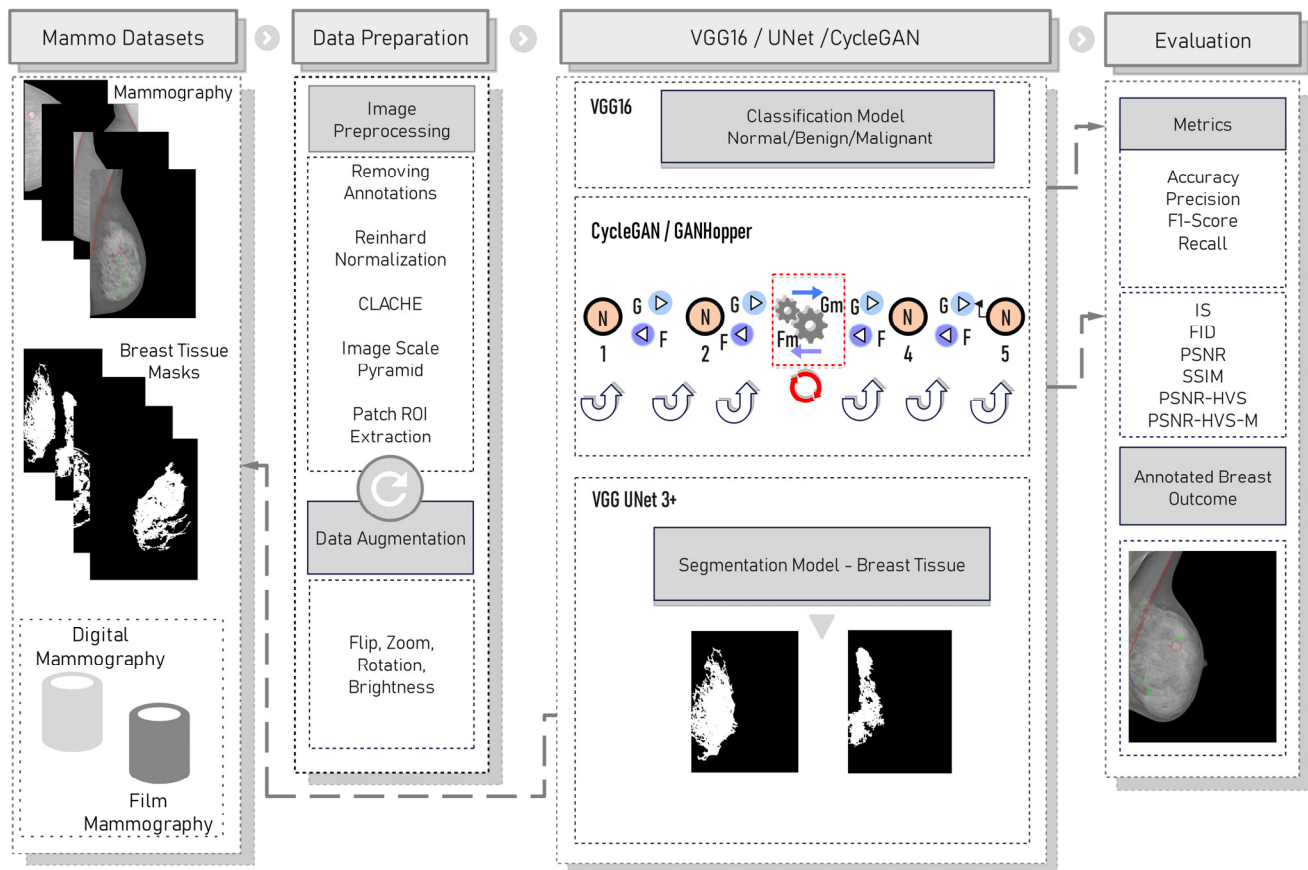


Figure 7. System Block Overview.

The system consists of four main modules: (a) The collection of mammographic datasets in which the Breast Tissue Masks are obtained by exploiting a UNet 3+ [38] based model. (b) the Data Preparation module where image preprocessing (removal of annotations, intra-transfer of datasets statistics [39], CLACHE [40,41], and patch extraction, i.e.,) on image scale pyramid basis, and Data Augmentation (Flip, Zoom, Rotation, Brightness Adjustments, i.e.,) takes place, (b) the Deep Learning module where breast tissue segmentation, breast density transformation and suspicious region detection is performed and finally (c) the evaluation module where annotation of abnormal regions on the given mammogram under examination is performed. To make our system more robust and to consider the differences between mammograms during examination (fully digital and scanned X-ray film), we thoroughly tested our system using five different datasets containing digital and film-scanned mammograms.

An image characteristic transfer method was applied to eliminate the differences in the intensity levels of the mammograms used in the databases, propagating INbreast dataset statistics to all other mammographic sets. This preprocessing normalization is critical because it can leverage large intensity variations typically attributed to using different scanners with variable parameters in the image digitization process. To ensure objective image comparison between different mammograms, a normalization algorithm is used in advance to modify the distribution of intensity values in each scan and match the selected baseline image. In this preprocessing step, adopted from [39], the color distribution of a mammographic captured image is mapped to that of a well-acquired target image (a.k.a Reinhard Normalization). The authors described a method for matching an image's color distribution to that of a target image using a linear transform in a perceptual color space to match the mean and standard deviation of each color channel in the two images in that color space. This method is also applicable for grayscale images, as the ones included in the

mammogram datasets, where color-discarding is performed to obtain a grayscale image as the outcome.

The discrete steps for the Reinhard Normalization are:

Step 1: Conversion of both source image X and target set image Y to the LAB space [42]. The baseline target, in our case, is the INbreast dataset.

Step 2: Compute the intensity variables $\{l_2, a_2, b_2\}$ of the processed image in the LAB color space. In the following equations, $\{l_1, a_1, b_1\}$ are the mean intensity variables of the target set in the LAB space, $\{l, a, b\}$ are the intensity variables of the source image, μ_g indicates the global mean of the image and σ_g represents the global standard deviation of the image.

$$l_2 = \mu_g(l_1) + (l - \mu_g(l)) \times (\sigma_g(l_1)/\sigma_g(l)) \quad (4)$$

$$a_2 = \mu_g(a_1) + (a - \mu_g(a)) \times (\sigma_g(a_1)/\sigma_g(a)) \quad (5)$$

$$b_2 = \mu_g(b_1) + (b - \mu_g(b)) \times (\sigma_g(b_1)/\sigma_g(b)) \quad (6)$$

Step 3: Conversion of the processed image Z from the LAB to the RGB color space and finally to grayscale.

However, the Reinhard Normalization method has certain limitations, which are outlined as follows:

1. The background luminance of the source image is not preserved in the processed image, as Equation (4) states. For this reason, the Reinhard method is applied only to the region inside the breast mask and not to the whole mammogram.
2. When the source image has greater contrast than the target image, it also transfers the contrast statistics from the target to the source image, which may lead to a processed image with a smaller contrast than the source image. This can be corrected by applying contrast-improving techniques such as CLACHE.
3. Due to the global transfer of the mean color from the target image to the source image, the same mean color is also transferred to all the pixels in the image. Thus, a sizeable homogeneous portion associated with white luminance in the source image can be affected by a faded color by this method. Histogram equalization and CLACHE can alleviate this.

To help radiologists detect abnormalities by improving image contrast while removing noise enhancement, a technique known as contrast-limited adaptive histogram equalization (CLAHE) is used. The CLAHE technique is used to enhance the contrast of images.

The main steps of the CLAHE technique are:

Step 1: Partition the image into non-overlapping and continuous tiles.

Step 2: Each tile's above-threshold clipped histogram is distributed evenly throughout the entire range of intensity values.

Step 3: Application of histogram equalization (HE) of each tile.

Step 4: Bilinear interpolation mapping between adjacent tiles. Each mapped pixel results from interpolating the four adjacent tiles' intensity maps.

The adaptive approach redistributes the image's brightness values by calculating numerous histograms, each corresponding to a different area. This differs from regular histogram equalization (HE) or adaptive histogram equalization (AHE). As a result, it can strengthen edge definition and local contrast in each region of the mammogram. On the other hand, details may be missed in some areas of a mammogram due to over-enhancement.

Since the used datasets lack this information (breast tissue masks), we estimate the breast tissue masks using a UNet 3+ network for breast tissue segmentation. In ref. [43], the UNet architecture was first suggested for segmenting biomedical image data. Feature maps are fully utilized by the UNet 3+ to increase computing performance and decrease network parameters while increasing accuracy. Compared to its predecessors, UNet and UNet++, UNet 3+ has fewer parameters because it redesigned skip connections to account for full-scale information. UNet 3+ boosted its breast tissue mask segmentation performance by

(1) using its deep supervision method and adjusting it to incorporate full-scale semantic data. (2) Modifying the dense connection architecture to better accommodate for segmentation using both high and low levels of detail from feature maps. A typical UNet 3+ is depicted in Figure 8.

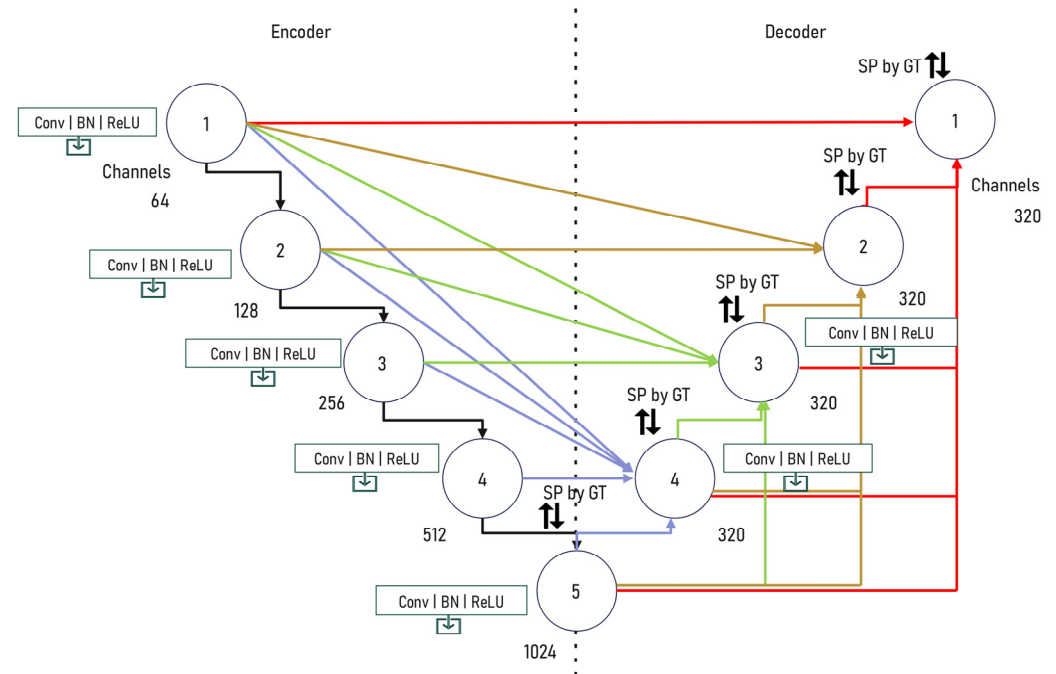


Figure 8. UNet 3+ architecture used for Breast Tissue Segmentation. (SP: denotes supervision by ground truth. Colored arrows represent two types of skip connections: downward arrows for full-scale intra connections and upward arrows for full-scale inter-connections. Black arrows indicate standard downsampling paths. The central dotted line separates encoder-decoder sections while highlighting the symmetrical UNet topology).

UNet3+ consists of five main parts: the encoder module (down-sampling), the decoder module (up-sampling), the full-scale skip connection module, the full-scale feature supervision module, and the classification guidance module.

The encoding part is the same as that of UNet. The encoder uses pooling and convolution to reduce the image size, which reduces the image resolution and causes some detailed information to be lost. The full-size jump connectivity changes the interconnections between the encoder and decoder and the internal connections within the decoder. Each decoder layer in UNet3+ incorporates small-scale and same-scale feature maps from the encoder and large-scale feature maps from the decoder, which capture fine-grained and coarse-grained semantics at the full scale. The Full-scale deep supervision is proposed on UNet3+ to generate a broadside output supervised by ground truth at each decoding layer.

In continuity, an image pyramid in different scales is constructed for each mammographic image, and a vast number of halve non-overlapping patches are extracted inside the breast tissue mask to feed the VGG16-based classifier. The classifier is responsible for classifying the extracted patches to Normal/Benign/Malignant.

Data scarcity is among medical imaging AI research's most common limiting factors in diagnosis tasks where a patch-based approach is followed, as here, extracting multiple samples from one image can aid in overcoming the data scarcity issue to some extent. As seen from the total statistics of the datasets used in Figure 6 and Table 1, patches from category ACR-D are less than the other three BI-RADs ACR categories. We have utilized the extraction of patches under multiple zoom scales in an overlapping manner and various augmentation variations, including rotation, variations in brightness, flipping, etc., to construct almost balanced pairs of ACR-A/ACR-D and ACR-B/ACR-C categories.

Although balancing the sets is not a strict requirement for a CycleGAN model to work, it is preferable. Otherwise, some pair inputs may be seen more times than others in the training phase. With the above approach, two unpaired sets were constructed. The first set contained patches of ACR-A and ACR-D categories, a total of 8 million patch images, while the second set had around 6 million patches of ACR-B and ACR-C categories. The second set was used to train only the middle CycleGAN node, while the first was used for training the gGANHopper model in charge of transforming a patch of type ACR-A to ACR-D type and vice-versa via the middle CycleGAN-based model.

7. Experimental Results

7.1. Image Quality Evaluation

To evaluate the quality of the synthetically produced images by the proposed GAN network, we have used the Fréchet Inception Distance (FID) [28] and the Inception Score (IS) [29] described by Equations (7) and (8). The FID and IS metrics were introduced to assess the quality and similarity of images generated by generative adversarial networks (GAN).

$$FID = ||\mu_r - \mu_g||^2 + Tr\left(\Sigma_r + \Sigma_g - 2\sqrt{(\Sigma_r \Sigma_g)}\right) \quad (7)$$

(a metric for quantifying the realism and diversity of images generated by GANs)

$$IS = \exp\left(E_{x \sim p_g} DKL(p(y|x)||p(y))\right) \quad (8)$$

(a measurement of distance between distributions of real and generated images by GANs)

The Inception Score was calculated using the Inception v3 neural network, which was trained on the ImageNet dataset for image classification. This score combines the confidence of the conditional class predictions for each synthetic image (quality) and the integral of the marginal probability of the predicted classes (diversity). The metric relies on extracting features from real and generated images using the Inception network and then comparing the distribution of the extracted features using the Kullback-Leibler (KL) divergence. The lowest IS can be zero, and the highest IS can be infinity. A higher IS is always better. The FID extends the inception distance calculation using the Fréchet distance to compare the actual and generated image feature distributions, which measures the distance between two multivariate Gaussian distributions. The lower FID score represents that the quality of images generated by the generator is higher and similar to the real ones. Figure 9 depicts the inception score for the used datasets.

Each dataset's Inception Score (IS) is low since the GAN model is trained on classes outside the Inception Network's known cases.

Figure 10 depicts the variation of the Fréchet inception distance (FID) for all the breast density-transformed datasets measured at frequent checkpoints during the training procedure. Since our method generates images with completely different label sets than those in ImageNet's training set, we expect the FID metric to be low in our experiments. A high score would indicate that the generated images lack intra-class diversity or that the generator uses training data repetitively.

Image quality assessment (IQA) is necessary to quantify differences between images accurately. Subjective and objective IQA techniques are generally used to assess an image's quality. In a subjective measure, an image is examined by persons (usually a group of experts) who then provide their assessment. This metric is considered the best approach as it allows uniformity in evaluating the images. However, conducting a subjective measure can be costly, time-consuming, and uncomfortable to yield a trustworthy result. This makes sense since human engagement is closely related to their ability, knowledge, and insight. For example, a radiology-trained physician is required to examine breast medical imaging.

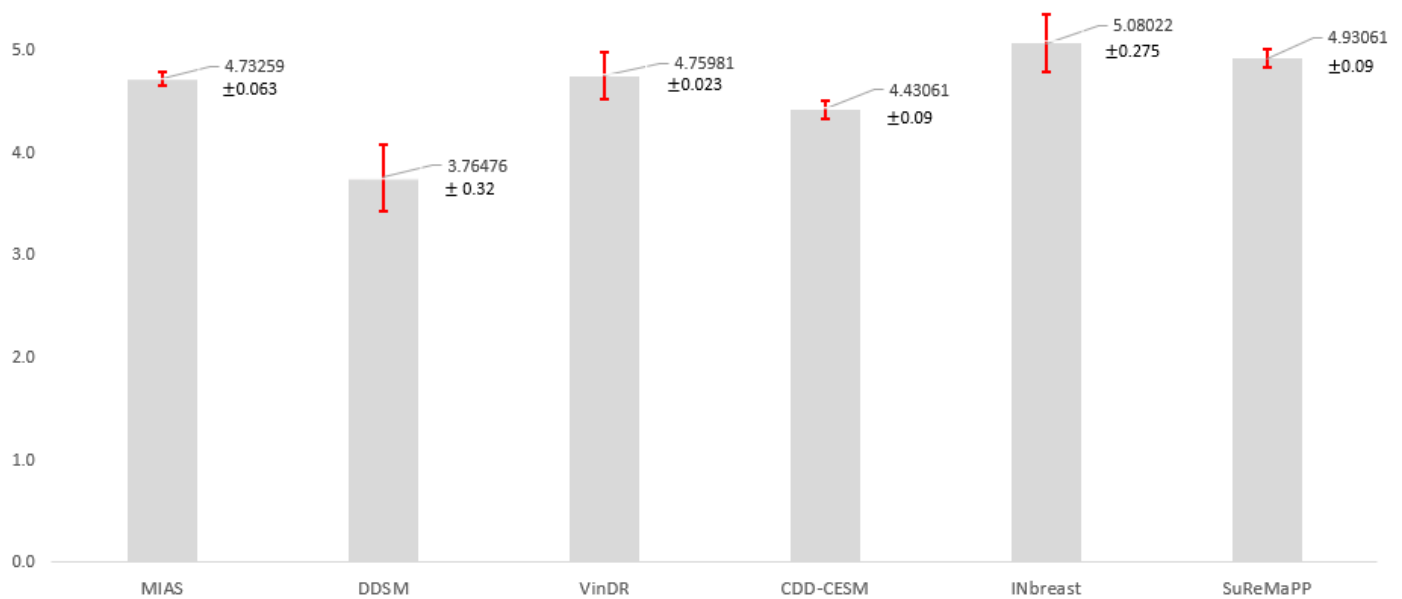


Figure 9. Fréchet Score variation during training and model checkpoints.

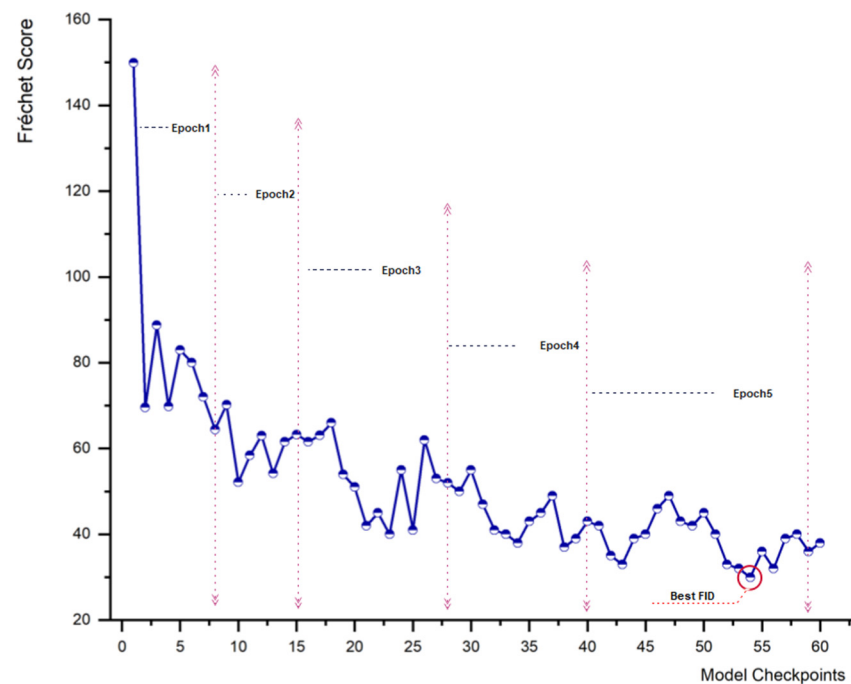


Figure 10. Fréchet Score variation during training and model checkpoints. The red circle marks the best FID found at epoch 6 (each epoch is limited in the area between the red arrow areas) and model checkpoint 54 during training.

Over the past few decades, various appealing metrics connected to objectively assessing such images have been established in recognition of the drawbacks of a subjective approach. The peak signal-to-noise ratio (PSNR) and the structural similarity measure (SSIM) [44] are the most often used metrics for evaluating image quality independently of the image field. While these metrics are acknowledged as a universal indicator of image quality, their sensitivity is lower than that of a human visual system (HVS). Because they are straightforward to compute and have the ability to decipher an image's physical meaning, they are regarded as standard measures. The human visual system's contrast perception is

incorporated into PSNR-HVS [45], an extension of PSNR. PSNR-HVS-M [46] builds upon PSNR-HVS by further accounting for visual masking.

PSNR concentrates on pixel-by-pixel comparison. Although a higher PSNR generally indicates that the transformation is of higher quality, in some cases, it may not. SSIM is correlated with the quality and perception of the human visual system. Instead of using traditional error summation methods, the SSIM measures the difference in the image transformations as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. SSIM values towards 1 denote transformations that produce images almost identical to the source image. PSNR performs poorly in discriminating structural content in images since various types of degradations applied to the same image can yield the same outcome. For PSNR-HVS, image differences are reflected in mean square errors weighted in the DCT domain with specific weights. Any changes in the contrast or shifts of the mean intensity led to non-zero MSE, resulting in smaller PSNR-HVS values. Larger values correspond to better visual quality regarding the source image.

For each ACR-transformed image patch of the testing set, we calculate the PSNR, SSIM, PSNR-HVS, and PSNR-HVS-M metrics in two experiments: The first experiment examines the GAN model's cycle consistency, and it is used to assess the quality of the generated patches. In this experiment, we transform each patch's density back to its initial by following the shortest path of the five-node GAN model. In particular, if the patch is of type ACR-B, we convert it to ACR-A via nodes 2 and 1 and back to ACR-B following the reverse path. This experiment establishes the maximum expected quality of the generated images. Table 2 shows the average IQA metrics for the datasets. This experiment sets the ceiling for the expected quality of the generated images. Table 3 displays the average IQA metrics for the datasets.

Table 2. Quality Metrics for GAN cycle consistency per dataset.

Dataset	PSNR	SSIM	PSNR-HVS	PSNR-HVS-M
MIAS	19.74	0.68	13.05	13.92
CBIS-DDSM	19.51	0.73	170.25	171.97
VinDR	23.44	0.76	22.50	23.10
CDD-CESM	23.05	0.76	23.61	24.05
INbreast	25.47	0.86	19.53	22.50
SuReMaPP	25.16	0.83	63.34	65.41

Table 3. Quality Metrics for GAN-transformed patches.

Dataset	PSNR	SSIM	PSNR-HVS	PSNR-HVS-M
MIAS	19.38	0.70	13.40	14.33
CBIS-DDSM	19.03	0.78	169.76	171.3
VinDR	26.10	0.85	24.3	25.1
CDD-CESM	21.36	0.86	21.35	22.74
INbreast	25.27	0.86	19.35	22.06
SuReMaPP	24.85	0.88	31.27	33.32

The following evaluation is performed on all transformed patches via the 6-hop CycleGAN/gGANHopper-based model, and it is used to denote the extent of changes applied when all the patches are converted to the ACR-A category. Table 3 displays the average IQA metrics for this experiment. From the comparisons of Tables 2 and 3, we can estimate the extent of changes imposed by the GAN model in the patches when applying the based breast density transformations.

For comparative purposes, we cite that if the bit depth is 8 bits, the accepted values for the PSNR in lossy image compression applications are between 30 and 50 dB, with higher values being better. A smaller PSNR-HVS than PSNR indicates that the distortions caused by the breast density modifications are more akin to spatially correlated noise, non-Gaussian (heavy-tailed) noise, or both [47]. A masking effect of the image content is

sufficient if PSNR-HVS-M is significantly larger than PSNR (this indicates that the image is presumably very textural). Conversely, if PSNR-HVS-M is about equal to or less than PSNR, distortions resemble spatially correlated noise, and/or there is no masking effect. Recall that distortion visibility thresholds are about 35–38 dB according to the metric PSNR and about 40–42 dB according to the metric PSNR-HVS-M [47].

7.2. Recognition Accuracy

To assess the performance of the suggested technique for each of the previously described mammographic databases, we employed the Precision, Recall, Accuracy, and F_1 -score metrics evaluated as follows:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \times 100\% \quad (9)$$

(The percentage of correct positive predictions to total positive predictions.)

$$\frac{\text{Recall}}{\text{Sensitivity}} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \times 100\% \quad (10)$$

(The ratio between the number of correct positive predictions and actual positive observations.)

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}} \times 100\% \quad (11)$$

(The ratio of the number of correct predictions to the total observations)

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (12)$$

(The F_1 score represents the harmonic mean of both Precision and Recall.)

Figures 11–16 demonstrate the CNN model's performance based on the VGG16 classifier before and after the ACR transformations. Across all the figures, the measured metrics for each dataset used have significantly improved. The results on the left of each vertical dotted line (i.e., in Figure 11 denoted as Before Breast Density Transformation) correspond to the initial CNN performance, while the right part of the dotted line (i.e., in Figure 11 marked as After Breast Density Transformation) corresponds to the CNN's performance after applying the proposed density transformations via our model.

Similarly, the results for the rest of the datasets used are presented in the following figures.

For the CBIS-DDSM dataset (Figure 12), the precision was boosted from 67% to 85% after gGANHopper transformations.

For the CDD-CESM dataset (Figure 14) and INbreast dataset (Figure 15), the precision was boosted from 79% to 95% and from 88% to 99% after gGANHopper transformations, respectively.

For the SuReMaPP dataset (Figure 16), the precision increased from 90% to 97% after gGANHopper transformations.

Figure 17 presents a graphical summary of the experimental results across all datasets used in this study. The interpretation of this figure is as follows:

The overall classification performance of the VGG16-based classifier improved significantly, increasing from an initial accuracy of 72% to 95% through three steps of analysis. In Step 1, before applying the gGANHopper transformations, the classifier showed varied error rates across different ACR categories. Specifically, patches in the ACR-C category exhibited the highest error rate (69.7%), followed by ACR-B (13.8%), ACR-D (13.3%), and ACR-A (3.3%). This indicates that the classifier was most error-prone for ACR-C patches and least for ACR-A. Notably, ACR-B and ACR-D demonstrated similar contributions to the classification errors.

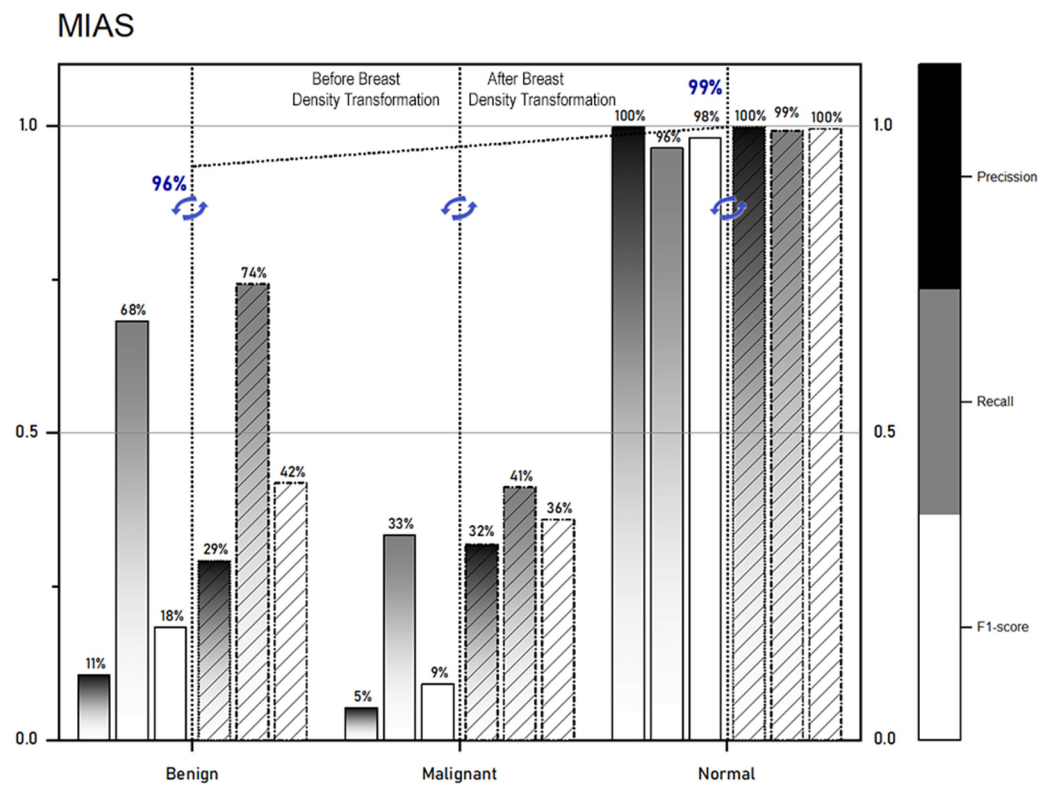


Figure 11. Precision, Recall, F1-score for MIAS dataset, before and after GAN-based model Breast Density transformations.

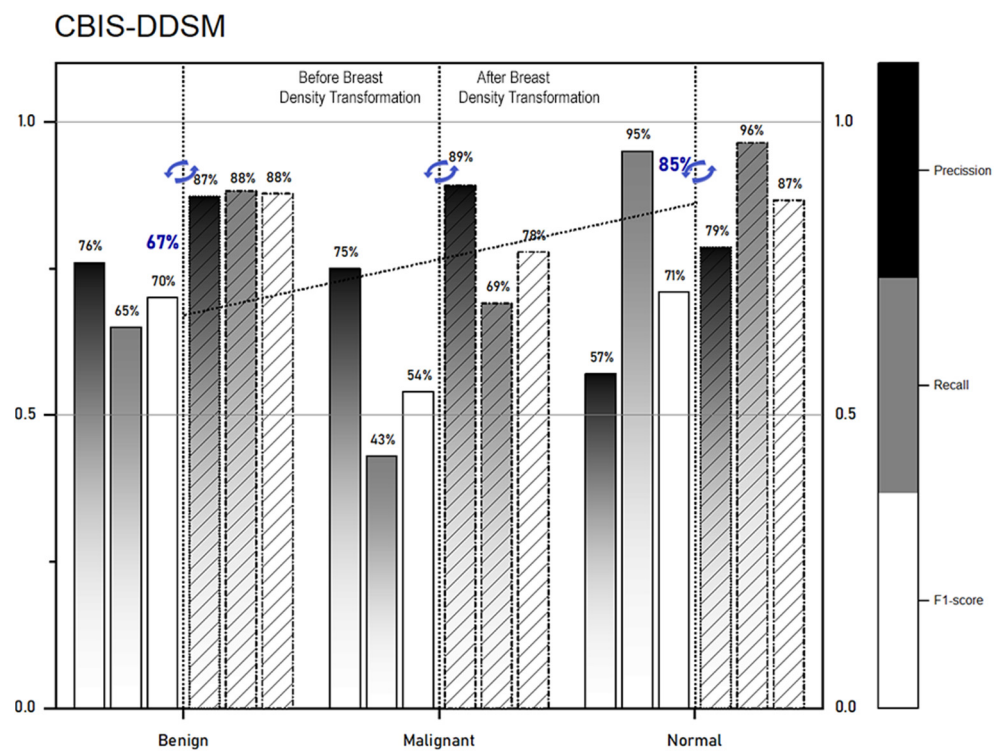


Figure 12. Precision, Recall, F1-score for CBIS-DDSM dataset, before and after GAN-based model Breast Density transformations.

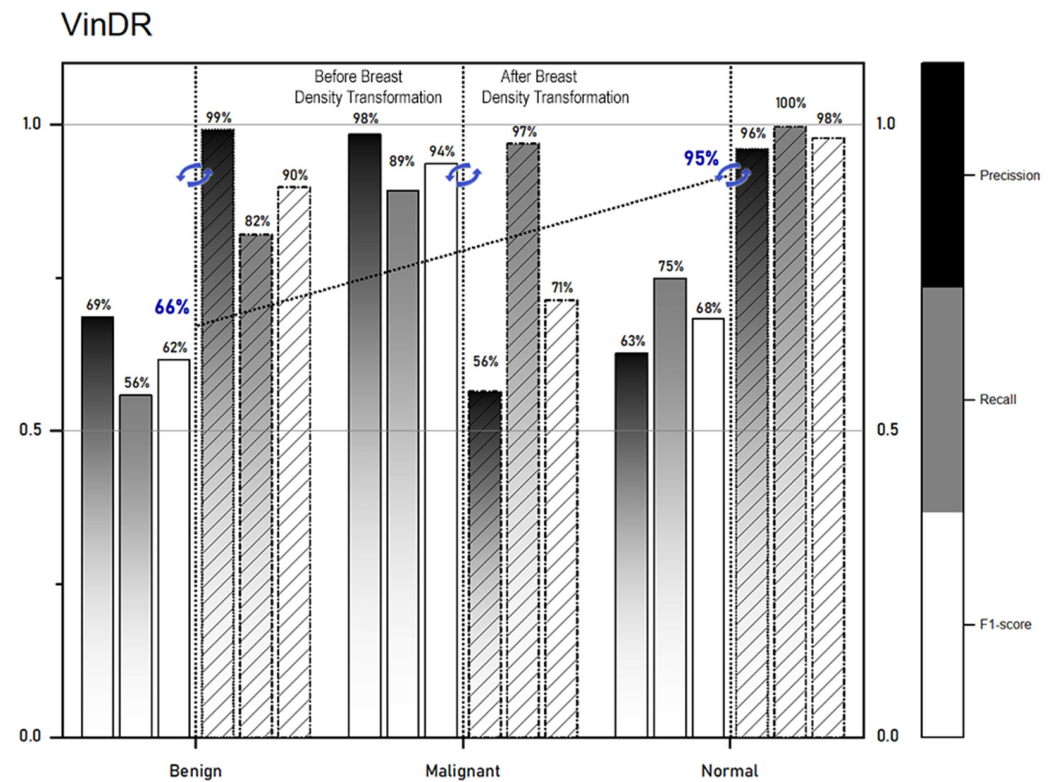


Figure 13. Precision, Recall, F1-score for VinDR dataset, before and after GAN-based model Breast Density transformations.

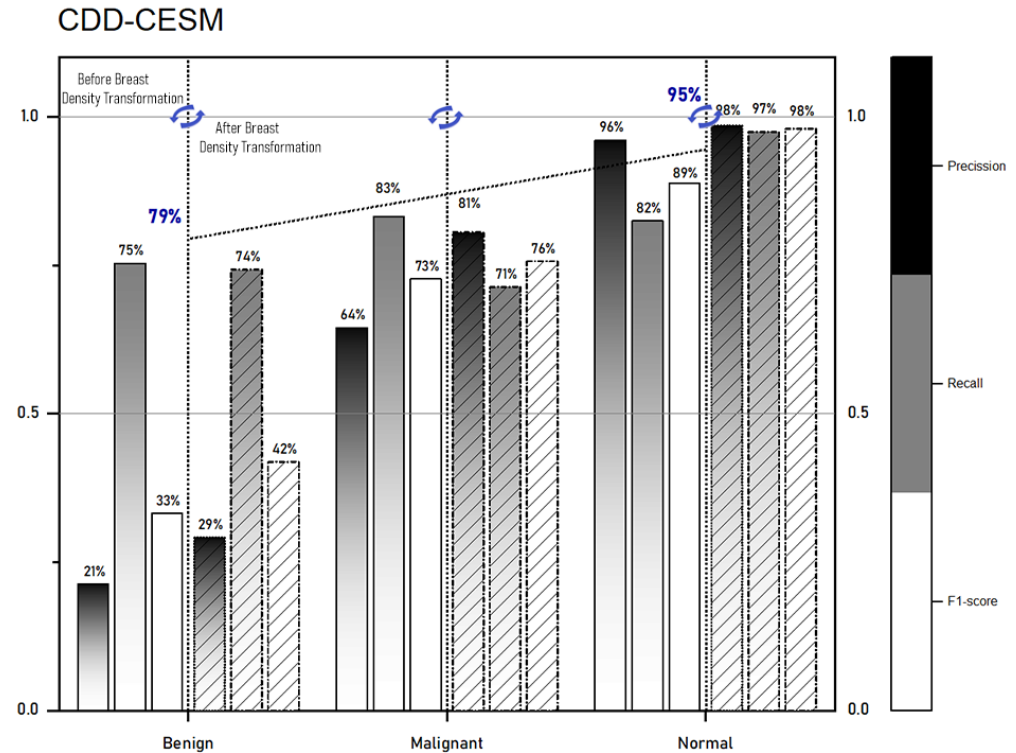


Figure 14. Precision, Recall, F1-score for CDD-CESM dataset, before and after GAN-based model Breast Density transformations.

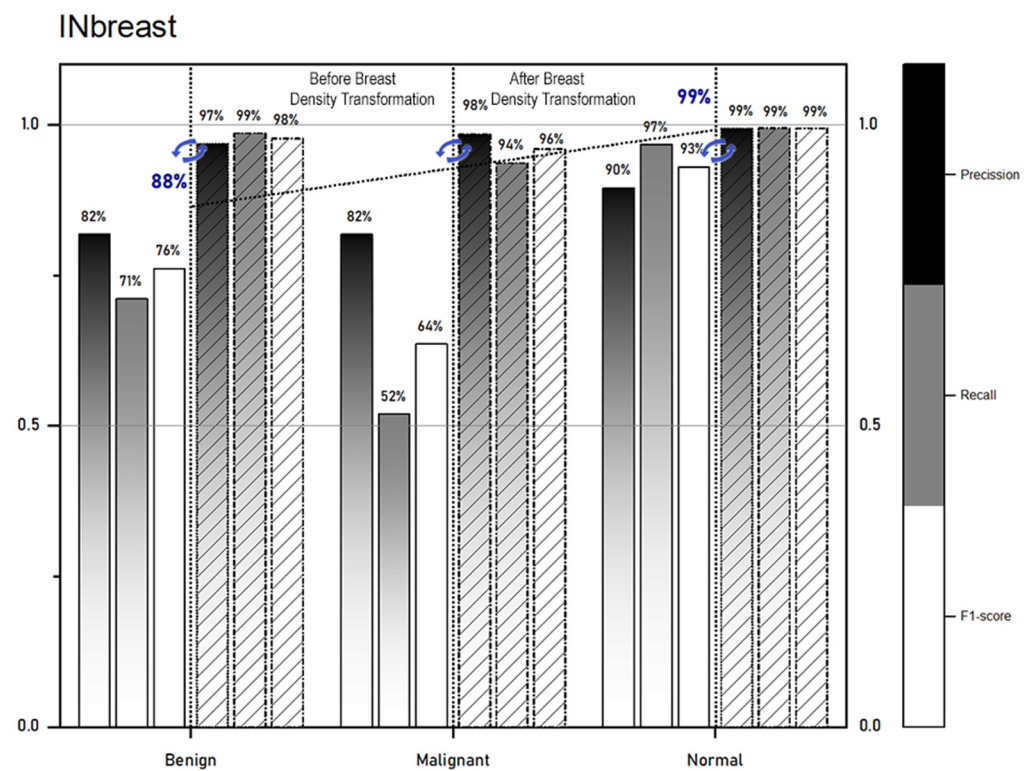


Figure 15. Precision, Recall, F1-score for INbreast dataset, before and after GAN-based model Breast Density transformations.

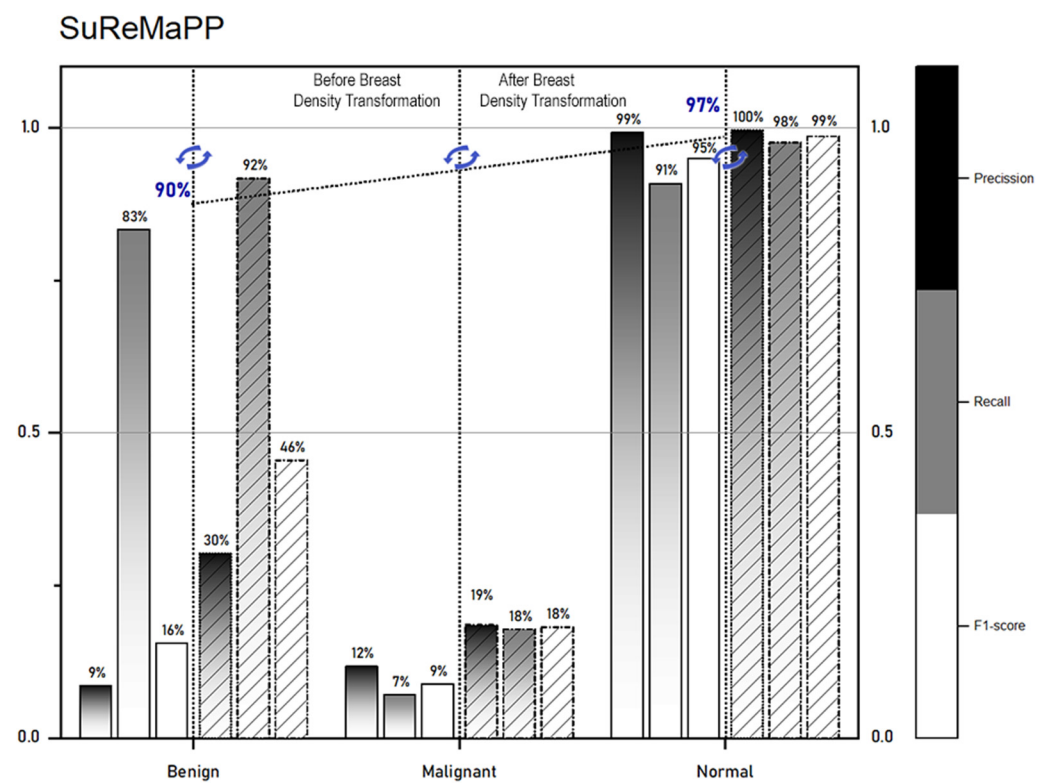


Figure 16. Precision, Recall, F1-score for SuReMaPP dataset, before and after GAN-based model Breast Density transformations.

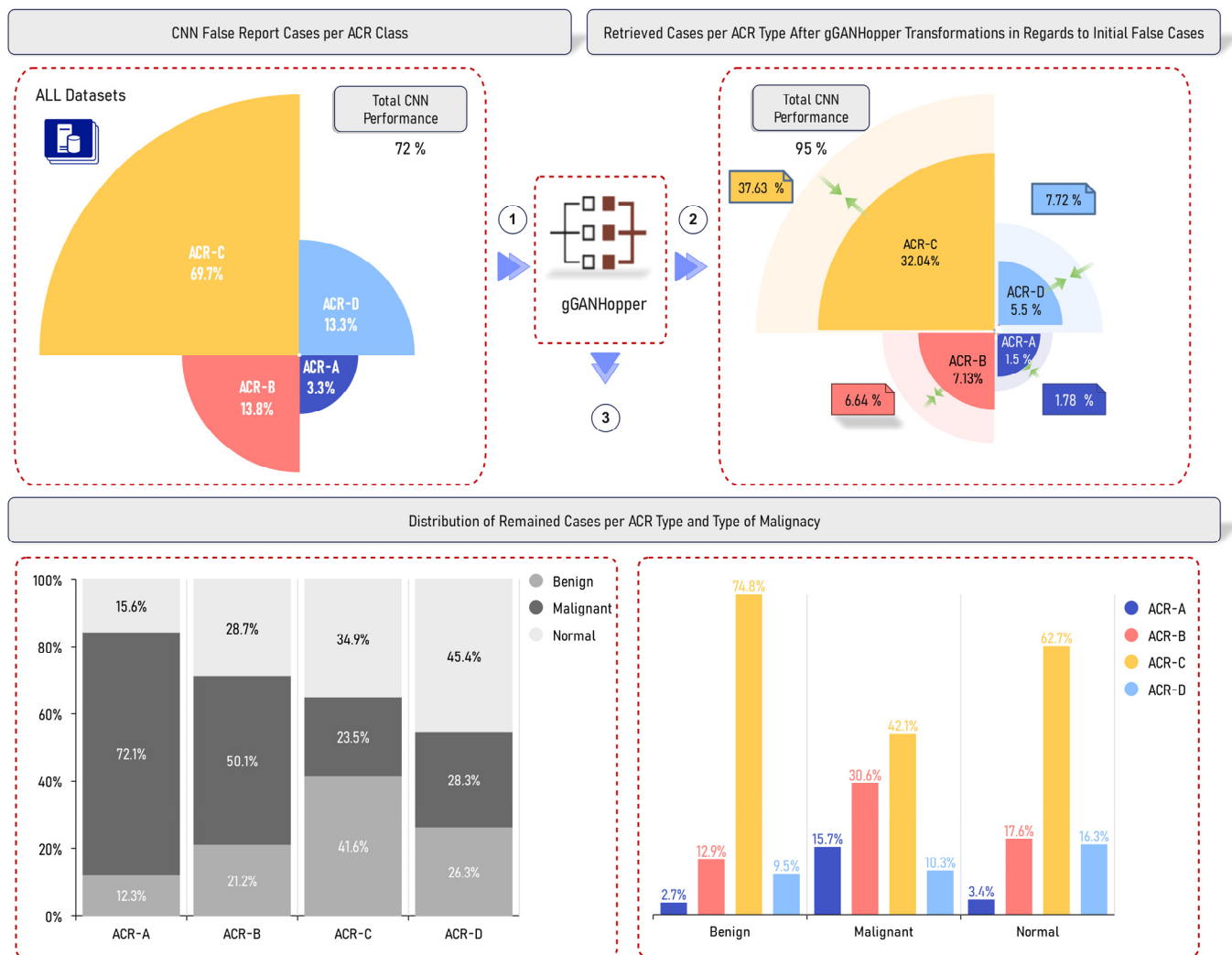


Figure 17. Impact of breast density transformations on classification performance. Top: CNN false report distribution by ACR class (left) and improvement after gGANHopper transformations (right). Bottom: Distribution of remaining misclassified cases by ACR type and malignancy class (left) and relative proportions across ACR categories (right).

In Step 2, the gGANHopper transformations were applied exclusively to the misclassified patches from Step 1. This approach improved the total classification accuracy to 95%. False responses were reduced by 37.63% for the ACR-C category, leaving only 32.04% of the initial ACR-C patches misclassified. These results highlight the value of gGANHopper-based breast density transformations in enhancing classification performance, as dense regions of interest (ROIs) often obscure critical features needed to distinguish between benign and malignant regions.

In Step 3, a detailed breakdown of the results from Step 2 is provided. The right lower bar graph illustrates the classification outcomes for the remaining patches across Benign, Malignant, and Normal categories. Specifically, the remaining misclassified ACR-C patches (32.04% from Step 2) were further decomposed into 41.6% Benign, 23.5% Malignant, and 34.9% Normal patches. Notably, the final misclassification errors were predominantly attributed to the ACR-B and ACR-C categories, with ACR-C showing a higher prevalence. In contrast, patches from the ACR-A category exhibited satisfactory classification performance, with minimal room for further improvement. However, patches from the ACR-D category were heavily impacted by masking effects, resulting in limited reversibility and significant degradation in classification performance.

The left lower bar graph provides additional insights, illustrating the redistribution of Normal ROI patches across ACR categories, with a reduction in ACR-D patches and an increase in ACR-A patches. Conversely, Benign and Malignant ROIs exhibit an opposing pattern across the ACR types. These findings highlight the complex interplay between breast density categories and classifier performance, demonstrating the impact of gGANHopper transformations on enhancing classification accuracy.

A depiction of the system response utilizing a full mammographic image can be seen in Figures 18 and 19. Figure 18 showcases six notable areas before and after the ACR category transformation to the ACR-A class. According to the database annotations, the findings are enclosed in the gray outline region. At first, the CNN classifier accurately identified four marked regions (Violet tick mark) as Benign/Malignant. The false responses of the CNN classifier (6 ROIs are erroneously recognized) are fed to the gGANHopper model to reverse the surrounding tissue's masking effect. Considering the ACR type of the ROI, the gGANHopper model regenerates the ROIs under examination to the ACR-A type. The CNN re-evaluates the transformed ROIs of this process, and the combined results are presented to the expert user. In the above example, 2 ROIs after the transformations to ACR-A are still classified wrongly by the CNN (Red circle with x).

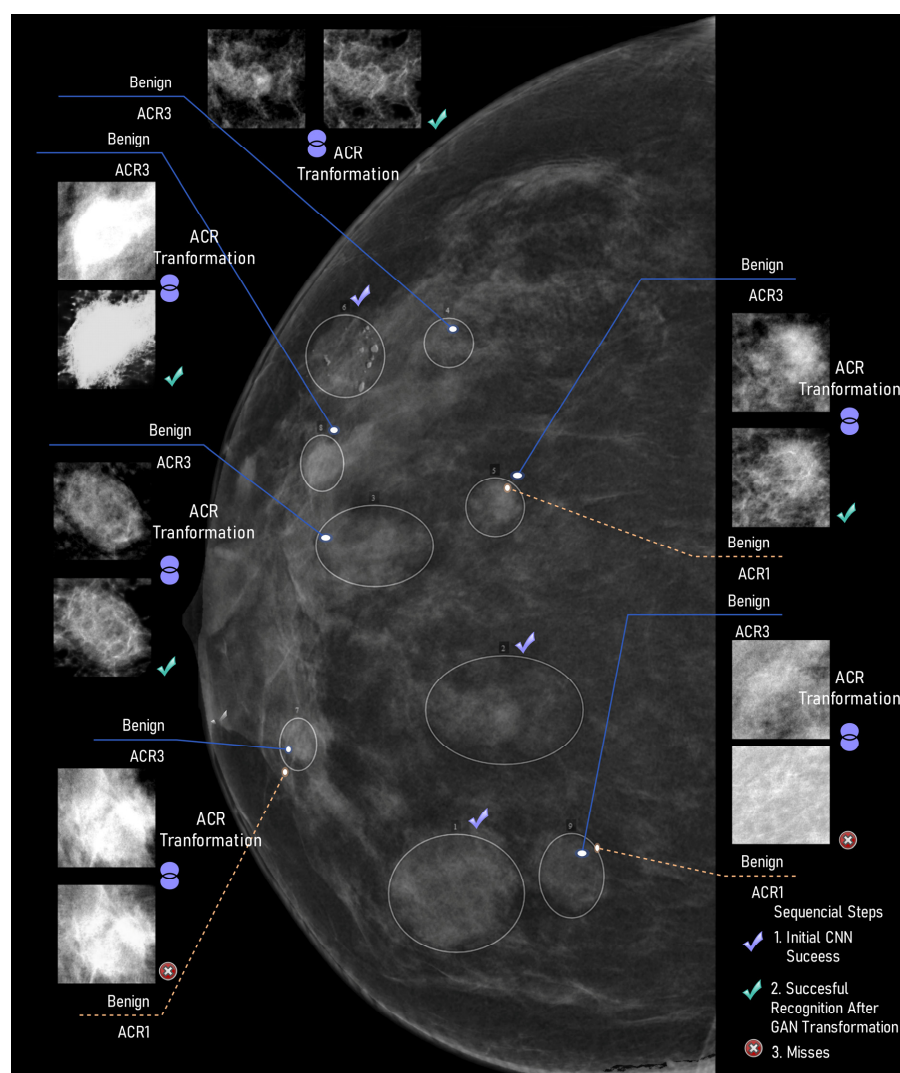


Figure 18. Visualization of system response before and after ROI-based breast density transformations. The CNN response is depicted for nine annotated ROIs. ACR transformations are performed for the faulty CNN verdict, marking the final CNN decision.

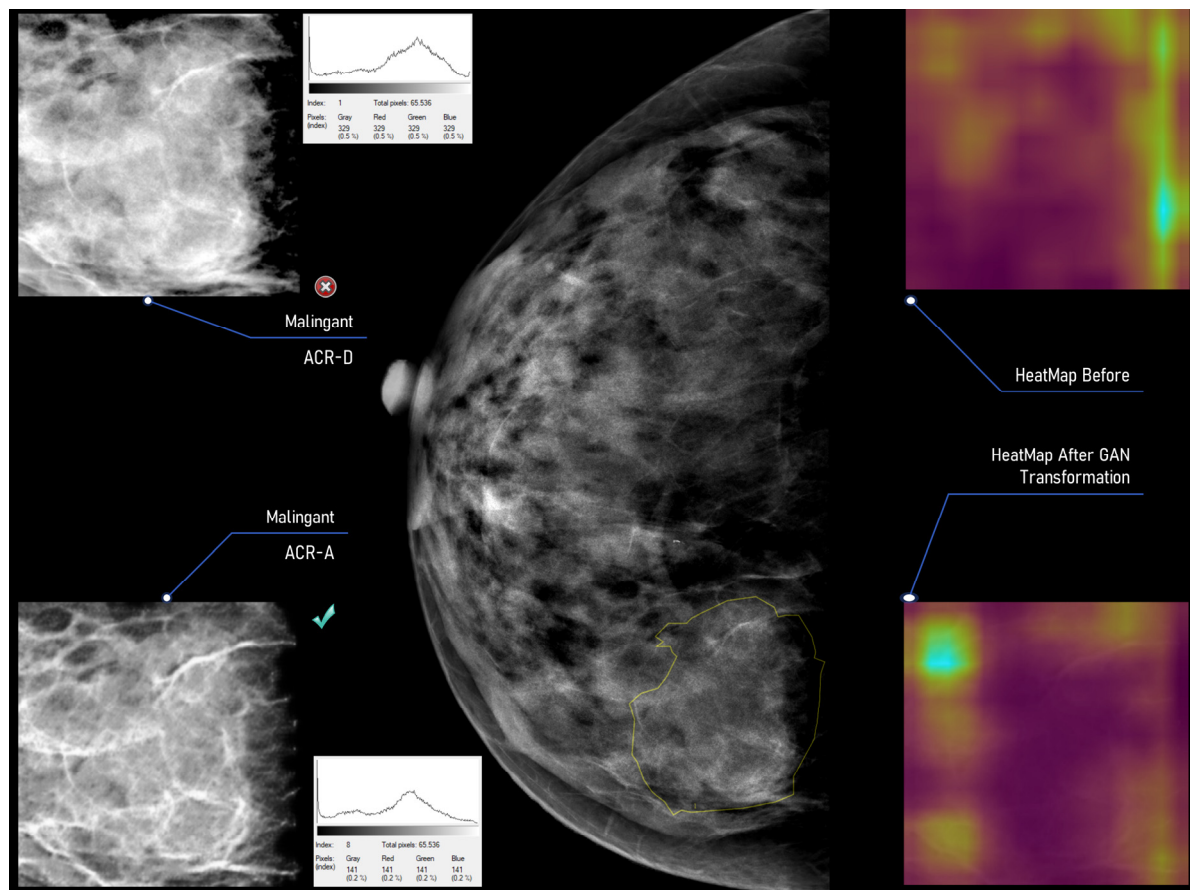


Figure 19. Annotated ROI with ACR-D characteristics, histogram, and heatmap for the ROI before and after ACR transformations.

We've included Figure 19, showcasing a histogram displaying intensity values for a marked ROI (Malignant ACR-D case) before and after its ACR transformation. Additionally, we've utilized Class Activation Mapping (as referenced in [48,49]) to highlight the specific areas within the input ROI that the CNN considers when making decisions. The resulting activation map is presented as a heatmap, where lighter regions indicate greater prominence in the predictive process than darker ones.

At this point, a necessary clarification is whether the above methodology can be applied in an image filter manner to enhance the classifier performance. The latter implies that the GAN-based breast tissue transformations will be used for all the extracted patches inside the breast tissue mask and not only for the faulty ones classified by CNN. Apart from the computational cost and the escalation in the time needed, using the class activation mapping visualization, one can comprehend the functionality and the number of changes the filter incurs. Because the number of extracted patches for all the datasets used is massive, we have randomly picked about 300 K patches from the total 12 M patches and fed the classifier. The outcome denotes that although applying the filter to the whole set of extracted patches inside the breast tissue mask may be slightly beneficial, the time and resources needed make it not applicable. Figure 20 depicts some representative cases from this experiment via the class map activation visualization.

The first column includes the input ROI patch (type, ACR class, and BIRADS class are given) before feeding the CNN classifier (but after the image processing steps); the second column displays the CAM, in other words, the regions from where the CNN classifier infers the final verdict (the brighter the areas the most significant they are). In contrast, the third and fourth columns correspond to the CAM of the same ROI patch after applying GAN-based breast transformations. From a simple overview, one can see the improvement of the

classifier discernment by observing the enlargement and enchantment of the prominent areas while maintaining the ROI image structure.

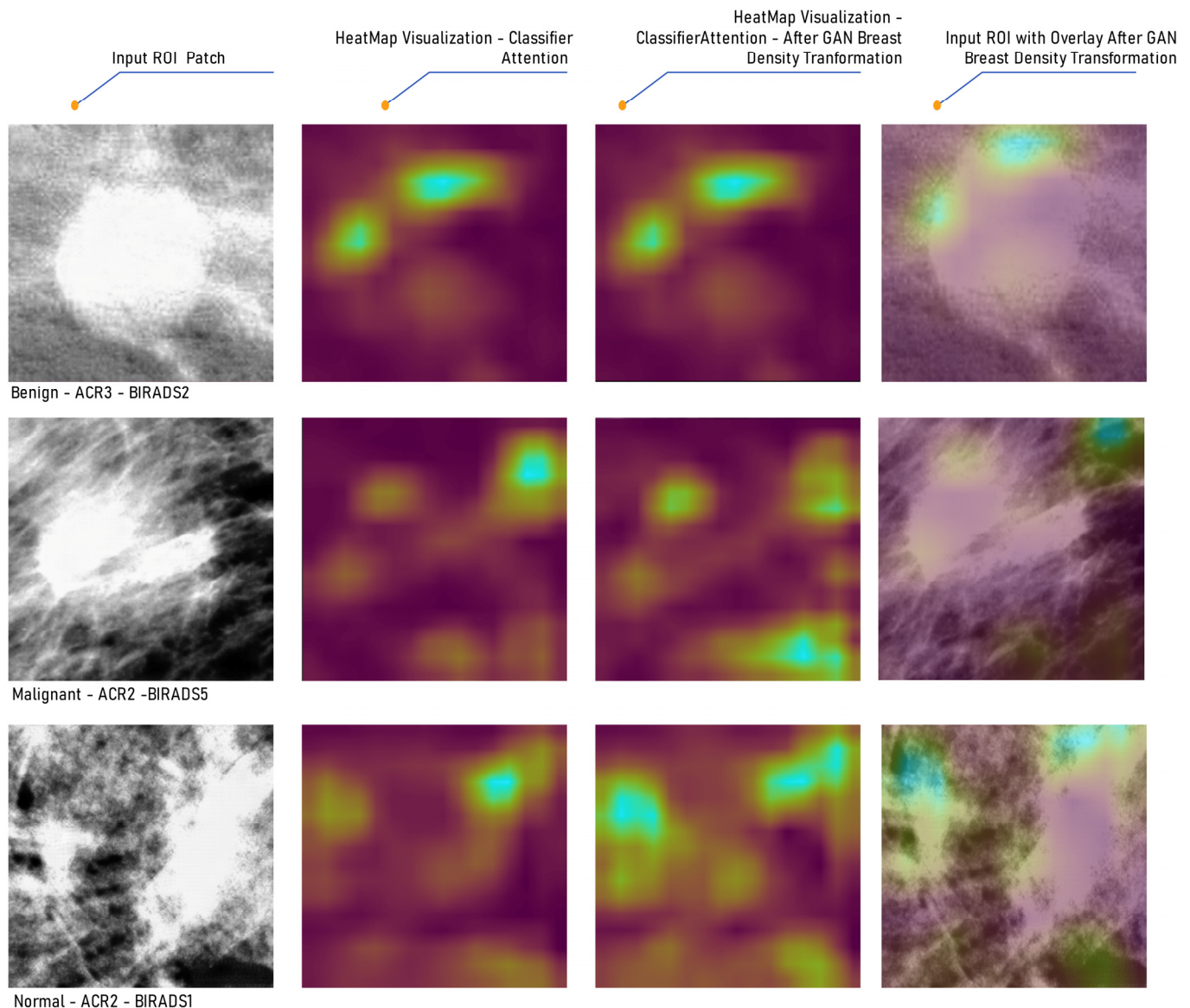


Figure 20. Visualization via heat maps of the application of ACR Transformations in all ROI patches rather than only to the faults of the classifier.

8. Discussion

Women with denser breast tissue are more likely to develop malignant regions than those with less dense tissue. Hence, breast tissue density is a known risk factor for cancer development. However, whether benign or malignant, breast abnormalities can often be disguised by glandular and connective tissue, making it difficult for computer-aided diagnosis tools and radiologists to detect them early in follow-up screening. Cancerous regions may be concealed by healthy tissue because connective tissue, glandular tissue, and cancers appear as white spots on mammography. Breast density evaluated by mammography is a significant independent risk factor for breast cancer. Dense breast tissue, which appears white or light grey on mammograms, contains more stroma, including collagen, and less fat [50,51]. Normal breast tissue includes 1–6% epithelial cells, contributing to density [52,53]. Some studies have found that the number of epithelial cells is higher in dense breasts, while [51] others have found no difference depending on breast density [53–55]. Our methodology considers the density of the analyzed region of interest (ROI) by making an effort to “uncover” and expose the tissue

effect's hidden features. In this study, we used two types of public screening mammography datasets, film and digital, to demonstrate the effectiveness of our proposed method.

The foundation of our approach is to transform the ACR density of ROI under examination to the equivalent of ACR-A to reveal underlying findings. We are more interested in ROIs of ACR density categories ACR-C, ACR-B, and ACR-D, which account for over 90% of all female breast cases. As expected, the density transformation of ROIs of type ACR-A back to ACR-A via the gGANHopper model (a cyclic transformation of an ROI that is, by nature, a foundation property of CycleGAN) contributes very little compared to the transformations of the other three breast density categories. We did not optimize the CNNs/GAN topologies regarding hyperparameters to focus only on reversing the ACR type of ROIs. CNN-based models with greater complexity, sophistication, and precision in the specific medical application can minimize the overall set to which the ACR transformations are applied, thus reducing the total time needed.

Furthermore, as we hypothesized, the gGANHopper model learned how to translate from high (ACR-D) to low (ACR-A) breast density, preserved the domain characteristics during translation, and gave an excellent total performance. During the evaluation phase, we examined the case of applying the method as a preprocessing step before feeding the CNN classifier, and although the outcome in terms of accuracy remained almost the same, the time needed was increased drastically. One must evaluate our method as a last resort to retrieve false responses of DL-based solutions. The method is advisable to be applied especially in dense tissue regions, which is considerably a smaller search space regarding the whole mammography, and the outcome is to be combined with that of the DL-based solution of detecting malignancies. Although more data are needed to thoroughly examine the extent of this reverse process, in all the datasets used, the overall percentage of successful recognition for Normal/Benign/Malignant ROIs improved significantly. Our approach was based on patches inside the breast density mask, which a UNet 3+ model provided. Still, we firmly believe that a similar approach with whole breast regions may work either by handling the area as a whole or via patch-segmented regions to several zoom scales.

In the case of unpaired images, quantitative evaluation of the transformed images is challenging since there are no standard or universal criteria for evaluating correctness [56]. Because of this, determining the quality of transformed images is a difficult and open subject whose metrics change based on the application's particular requirements. Most previously published work depends on application-specific metrics or human volunteers visually examining the modified images. Analyzing the transformed image visually remains the most popular and natural way to assess the quality of the transformation. In terms of IQA metrics associated with the quality of the images generated by the GAN model, one must consider that IQA metrics are divided into two categories: (i) subjective assessment, which is judged by human beings (radiologists in this application), and (ii) objective assessment, which is computed with mathematical algorithms. We stick with the second category since we are not replacing the original ROI and presenting the modified mammography with the generated ROI, but we feed the classifier with the GAN-based transformed ROI and reassess the specific area. The generated ROIs produced by the GAN model are 8-bit grayscale and 256×256 in size, and even the most trained human expert eye cannot confidently evaluate them as normal, benign, or malignant. The IQA metrics in medical images have two significant issues: they do not identify any abnormalities in the areas, and the image quality score is not a valid indicator of diagnostic accuracy. IQA metrics and diagnostic quality evaluation are different [57]. GANs are notorious for introducing artifacts in the output images, which is another reason IQA metrics must not be taken strictly. The fabricated textures introduced by GAN models undermine the integrity of the images and degrade the quality of the information they carry. These artifacts can be resolved with longer training epochs and more unpaired image samples. From the comparison of Tables 1 and 2, we can see that the GAN-based breast density transformations have a negligible impact (a deterioration of metrics is monitored since we are compulsory transforming the patches

from any ACR type to ACR-A) on IQA metrics, denoting that the changes performed do not alter in great extent the ROI patch and support our allegations, that the GAN transformations preserve the source domain characteristics. Also, domain disparities between the source and the target domain (here, the four ACR categories) significantly impact the perceptual realism of the synthetic patches.

Regarding performance, for film mammography datasets (MIAS, CBIS-DDSM), the accuracy improvement is up to 18%, which can be attributed to the GAN model's ACR reverse transformation process acting as an image enhancement step. However, the image quality of these datasets is inferior to that of fully digital ones (VinDR, SuReMaPP, INbreast, CDD-CESM), where the accuracy improvement is up to 29% at most, as these mammograms have better quality with no extreme intensity variations. When dealing with datasets acquired with equipment variations in hardware and time, the Reinhard Normalization and CLACHE techniques produce a common reference and make mammography sets acquired under different environments usable in mixed CNN-based solutions. Last, for all the datasets used (both film and digital), according to Figure 17, the increase in performance is 23% in total when the GAN-based breast density transformations are applied. The number of remaining patches of type ACR-A is below 2%, while the prevailing type of the remaining patches is ACR-C, with 32% for the categories Normal/Benign/Malignant. The CNN classifier was shown to be error-prone, especially for patches in categories ACR-C (69.7%), ACR-B (13.8%), and ACR-D (13.3%), since dense regions can mask key features needed to distinguish abnormalities.

Our study has a few limitations; one is the need for a perceptual study conducted by radiologists regarding the produced GAN-based patch images. This perceptual study will provide a post hoc explanation to comprehend the approach's learned mapping. Our method does not replace manual healthcare diagnosis but rather provides a tool that can be used by expert radiologists as a second (or, in some cases, third) reader to assist them in their diagnostic task. Although the advanced mathematics that underpins deep learning training algorithms is conceptually straightforward, their designs are more of a "black box." Another issue is the availability of healthy ACR-D type mammograms (or regions) for training the GAN models. Because few annotated digitized datasets are available to the public, a limited set of mammograms must be used. Finally, another difficulty stems from the problem's imbalanced nature; the ratio between the various ACR classes is far from equal, necessitating additional datasets.

9. Conclusions

Breast density can undoubtedly affect the accuracy of routine breast cancer detection methods, such as screening mammography. Early detection and classification of breast tumors are increasingly required to reduce the risk of death among cancer patients. Normal dense breast tissue appears white, and breast malignancies or tumors appear white, so dense tissue can conceal some tumors. In this study, we propose a method to reverse the masking effect of breast density levels and increase the performance of the DL classifier by projecting the misclassified ROIs in the breast density category ACR-A via GAN-based models. Consistent with our experimental findings, the gGANHopper-based approach successfully learned to transform ROI's breast density from high to low (and vice versa, although the first transformation is of more significant concern since we are interested in revealing details masked by tissue) while preserving all domain features necessary for accurate classification of cases to Normal/Benign/Malignant. In the future, we aim to apply the same concept to the whole breast region, considering the local type of Mammographic Breast Density (MBD) and the surrounding information from neighboring areas.

In summary, we developed a CycleGAN/gGANHopper-based approach to generate plausible mammogram ROI patches with ACR-A breast density appearance using a learned transformation between any of the other three ACR types. We showed that GAN-generated ROIs can help increase the classifier's performance by unveiling findings that otherwise would have been discarded. Our technique could help doctors and radiologists identify

worrisome ROIs and arrange treatments early, perhaps preventing complications and treatment challenges. Further evaluation of the clinical realism of the synthetic patches produced is needed.

Author Contributions: Conceptualization, D.A., G.A. and A.K.; methodology, A.K.; software, D.A., G.A. and I.C.; validation, G.A. and I.C.; formal analysis, D.A. and A.K.; investigation, D.A., G.A. and I.C.; resources, G.A. and I.C.; data curation, D.A. and G.A.; writing—original draft preparation, D.A. and A.K.; writing—review and editing, D.A. and A.K.; visualization, G.A. and I.C.; supervision, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: There are no ethical implications on the public dataset.

Informed Consent Statement: There are no ethical implications on the public dataset.

Data Availability Statement: All the datasets used are publicly available.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 73–76. [\[CrossRef\]](#)
4. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017. [\[CrossRef\]](#)
5. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017. [\[CrossRef\]](#)
6. Lira, W.; Merz, J.; Ritchie, D.; Cohen-Or, D.; Zhang, H. GANHopper: Multi-hop GAN for Unsupervised Image-to-Image Translation. In *Computer Vision—ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020*; Part XXVI; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer International Publishing: Cham, Switzerland, 2020. [\[CrossRef\]](#)
7. Gemici, A.A.; Bayram, E.; Hocaoglu, E.; Inci, E. Comparison of breast density assessments according to BI-RADS 4th and 5th editions and experience level. *Acta Radiol. Open* **2020**, *9*, 2058460120937381. [\[CrossRef\]](#)
8. Weigel, S.; Heindel, W.; Heidrich, J.; Hense, H.W.; Heidinger, O. Digital mammography screening: Sensitivity of the programme dependent on breast density. *Eur. Radiol.* **2017**, *27*, 2744–2751. [\[CrossRef\]](#)
9. Wanders, J.O.P.; Holland, K.; Veldhuis, W.B.; Mann, R.M.; Pijnappel, R.M.; Peeters, P.H.M.; van Gils, C.H.; Karssemeijer, N. Volumetric breast density affects performance of digital screening mammography. *Breast Cancer Res. Treat.* **2017**, *162*, 95–103. [\[CrossRef\]](#)
10. Sexauer, R.; Hejduk, P.; Borkowski, K.; Ruppert, C.; Weikert, T.; Dellas, S.; Schmidt, N. Diagnostic accuracy of automated ACR BI-RADS breast density classification using deep convolutional neural networks. *Eur. Radiol.* **2023**, *33*, 4589–4596. [\[CrossRef\]](#)
11. Mandelson, M.T.; Oestreicher, N.; Porter, P.L.; White, D.; Finder, C.A.; Taplin, S.H.; White, E. Breast density as a predictor of mammographic detection: Comparison of interval- and screen-detected cancers. *J. Natl. Cancer Inst.* **2000**, *92*, 1081–1087. [\[CrossRef\]](#)
12. Buist, D.S.M.; Porter, P.L.; Lehman, C.; Taplin, S.H.; White, E. Factors contributing to mammography failure in women aged 40–49 years. *J. Natl. Cancer Inst.* **2004**, *96*, 1432–1440. [\[CrossRef\]](#)
13. Destounis, S.; Johnston, L.; Highnam, R.; Arieno, A.; Morgan, R.; Chan, A. Using Volumetric Breast Density to Quantify the Potential Masking Risk of Mammographic Density. *AJR Am. J. Roentgenol.* **2017**, *208*, 222–227. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Anyfantis, D.; Koutras, A.; Apostolopoulos, G.; Christoyianni, I. Breast Density Transformations Using CycleGANs for Revealing Undetected Findings in Mammograms. *Signals* **2023**, *4*, 421–438. [\[CrossRef\]](#)
15. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [\[CrossRef\]](#)
16. Osuala, R.; Kushibar, K.; Garrucho, L.; Linardos, A.; Szafranowska, Z.; Klein, S.; Glocker, B.; Diaz, O.; Lekadir, K. Data synthesis and adversarial networks: A review and meta-analysis in cancer imaging. *Med. Image Anal.* **2023**, *84*, 102704. [\[CrossRef\]](#)
17. Guan, S. Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. *J. Med. Imaging* **2019**, *6*, 10. [\[CrossRef\]](#)
18. Wu, E.; Wu, K.; Lotter, W. Synthesizing lesions using contextual GANs improves breast cancer classification on mammograms. *arXiv* **2020**, arXiv:2006.00086.

19. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
20. Garrucho, L.; Kushibar, K.; Osuala, R.; Diaz, O.; Catanese, A.; del Riego, J.; Bobowicz, M.; Strand, F.; Igual, L.; Lekadir, K. High-resolution synthesis of high-density breast mammograms: Application to improved fairness in deep learning based mass detection. *Front. Oncol.* **2023**, *12*, 1044496. [[CrossRef](#)]
21. Lee, J.; Mustafaev, T.; Nishikawa, R.M. Impact of GAN artifacts for simulating mammograms on identifying mammographically occult cancer. *J. Med. Imaging* **2023**, *10*, 1044496. [[CrossRef](#)]
22. Yamazaki, A.; Ishida, T. Two-View Mammogram Synthesis from Single-View Data Using Generative Adversarial Networks. *Appl. Sci.* **2022**, *12*, 12206. [[CrossRef](#)]
23. Rautela, K.; Kumar, D.; Chahar, V. Improved GAN for image resolution enhancement using ViT for breast cancer detection. *Int. J. Imaging Syst. Technol.* **2023**, *34*, e22998. [[CrossRef](#)]
24. Park, S.; Lee, K.H.; Ko, B.; Kim, N. Unsupervised anomaly detection with generative adversarial networks in mammography. *Sci. Rep.* **2023**, *13*, 2925. [[CrossRef](#)] [[PubMed](#)]
25. Oyelade, O.N.; Ezugwu, A.E.; Almutairi, M.S.; Saha, A.K.; Abualigah, L.; Chiroma, H. A generative adversarial network for synthetization of regions of interest based on digital mammograms. *Sci. Rep.* **2022**, *12*, 6166. [[CrossRef](#)] [[PubMed](#)]
26. D'Orsi, C.J. *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*; American College of Radiology: Reston, VA, USA, 2014.
27. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Computer Vision—ECCV 2016, Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016*; Springer: Berlin/Heidelberg, Germany, 2016.
28. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6636–6637.
29. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved Techniques for Training GANs. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2234–2242.
30. Suckling, J.; Parker, J.; Dance, D.R.; Astley, S.; Hutt, I.W.; Boggis, C.R.M.; Ricketts, I.W.; Stamatakis, E.A.; Cerneaz, N.; Kok, S.; et al. Mammographic Image Analysis Society (MIAS) Database v1.21. 2015. Available online: <https://api.semanticscholar.org/CorpusID:62937784> (accessed on 1 September 2024).
31. Lee, R.S.; Gimenez, F.; Hoogi, A.; Miyake, K.K.; Gorovoy, M.; Rubin, D.L. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci. Data* **2017**, *4*, 170177. [[CrossRef](#)]
32. Lee, R.S.; Gimenez, F.; Hoogi, A.; Rubin, D. Curated breast imaging subset of Digital Database for Screening Mammography (CBIS-DDSM) [Data set]. *Cancer Imaging Arch.* **2016**. [[CrossRef](#)]
33. Moreira, I.C.; Amaral, I.; Domingues, I.; Cardoso, A.; Cardoso, M.J.; Cardoso, J.S. INbreast. *Acad. Radiol.* **2012**, *19*, 236–248. [[CrossRef](#)]
34. Bruno, A.; Ardizzone, E.; Vitabile, S.; Midiri, M. A novel solution based on scale invariant feature transform descriptors and deep learning for the detection of suspicious regions in mammogram images. *J. Med. Signals Sens.* **2020**, *10*, 158. [[CrossRef](#)]
35. Khaled, R.; Helal, M.; Alfarghaly, O.; Mokhtar, O.; Elkorany, A.; El Kassas, H.; Fahmy, A. Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. *Sci. Data* **2022**, *9*, 122. [[CrossRef](#)]
36. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [[CrossRef](#)]
37. Nguyen, H.T.; Nguyen, H.Q.; Pham, H.H.; Lam, K.; Le, L.T.; Dao, M.; Vu, V. VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Sci. Data* **2023**, *10*, 277. [[CrossRef](#)] [[PubMed](#)]
38. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Wu, J. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In *Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 4–8 May 2020.
39. Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. *IEEE Comput. Graph Appl.* **2001**, *21*, 34–41. [[CrossRef](#)]
40. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; Romeny, B.T.H.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process* **1987**, *39*, 355–368. [[CrossRef](#)]
41. Qiao, T.; Ren, J.; Wang, Z.; Zabalza, J.; Sun, M.; Zhao, H.; Li, S.; Benediktsson, J.A.; Dai, Q.; Marshall, S. Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 119–133. [[CrossRef](#)]
42. Ruderman, D.L.; Cronin, T.W.; Chiao, C.-C. Statistics of cone responses to natural images: Implications for visual coding. *J. Opt. Soc. Am. A* **1998**, *15*, 2036–2045. [[CrossRef](#)]
43. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October*; Springer: Cham, Switzerland, 2015.
44. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 5–9. [[CrossRef](#)]

45. Egiazarian, K.; Astola, J.; Ponomarenko, N.; Lukin, V.; Battisti, F.; Carli, M. Two New Full-Reference Quality Metrics based on HVS. In Proceedings of the Second International Workshop on Video Processing and Quality Metrics, Scottsdale, AZ, USA, 22–24 January 2006.
46. Ponomarenko, N.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Carli, M. Modified image visual quality metrics for contrast change and mean shift accounting. In Proceedings of the 2011 11th International Conference—The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2011, Polyana-Svalyava, Ukraine, 23–25 February 2011.
47. Ponomarenko, N.; Silvestri, F.; Egiazarian, K.; Carli, M.; Astola, J.; Lukin, V. On Between-coefficient Contrast Masking of DCT Basis Functions. In Proceedings of the Third International Workshop on Video Processing and Quality, Scottsdale, AZ, USA, 25–26 January 2007.
48. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *arXiv* **2016**, arXiv:1610.02391. [[CrossRef](#)]
49. Chattopadhyay, A.; Sarkar, A.; Howlader, P.; Balasubramanian, V.N. Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018. [[CrossRef](#)]
50. Ng, K.; Lau, S. Vision 20/20: Mammographic breast density and its clinical applications. *Med. Phys.* **2015**, *42*, 7059–7077. [[CrossRef](#)]
51. McCormack, V.A.; Silva, I.D.S. Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis. *Cancer Epidemiol. Biomark. Prev.* **2006**, *15*, 1159–1169. [[CrossRef](#)]
52. Bertrand, K.A.; Scott, C.G.; Tamimi, R.M.; Jensen, M.R.; Pankratz, V.S.; Norman, A.D.; Visscher, D.W.; Couch, F.J.; Shepherd, J.; Chen, Y.-Y.; et al. Dense and Nondense Mammographic Area and Risk of Breast Cancer by Age and Tumor Characteristics. *Cancer Epidemiol. Biomark. Prev.* **2015**, *24*, 798–809. [[CrossRef](#)]
53. Bertrand, K.A.; Tamimi, R.M.; Scott, C.G.; Jensen, M.R.; Pankratz, V.S.; Visscher, D.; Norman, A.; Couch, F.; Shepherd, J.; Fan, B.; et al. Mammographic density and risk of breast cancer by age and tumor characteristics. *Breast Cancer Res.* **2013**, *15*, R104. [[CrossRef](#)] [[PubMed](#)]
54. Yaghjian, L.; Colditz, G.A.; Collins, L.C.; Schnitt, S.J.; Rosner, B.; Vachon, C.; Tamimi, R.M. Mammographic Breast Density and Subsequent Risk of Breast Cancer in Postmenopausal Women According to Tumor Characteristics. *JNCI J. Natl. Cancer Inst.* **2011**, *103*, 1179–1189. [[CrossRef](#)] [[PubMed](#)]
55. Boyd, N.F.; Huszti, E.; Melnichouk, O.; Martin, L.J.; Hislop, G.; Chiarelli, A.; Yaffe, M.J.; Minkin, S. Mammographic features associated with interval breast cancers in screening programs. *Breast Cancer Res.* **2014**, *16*, 417. [[CrossRef](#)]
56. Borji, A. Pros and Cons of GAN Evaluation Measures. *Comput. Vis. Image Underst.* **2018**, *179*, 41–65. [[CrossRef](#)]
57. Kumar, R.; Rattan, M. Analysis of Various Quality Metrics for Medical Image Processing. *Int. J. Adv. Res. Comput. Sci.* **2012**, *2*, 137–144.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.