

## A deep learning approach for virtual contrast enhancement in Contrast Enhanced Spectral Mammography

Aurora Rofena <sup>a</sup>,<sup>1</sup> Valerio Guerrasi <sup>a</sup>,<sup>1</sup> Marina Sarli <sup>b</sup>, Claudia Lucia Piccolo <sup>b</sup>, Matteo Sammarra <sup>b</sup>, Bruno Beomonte Zobel <sup>b,c</sup>, Paolo Soda <sup>a,d</sup>,<sup>\*</sup>

<sup>a</sup> Unit of Computer Systems & Bioinformatics, Department of Engineering University Campus Bio-Medico, Rome, Italy

<sup>b</sup> Department of Radiology, Fondazione Policlinico Campus Bio-Medico, Rome, Italy

<sup>c</sup> Department of Radiology, University Campus Bio-Medico, Rome, Italy

<sup>d</sup> Department of Radiation Sciences, Radiation Physics, Biomedical Engineering, Umeå University, Sweden



### ARTICLE INFO

#### Keywords:

Virtual contrast enhancement  
Contrast enhanced spectral mammography  
CESM  
Image-to-image translation  
Generative adversarial network

### ABSTRACT

Contrast Enhanced Spectral Mammography (CESM) is a dual-energy mammographic imaging technique that first requires intravenously administering an iodinated contrast medium. Then, it collects both a low-energy image, comparable to standard mammography, and a high-energy image. The two scans are combined to get a recombined image showing contrast enhancement. Despite CESM diagnostic advantages for breast cancer diagnosis, the use of contrast medium can cause side effects, and CESM also beams patients with a higher radiation dose compared to standard mammography. To address these limitations, this work proposes using deep generative models for virtual contrast enhancement on CESM, aiming to make CESM contrast-free and reduce the radiation dose. Our deep networks, consisting of an autoencoder and two Generative Adversarial Networks, the Pix2Pix, and the CycleGAN, generate synthetic recombined images solely from low-energy images. We perform an extensive quantitative and qualitative analysis of the model's performance, also exploiting radiologists' assessments, on a novel CESM dataset that includes 1138 images. As a further contribution to this work, we make the dataset publicly available. The results show that CycleGAN is the most promising deep network to generate synthetic recombined images, highlighting the potential of artificial intelligence techniques for virtual contrast enhancement in this field.

### 1. Introduction

Contrast-Enhanced Spectral Mammography (CESM) is a dual-energy breast imaging technique classified as a level II breast diagnostic. Unlike standard mammography, also referred to as full-field digital mammography (FFDM), CESM involves the injection of an iodinated contrast medium that diffuses into the tumor tissue, enhancing lesion visibility (Jochelson and Lobbes, 2021). This results in improved diagnostic accuracy, especially in patients with heterogeneously dense and extremely dense breast tissue, which in FFDM may obscure small masses and reduce sensitivity (James and Tennant, 2018). Furthermore, CESM is a valid alternative to contrast-enhanced breast magnetic resonance imaging (MRI), being more specific, accessible, cheaper, and preferred by patients (James and Tennant, 2018). During CESM, both breasts undergo compression in craniocaudal and mediolateral oblique projections, capturing low-energy (LE) and high-energy (HE) images in rapid succession post-contrast injection. The LE image is

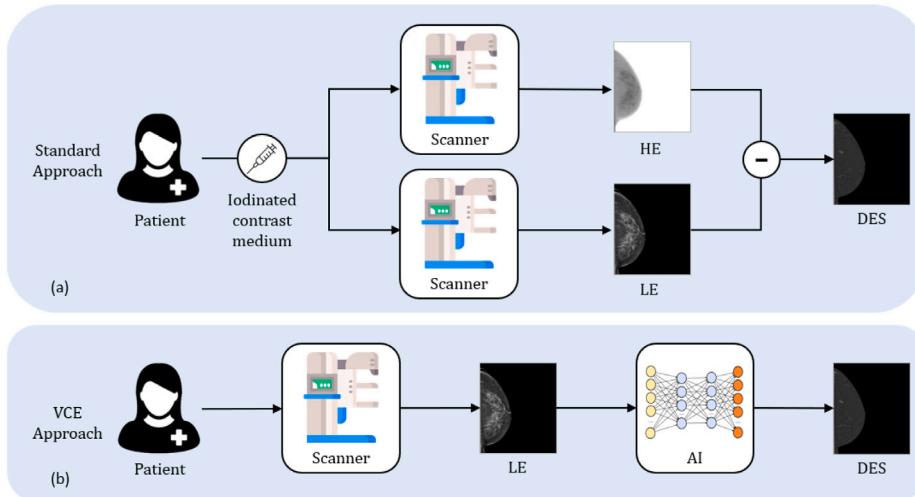
comparable with FFDM and is combined with the HE image through post-processing to generate a dual-energy subtracted (DES) image in which areas of lesion contrast enhancement are highlighted (Jochelson and Lobbes, 2021). However, despite its benefits, the widespread adoption of CESM may be limited by two issues: (I) the administration of iodinated contrast medium carries risks of adverse reactions, including hypersensitivity reactions and contrast-induced nephropathy (CIN); (II) CESM entails a higher radiation dose compared to FFDM due to the dual-energy exposure for each projection (Patel et al., 2018).

A promising solution to address these limitations is to leverage Artificial Intelligence (AI) techniques, particularly Deep Learning (DL). Recent advancements in DL have facilitated Image-to-Image translation tasks in medical imaging, including Virtual Contrast Enhancement (VCE), which involves generating synthetic contrast-enhanced images from non-contrast images. While previous studies have explored VCE

\* Corresponding author.

E-mail address: [p.soda@unicampus.it](mailto:p.soda@unicampus.it) (P. Soda).

<sup>1</sup> These authors contributed equally to this work.



**Fig. 1.** Illustration of DES images generation. (a) Standard approach with DES image obtained by injecting iodinated contrast medium into the patient and acquiring both the LE and HE images. (b) VCE approach proposed by our work, with DES images generated directly from the LE images, without the need to inject iodinated contrast medium or acquire HE images.

for MRI (Kleesiek et al., 2019) and computed tomography (CT) images (Choi et al., 2021; Xie et al., 2021), the application of this approach to CESM remains unexplored. The use of DL techniques to perform VCE on CESM would involve investigating the synthesis of DES images directly from LE images. Given the equivalence between LE images and FFDM images, as supported in Francescone et al. (2014), this approach would offer three significant benefits. First, it may eliminate the necessity of administering contrast medium, that may have adverse reaction. Second, it may reduce patient radiation exposure since the HE image would not need to be acquired. Third, employing AI models to generate DES images from LE images may significantly reduce waiting times, costs, and stress when a patient has already undergone FFDM, but CESM is further needed for the diagnosis, treatment, or intervention.

Thus, as depicted in Fig. 1, our goal is to investigate if AI can be used as a possible tool for VCE on CESM, by training and validating established DL models. As part of our efforts to advance research in the CESM field, we release the dataset collected and used in this study. This marks a significant milestone, as it is the first publicly available CESM dataset in DICOM format, with data anonymized to ensure privacy. The dataset consists of 1138 images in DICOM format, comprising 569 LE and 569 DES images, collected from 105 patients. All relevant non-sensitive DICOM tags are accessible, and the dataset includes information extracted from medical reports and biopsy records. Furthermore, to ensure the reproducibility of our experiments in support of research, we have decided to make our code public.<sup>2</sup>

To summarize, our contributions are:

- Present an evaluation of three state-of-the-art DL models in executing the novel task of VCE for CESM images. Through quantitative and qualitative assessments, we aim to define the baseline performance for this specific task;
- Release of the first dataset of CESM examination images in DICOM format, complete with information from medical reports and biopsies;
- Release of code used in our experiments.

The rest of the manuscript is organized as follows: Section 2 presents the background of the study. Section 3 outlines the publicly released

dataset used for the study. Section 4 proposes the methods used, including neural network architectures, image pre-processing, and the training and evaluation process. Section 5 presents and discusses the results obtained, while Section 6 offers our conclusions.

## 2. Background

In this Section, we describe the CESM technique, highlighting its growing relevance in clinical applications, while also addressing the limitations that could impede its widespread adoption in the clinical setting. Additionally, we delve into the current advancements in image-to-image translation for VCE across diverse medical imaging modalities, pivotal to our investigation into its application to CESM images in our study.

### 2.1. CESM technique

The CESM technique involves compressing both breasts in cranio-caudal and mediolateral oblique projections for approximately two minutes after contrast medium injection. During each compression, a LE and HE image are acquired in quick succession, and the exam is completed within 10 min of the start of contrast administration. The LE image is acquired below the K-edge of iodine (33.2 keV). Thus, as supported in Francescone et al. (2014), it is blind to iodine and comparable to FFDM, despite the presence of iodinated contrast medium in the breast. Conversely, the HE image is acquired above the K-edge of iodine, and it is not suitable for diagnostic purposes on its own. However, employing logarithmic subtraction in post-processing, it is combined with the LE image to obtain the recombined image or DES image. This DES image suppresses parenchymal tissue, revealing only areas of contrast enhancement, thereby enhancing lesion visibility and improving diagnostic accuracy. Thus, CESM provides radiologists with LE and DES image pairs for diagnosis (Jochelson and Lobbes, 2021).

### 2.2. CESM clinical applications

Because of its characteristics, CESM can be used in several clinical scenarios. It serves as a valuable tool in high-risk screening, further evaluation of extremely dense breast tissue, diagnostic assessment of suspicious lesions, breast cancer staging, surgical planning, and assessment of treatment response (Bhimani et al., 2017). In particular,

<sup>2</sup> [https://github.com/cosbidev/VCE\\_CESM](https://github.com/cosbidev/VCE_CESM)

CESM is a successful digital breast imaging method for planning surgery since it provides a clear map of the lesion extent and its relation to other structures. In this context, it has been demonstrated that CESM led to a significant improvement in detecting multifocal-multicentric cancers compared to FFDM. This capability is essential to determine the methods of surgical treatment for the patients, leading CESM to have a crucial influence on surgical decision (Lorek et al., 2021). CESM also plays a crucial role in evaluating the response to neoadjuvant chemotherapy, which is the standard treatment for locally advanced breast carcinomas. Accurate assessment of this response is fundamental to increase the chances of successful breast-conserving surgery and to avoid local recurrence (Elkassas et al., 2022). It has been demonstrated that CESM is as reliable as MRI in assessing the response to neoadjuvant chemotherapy and may be an alternative when MRI is contraindicated or its availability is limited (Iotti et al., 2017). Furthermore, CESM-guided biopsy has emerged as a promising technique, offering a viable alternative to MRI-guided breast biopsy. Recent studies have showcased its efficacy, demonstrating that CESM-guided biopsy is well tolerated by patients and feasible due to its fast execution and high success rates (Sammarra et al., 2024; James, 2022).

### 2.3. CESM limitations

Two factors may limit the adoption of CESM as a widespread screening technique. First, although the administration of iodinated contrast medium is essential for the performance of CESM, adverse effects may occur, such as hypersensitivity reactions and the possibility of CIN (Pasternak and Williamson, 2012). The estimated incidence of hypersensitivity reactions ranges from 0.7% to 3.1%, with severe reactions occurring in only 0.02% to 0.04% of administrations (Patel et al., 2018). CIN is defined as a decline in renal function within three days of iodine-based contrast medium administration. Its pathophysiology remains poorly understood, but pre-existing renal dysfunction appears to contribute (Zamora et al., 2021). Second, CESM requires a double energy exposure for each projection, i.e., it collects both the LE and HE images, resulting in a higher radiation dose compared to FFDM (Patel et al., 2018). To overcome these limitations, reducing the administration of the contrast medium and the amount of radiation using AI techniques is a viable solution that can now be investigated, thanks to the recent progress in DL and generative approaches, in particular.

### 2.4. Image-to-image translation for virtual contrast enhancement

In recent years, the use of deep neural networks has become dominant in image-to-image translation for medical images (Kaji and Kida, 2019). Image-to-image translation involves converting images from the source domain  $X$  to the target domain  $Y$ . This approach is implemented in medical image analysis, including tasks such as noise reduction (Yang et al., 2017), super-resolution (Umebara et al., 2018), cross-modality synthesis (Liu et al., 2021a), and VCE, which is the goal of our work. Previous studies have explored the application of deep neural networks in VCE for MRI and CT images. The aim was to generate synthetic contrast-enhanced images from non-contrast images, with the goal of evaluating the feasibility of reducing contrast media usage while preserving image quality. In this context, Kleesiek et al. (2019) trained a fully convolutional autoencoder with dropout to predict contrast enhancement from non-contrast multiparametric brain MRI scans, with 10 channels used for the model input. Choi et al. (2021) evaluated the 3D implementation of Pix2Pix for generating synthetic contrast-enhanced chest CT from non-contrast chest CT. Xie et al. (2021) exploited a cycle-consistent generative adversarial network (CycleGAN) framework to learn a mapping from non-contrast CT to contrast-enhanced CT, using a residual U-Net as the generator of the CycleGAN.

However, no one has yet analyzed VCE within the context of CESM. Nonetheless, we firmly believe that investigating this approach is a valid solution to overcome the use of contrast media and reduce exposure to dual energy.

## 3. Materials

In this work, we collected and utilized an in-house dataset, named CESM@UCBM in the following, which we make publicly available<sup>3</sup> after anonymizing sensitive data. The images are stored in DICOM format, thus including imaging parameters, acquisition details, and annotations, which are useful for the analyses.

**Table 3.1** summarizes the main characteristics of the dataset. It consists of 1138 images, divided into 569 LE images and 569 DES images. These images were obtained from 105 patients aged between 31 and 90 years, with an average age of 57.4 years and a standard deviation of 11.5 years. They were enrolled in the study under the Ethical Committee approval PAR 51.23 OSS. All the exams were performed at the Fondazione Policlinico Universitario Campus Bio-Medico in Rome between September 2021 and October 2022, using the Senograph Pristina digital mammography (GE Healthcare). The size of 984 images is 2850 × 2396 pixels, while the remaining 154 images have a size of 2294 × 1916 pixels. Among the 1138 images in the CESM@UCBM dataset:

- 284 images show the craniocaudal projection of right breasts, with 142 LE images and 142 DES images;
- 282 images show the craniocaudal projection of left breasts, with 141 LE images and 141 DES images;
- 292 images show the mediolateral oblique projection of right breasts, with 146 LE images and 146 DES images;
- 280 images show the mediolateral oblique projection of left breasts, with 140 LE and 140 DES images.

Late-phase acquisitions are available for 86 patients, implying that the image of the same breast in the same projection was acquired multiple times within a few minutes. Unilateral acquisitions were performed for 7 patients.

Furthermore, for each patient, we extract information from the medical report and the biopsy, if available. This includes the classification of breast composition with one of the 4 categories ( $a, b, c, d$ ) defined by the fifth edition of the Breast Imaging Reporting and Data System (BI-RADS) released by the American College of Radiology (ACR) (D'orsi et al., 2003). The category  $a$  indicates almost entirely fatty breasts and characterizes 130 images, of which 35% show contrast enhancement. The category  $b$  identifies breasts with scattered areas of fibroglandular density and characterizes 360 images, of which 33% show contrast enhancement. The category  $c$  indicates heterogeneously dense breast tissue and characterizes 414 images, of which 39% show contrast enhancement. Finally, the category  $d$  identifies extremely dense breasts and characterizes 190 images, of which 36% show contrast enhancement. Instead, for 44 images, the ACR category is not reported. **Fig. 2** compares 4 pairs of LE and DES images belonging to the CESM@UCBM dataset, which differ in the ACR category.

In addition, the reports for 31 patients are completed with the BI-RADS descriptors (from 1 to 6) associated with the probability of malignancy of the lesion and the actions to be taken to conclude the diagnostic-therapeutic management process. Based on biopsies, malignant lesions were identified in 258 images, 129 of which were LE and 129 of which were DES. Benign lesions were identified in 104 images, 52 LE and 52 DES. Borderline lesions were identified in 16 images, 8 LE and 8 DES. The remaining images did not show any tumor-related abnormalities.

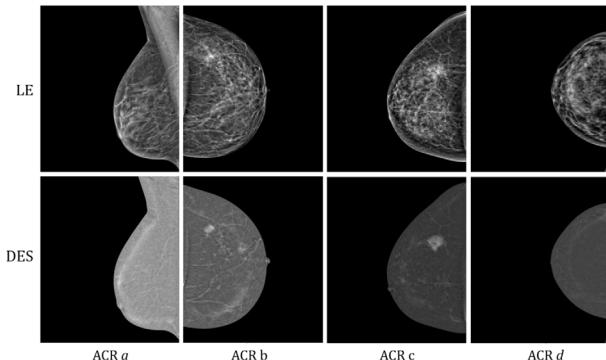
## 4. Methods

We design the VCE task on CESM images as an image-to-image translation problem, where the source domain  $X$  consists of LE images, and the target domain  $Y$  consists of DES images. Considering  $x$  as a generic LE image from  $X$ , and  $y$  as the corresponding DES image

<sup>3</sup> <http://www.cosbi-lab.it/cesmucbm/>

**Table 3.1**  
Characteristics of the CESM@UCBM dataset.

General Info	Period	September 2021–October 2022
	Source	Fondazione Policlinico UCBM
	Patients	105
	Total images	1138
Image type	LE	569 images
	DES	569 images
Image view	Craniocaudal	566 images
	mediolateral oblique	572 images
Image size	2850 × 2396	984 images
	2294 × 1916	154 images
Patients' age	<50 years	31 patients
	50–59 years	37 patients
	60–69 years	19 patients
	≥70 years	18 patients
Biopsy	Malignant	258 images
	Benign	104 images
	Borderline	16 images
ACR category	a	130 images
	b	360 images
	c	414 images
	d	190 images
	Not reported	44 images
BI-RADS descriptors	1–6	31 patients



**Fig. 2.** From left to right, pairs of LE (above) and DES images (below) from the CESM@UCBM dataset with ACR categories a, b, c, d.

from  $Y$ , we implement DL models  $G$  that take  $x$  as input and attempt to generate the corresponding virtual contrast-enhanced DES images, i.e.,  $\hat{y} = G(x)$ , minimizing the distance between  $\hat{y}$  and  $y$ .

A schematic representation of the methods is shown in Fig. 3 and, accordingly, in this Section, we present the image pre-processing, the AI models, and the training and evaluation process.

#### 4.1. Pre-processing

To ensure that data are consistent and homogeneous, we first applied zero-padding to make images squared, second we applied contrast stretching to adjust the brightness of the images, third we normalized the pixel values bringing them in the range  $[0, 1]$ , and fourth we resized the images to  $256 \times 256$ .

#### 4.2. AI models

The AI models used in this work to perform the VCE task on CESM images are an autoencoder and two GANs, the Pix2Pix (Isola et al., 2017) and the CycleGAN (Zhu et al., 2017), since these networks have accelerated the use of deep neural networks in image-to-image translation (Kaji and Kida, 2019) and are consistent with the state of the art of VCE, as reported in Section 2.4. Representative diagrams of the architectures used are in Fig. 4.

#### 4.2.1. Autoencoder

The Autoencoder architecture  $G$  consists of four convolutional encoder blocks and four convolutional decoder blocks, with bypass connection and residual connection. Each block is equipped with three convolutional layers with  $3 \times 3$  kernels, each followed by batch normalization and ReLU activation function. Only the last decoder block is an exception, as it consists of a convolutional layer with a  $1 \times 1$  kernel, followed by a sigmoid activation function. In the first half of the network, each block is followed by a max pooling layer, while an upsampling layer is used in the later half of the network. The total amount of parameters for the Autoencoder is  $1.6 \cdot 10^6$ . The encoder part takes as input an image  $x$  and compresses it into a lower-dimension internal representation. The decoder part decodes the compact representation provided by the encoder and generates the output image  $\hat{y}$ . The loss function minimizes the difference between the generated image  $\hat{y} = G(x)$  and the corresponding target image  $y$ , enabling the optimization of the network parameters as follows:

$$G^* = \arg \min_G \mathbb{E}_{x,y}[\|y - G(x)\|] \quad (1)$$

where  $G^*$  is the optimal model that minimizes the pixel-wise difference denoted as  $\|\cdot\|$ .

#### 4.2.2. Pix2Pix

The Pix2Pix (Isola et al., 2017) is a conditional GAN specifically designed for image-to-image translation with paired datasets. It comprises a generator network  $G$  and a discriminator network  $D$ , as in a conventional GAN. The generator  $G$  learns the mapping from the  $X$  domain to the  $Y$  domain. The  $L1$  loss  $\mathcal{L}_{L1}(G)$  is used to minimize the difference between the generated output  $G(x) = \hat{y}$  and the real image  $y$ , such that  $G(x) = \hat{y} \approx y$ . It is formulated as:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1] \quad (2)$$

On the other hand, the discriminator  $D$  tries to distinguish between the real image  $y$  and the generated output  $\hat{y}$ . The generator and discriminator are thus trained in a competitive way using an adversarial loss  $\mathcal{L}_{cGAN}(G, D)$ , given by:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log(D(x, y))] + \mathbb{E}_x[\log(1 - D(x, G(x)))] \quad (3)$$

The generator tries to minimize this function, learning to fool the discriminator, and the discriminator tries to maximize this function, learning to distinguish real images from synthetic images. The Pix2Pix full objective is formulated as

$$G^* = \arg \min_G \max_D [\mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)] \quad (4)$$

where  $\lambda$  balances the relative contribution of the two loss terms. As presented in Isola et al. (2017), both the generator and discriminator network are constructed with convolutional blocks, which comprise the convolutional layer, batch normalization, and ReLU activation function. The generator architecture is based on the U-Net autoencoder (Isola et al., 2017), with skip connections concatenating encoder block  $i$  and decoder block  $n - i$ , where  $n$  is the total number of blocks. The discriminator architecture is a convolutional PatchGAN classifier that tries to classify if each  $70 \times 70$  patch in an image is real or fake. By averaging over all image's patches, the overall response of the discriminator is obtained. The total amount of parameters for Pix2Pix is  $5.7 \cdot 10^7$ .

#### 4.2.3. CycleGAN

The CycleGAN is a type of GAN that learns to translate images from one domain to another without the need for paired samples between the two domains (Zhu et al., 2017). It comprises two generators,  $G$  and  $F$ , and two discriminators,  $D_y$  and  $D_x$ . The generator  $G$  learns the mapping from the  $X$  domain to the  $Y$  domain, while the generator  $F$  learns the mapping from the  $Y$  domain to the  $X$  domain. The discriminator  $D_y$  tries to distinguish between the real image  $y$  in the  $Y$  domain and

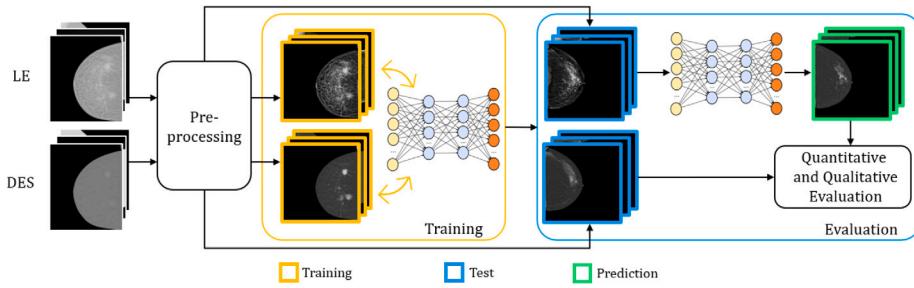
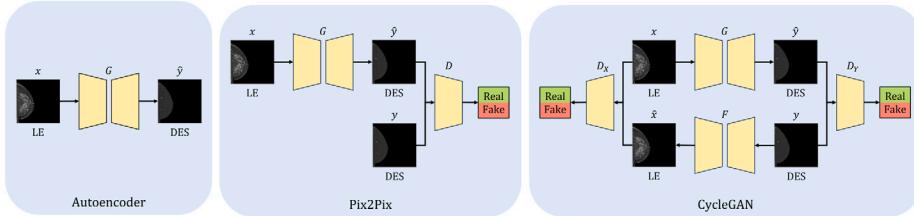


Fig. 3. Schematic representation of the methodology.

Fig. 4. Representative diagrams of Autoencoder, Pix2Pix and CycleGAN translating from LE to DES images. Symbols:  $x$  is the input image,  $\hat{y}$  is the output image,  $y$  is the target image,  $G$  and  $F$  are the generators,  $D$  are the discriminators.

the generated image  $y = \hat{G}(x)$ . In the same way, the discriminator  $D_x$  tries to distinguish between the real image  $x$  in the  $X$  domain and the generated image  $\hat{x} = F(y)$ . To minimize the difference between the distributions of real and generated images, the adversarial loss  $\mathcal{L}_{GAN}(G, D_y)$  is used to adversarially train the generator  $G$  and the discriminator  $D_y$ , with  $G$  trying to maximize  $\mathcal{L}_{adv}(G, D_y)$  to fool  $D_y$ , while  $D_y$  tries to minimize it. It is expressed as:

$$\mathcal{L}_{GAN}(G, D_y) = \mathbb{E}_y[\log(D_y(y))] + \mathbb{E}_x[\log(1 - D_y(G(x)))] \quad (5)$$

Similarly, the adversarial loss  $\mathcal{L}_{GAN}(F, D_x)$  is used to train adversarially the generator  $F$  and the discriminator  $D_x$ . To have consistent mappings, the generators should satisfy the forward cycle consistency, being able to bring  $x$  back to the original image, i.e.  $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ , and the backward cycle consistency, that is  $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ . This behavior is encouraged using a cycle consistency loss  $\mathcal{L}_{cyc}(G, F)$ , that helps minimize the difference between  $x$  and  $F(G(x))$ , and between  $y$  and  $G(F(y))$ . It is formulated as:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_y[\|G(F(y)) - y\|_1] + \mathbb{E}_x[\|F(G(x)) - x\|_1] \quad (6)$$

Moreover, an identity mapping loss  $\mathcal{L}_{id}(G, F)$  can be used to ensure that when real samples of the target domain are provided as input to the generator, they are identically mapped to the output. It is formulated as:

$$\mathcal{L}_{id}(G, F) = \mathbb{E}_y[\|G(y) - y\|_1] + \mathbb{E}_x[\|F(x) - x\|_1] \quad (7)$$

The CycleGAN full objective can thus be expressed as:

$$\begin{aligned} G^*, F^* = \arg \min_{G, F} \max_{D_x, D_y} & [\mathcal{L}_{GAN}(G, D_y) + \mathcal{L}_{GAN}(F, D_x) \\ & + \lambda_1 \mathcal{L}_{cyc}(G, F) + \lambda_2 \mathcal{L}_{id}(G, F)] \end{aligned} \quad (8)$$

where  $\lambda_1$  and  $\lambda_2$  balance the relative contribution of the loss terms.

The architecture of the generators initiates with three convolutional blocks: one dedicated to feature extraction from the input image and two for down-sampling. Each block is composed of a convolutional layer, followed by instance normalization and ReLU activation function. These three convolutional blocks are succeeded by multiple Residual Blocks, each comprised of two convolutional layers, crucial for preserving significant details during image transformation. Following these are two convolutional blocks for up-sampling, implemented with transposed convolutional layers, subsequently followed by instance

normalization and ReLU activation function. The architecture concludes with a single final convolutional layer. On the other hand, the architecture of the discriminators is that of a PatchGAN, examining  $70 \times 70$  patches to classify the image as real or synthetic. The overall architecture is in line with the one presented in Zhu et al. (2017). The total amount of parameters for CycleGAN is  $2.8 \cdot 10^7$

#### 4.3. Training

The Autoencoder, Pix2Pix, and CycleGAN models were trained and evaluated on the pre-processed CESM@UCBM dataset so that they take LE images  $x$  as input and generate the virtually contrast-enhanced DES images  $\hat{y}$  as output, as similar as possible to the corresponding real DES images  $y$ . All the experiments are computed in 10-fold cross-validation, with the training, validation, and test sets following the respective proportions: 0.8, 0.1, and 0.1. To facilitate the training, a pre-training on the publicly available dataset of CESM images (Khaled et al., 2022) was performed, following the same pre-processing as described in Section 4.1.

To prevent overfitting, we proposed a random data augmentation on the images of the training set in terms of vertical or horizontal shift (by a maximum of  $\pm 10\%$  of the original dimension), zoom (by a maximum of  $\pm 10\%$ ), horizontal flip, and rotation (by a maximum of  $\pm 15^\circ$ ). Each model is trained for a maximum of 200 epochs using an early stopping criterion of 50 epochs following the validation loss.

To train the Autoencoder, we used the  $L1$  loss, and we chose the Adam optimizer for optimizing the model, with a learning rate of  $10^{-3}$ , a weight decay of  $10^{-5}$ , a beta of 0.9 and a momentum of 1.

For the Pix2Pix, the generator network  $G$  and the discriminator network  $D$  were trained simultaneously. Referring to Eq. (4), we used the binary cross entropy as the adversarial loss  $\mathcal{L}_{GAN}(G, D)$ , and we added the  $L1$  loss  $\mathcal{L}_{L1}(G)$  with a  $\lambda$  factor of 100 for the generator's training. For both the generator network and the discriminator network, we used the Adam optimizer with a learning rate of  $2 \cdot 10^{-4}$ , a weight decay of  $10^{-5}$ , a beta of 0.5 and a momentum of 1.

For the CycleGAN, the two generator networks,  $G$  and  $F$ , and the two discriminator networks,  $D_y$  and  $D_x$ , are trained simultaneously. Typically, CycleGAN can work with unpaired datasets, but in this work, having access to a paired dataset, we maintained image pairing in the loss computation, comparing the virtually produced DES images to the corresponding real DES images. Referring to Eq. (8), we utilized the

mean squared error for both the adversarial losses  $\mathcal{L}_{GAN}(G, D_y)$  and  $\mathcal{L}_{GAN}(F, D_x)$ . Furthermore, we employed the  $L1$  loss for both the cycle consistency loss, with a  $\lambda_1$  factor of 10, and the identity mapping loss, with a  $\lambda_2$  factor of 5. For both generator and discriminator networks, we used the Adam optimizer with a learning rate of  $10^{-5}$ , a weight decay of  $10^{-5}$ , a beta of 0.5, and a momentum of 1.

For all the models, we did not further investigate any other hyperparameter configuration since their tuning is out of the scope of this manuscript. Nevertheless, the “No Free Lunch” Theorem for optimization states that no universal set of hyperparameters will optimize a model’s performance across all possible datasets.

#### 4.4. Quantitative evaluation

The quantitative analysis was performed by calculating four metrics between the synthetic DES images  $\hat{y}$  and the target DES images  $y$ . They are mean squared error (MSE), peak-signal-to-noise ratio (PSNR), structural similarity index (SSIM), and visual information fidelity (VIF).

The MSE measures the mean squared difference between the pixel values of the target image  $y$  and the reconstructed image  $\hat{y}$ , thus is formulated as:

$$\text{MSE}(y, \hat{y}) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - \hat{y}_{ij})^2 \quad (9)$$

where  $m$  and  $n$  are the number of rows and columns in the images, respectively, and  $y_{ij}$  and  $\hat{y}_{ij}$  represent the pixels elements at the  $i$ th row and  $j$ th column of  $y$  and  $\hat{y}$ , respectively. It varies in the range  $[0, \infty]$ ; the lower its value, the higher the quality of the reconstructed image.

The PSNR is defined as the ratio of the maximum possible power of a signal to the power of the noise corrupting the signal. In our case, the signal is the target image, while the noise is the error introduced by its reconstruction. The PSNR is typically expressed in decibels (dB), where a PSNR value of 30 dB is considered excellent quality, a value of 27 dB is considered good quality, a value of 24 dB is considered poor quality, and a value of 21 dB is considered bad quality (Ponomarenko et al., 2008). It is commonly expressed as a function of the MSE as follows:

$$\text{PSNR}(y, \hat{y}) = 10 \cdot \log_{10} \left( \frac{\max^2(y)}{\text{MSE}(y, \hat{y})} \right) \quad (10)$$

The VIF (Sheikh and Bovik, 2006) evaluates the quality of visual information in the synthetic image  $\hat{y}$  compared to the target image  $y$ . It is derived by modeling the Human-Visual-System in the wavelet domain and is formulated as:

$$\text{VIF}(y, \hat{y}) = \frac{\sum_{j \in \text{subbands}} I(y^j)}{\sum_{j \in \text{subbands}} I(\hat{y}^j)} \quad (11)$$

where  $I(y^j)$  and  $I(\hat{y}^j)$  represent the information ideally extracted by the brain from a certain subband in the target image and synthetic image, respectively. It varies in the range  $[0, 1]$ , the higher its value, the higher the quality of the synthetic image (Sheikh and Bovik, 2006).

The SSIM (Wang et al., 2004) measures the similarity between two images by comparing luminance, contrast, and structure. The formulation is as follows:

$$\text{SSIM}(y, \hat{y}) = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)} \quad (12)$$

where  $\mu_y$  and  $\mu_{\hat{y}}$  represent the means of the images  $y$  and  $\hat{y}$  respectively, while  $\sigma_y$  and  $\sigma_{\hat{y}}$  represent their standard deviations.  $C_1$  and  $C_2$  are small constants used for stabilization. The SSIM varies in the range  $[0, 1]$ , the higher its value, the greater the similarity between the two images (Wang et al., 2004).

#### 4.5. Qualitative evaluation

The goal of the qualitative analysis is to have a medical assessment of the synthetic DES images generated by the models. It consists of two

visual Turing tests presented to four expert radiologists with experience ranging from 8 to 38 years. For both assessments, we set up a replicated radiology reading room environment, and we randomly selected images from 50 patients of the CESM@UCBM test sets from the different cross-validation folds. To have a global consensus, the final human-based decisions were given by majority voting weighted by the years of experience of the radiologists (Jones and Hunter, 1995).

The first visual Turing test helps find the DL model producing the most realistic synthetic DES images. To this end, we showed radiologists side-by-side synthetic DES images generated by the Autoencoder, Pix2Pix, and CycleGAN for the same LE input image. Then, each radiologist had to independently choose the most realistic image without knowing which model generated it.

The second visual Turing test aims to study not only if synthetic images are perceived as real ones but also if they can be used to assign the BI-RADS score, a measure indicating patients’ risk of developing breast cancer (Section 3). We set the input size of the model winning the first visual Turing test to  $512 \times 512$ , and then we retrained it. Note that this is a higher image size than the one used in previous experiments, i.e.,  $256 \times 256$  as reported in Section 4.1: while this has the beneficial effect of producing more realistic images, it substantially increases the computational costs. Hence, we provided the four radiologists with both real DES images and these newly generated synthetic DES images. All the images are presented one at a time, in random order: the radiologist specified whether the displayed image is real or synthetic, and he/she scored the BI-RADS.

## 5. Results and discussion

In this Section, we present and discuss the results obtained from the quantitative and the qualitative evaluations, conducted as described in Sections 4.4 and 4.5, respectively. Given that this is the first time AI is being utilized to execute VCE on CESM images, we establish these results as the baseline performance.

### 5.1. Quantitative evaluation

**Table 5.1** shows the outcomes of the quantitative analysis performed for the Autoencoder, Pix2Pix, and CycleGAN on the whole test sets. For each model, the metrics MSE, PSNR, VIF, and SSIM are computed between all the target DES images of the test sets from the 10 folds and their corresponding synthetic DES images.

The values of these metrics are presented as the mean value  $\pm$  standard deviation calculated across the different folds. For each metric, the most performing model is highlighted in bold. Furthermore, we performed paired t-tests for each metric to determine any statistical differences in the performances of the models. The results are presented in **Table 5.2** with symbols \*\*\*, \*\*, or \*, indicating p-values lower than 0.001, 0.01, or 0.05, respectively. Empty cells indicate instances where the p-value exceeds 0.05. Referring to **Table 5.1**, the Autoencoder is the worst-performing architecture according to MSE, PSNR, and SSIM, and this is supported by statistical results in the aforementioned **Table 5.2**. As shown in **Fig. 5**, the Autoencoder produces images that are generally blurry, often too similar to the input image rather than the corresponding target image, or characterized by the presence of non-physiological hyperintense areas. These results are in line with expectations, as autoencoders were among the first methods used for image-to-image translation, and it is known that they have the limitation of generating blurry images, partly due to the element-wise criteria typically used, such as  $L1$  loss and  $L2$  loss (Yoo et al., 2019; Liu et al., 2021b). Referring to **Table 5.1**, Pix2Pix obtains the worst mean value of VIF, but also the best mean value of SSIM, and both these results are supported by statistical analysis in **Table 5.2**. On the other hand, CycleGAN obtains the best mean values of MSE, PSNR, and VIF, but, as shown in **Table 5.2**, it is not always statistically supported. As displayed in **Fig. 5**, both Pix2Pix and CycleGAN exhibit superior fidelity to the

**Table 5.1**

Quantitative analysis performed for Autoencoder, Pix2Pix, and CycleGAN on the whole test set. Each tabular denotes the mean  $\pm$  standard deviation from the 10-folds.

Model	MSE $\downarrow$	PSNR $\uparrow$	VIF $\uparrow$	SSIM $\uparrow$
Autoencoder	0.0083 $\pm$ 0.0036	23.2775 $\pm$ 2.4942	0.1581 $\pm$ 0.0212	0.5365 $\pm$ 0.2104
Pix2Pix	0.0042 $\pm$ 0.0008	26.3476 $\pm$ 0.7640	0.1129 $\pm$ 0.0106	<b>0.8575 <math>\pm</math> 0.0132</b>
CycleGAN	<b>0.0038 <math>\pm</math> 0.0010</b>	<b>26.4866 <math>\pm</math> 0.9206</b>	<b>0.1840 <math>\pm</math> 0.0132</b>	0.8492 $\pm$ 0.0131

**Table 5.2**

Paired t-test results between all pairs of models for the considered metrics. Symbols \*\*\*, \*\*, or \*, indicate p-values lower than 0.001, 0.01, or 0.05, respectively. Empty cells indicate instances where the p-value exceeds 0.05.

Compared models	MSE	PSNR	VIF	SSIM
Autoencoder - Pix2Pix	**	**	***	***
Autoencoder - CycleGAN	**	**		***
Pix2Pix - CycleGAN			***	**

target images in comparison to the Autoencoder. However, the images produced by Pix2Pix do not have sufficient informative content, as the low VIF value proves. Frequently, the output of Pix2Pix exhibits a deficiency in accurately representing the virtual enhancement of contrast by the lesion, alongside discrepancies in texture compared to the target image. On the contrary, despite statistical analyses not confirming the superiority of CycleGAN, Fig. 5 suggests that CycleGAN is the most promising network for generating synthetic DES images, as it demonstrates superior capability in reconstructing contrast enhancement. It is worth noting that such results agree with those reported by Azarfar et al. who implemented VCE on CT (Azarfar et al., 2023): indeed, they proved that the CycleGAN outperforms conditional GANs in terms of performance, likely due to its ability to learn heuristically from all the images in the paired dataset without being restricted to specific pixel relationships as the Pix2Pix.

Following the comprehensive evaluation of model performance through metric calculations across the entire test datasets, we need to assess the robustness of these models across varying mammographic densities. Indeed, it is crucial to ensure that model performance remains consistent even in the most critical cases, represented by ACR categories *c* (heterogeneously dense breast tissue) and *d* (extremely dense breasts). Indeed, heterogeneously or extremely dense breast parenchyma can cause tumor lesions not to be clearly identifiable in images generated by FFDM. Therefore, in these cases, there is a risk that DL models may find it more complex to generate virtual DES images using only LE images, which are analogous to FFDM images. For these reasons, we also compute the MSE, PSNR, VIF, and SSIM for each model by distinguishing the test sets' images according to the breasts' ACR categories *a*, *b*, *c*, and *d* which define mammographic density. Table 5.3 reports the outcomes of this analysis, with the metrics values reported as mean value  $\pm$  standard deviation calculated on the different folds. Within each model, the values of a specific metric are compared across the different categories, with the best values highlighted in bold. However, our analysis indicates that there are no statistically significant differences (Kruskal-Wallis *p*-value  $> 0.05$ ) in the values of the same metric when calculated for the four categories within each model. Additionally, it is worth noting that performance does not deteriorate in the critical cases *c* and *d*. Hence, these findings suggest that the implemented models demonstrate robustness across the different mammographic densities.

## 5.2. Qualitative evaluation

In addition to the quantitative analysis, the qualitative analysis carried out with the support of radiologists has been crucial in evaluating the performance of the DL models. The first visual Turing test was proposed to identify the model capable of generating the most realistic DES images. In this regard, for each set of images shown,

all radiologists consistently determined that the images most realistic were those generated by the CycleGAN, choosing it as the most reliable network for all the cases. For this reason, we further investigated the CycleGAN, and we conducted the second visual Turing test by using both the real DES images and the CycleGAN-generated DES images. The results are graphically represented in Fig. 6, which shows the True Positive Rate and True Negative Rate obtained by radiologists in distinguishing real images from synthetic images, as well as the Accuracy in associating the BI-RADS descriptor on real and synthetic images. In order to calculate the True Positive Rate and True Negative Rate, we have assigned positive samples to real DES images  $y$  and negative samples to synthetic DES images  $\hat{y}$ . Therefore, it is desirable to achieve a high True Positive Rate and a low True Negative Rate. The True Positive Rate is higher when radiologists are better at labeling real DES images, reflecting their reliability. The True Negative Rate is higher when radiologists are better at labeling synthetic DES images. Consequently, a low True Negative Rate indicates that the network generates realistic synthetic DES images that can confuse radiologists into thinking they are real.

On the other hand, we evaluated the Accuracy of assigning the BI-RADS descriptor separately for real DES images and synthetic DES images. In both subsets, it was calculated as the ratio of correctly assigned descriptors to the total number of descriptors assigned. Therefore, obtaining a high Accuracy value is desirable for both real and synthetic images.

The results show that a True Positive Rate of 100% has been achieved, meaning that all true DES images were correctly recognized, and a True Negative Rate of 15.4%, indicating that only 15.4% of the synthetic DES images were identified as fake, while the remaining 84.6% were classified as real.

Additionally, the Accuracy in assigning the BI-RADS descriptor was found to be 80% for both real DES images and synthetic DES images. Thus, the percentage of error on real DES images is the same as the error on synthetic DES images.

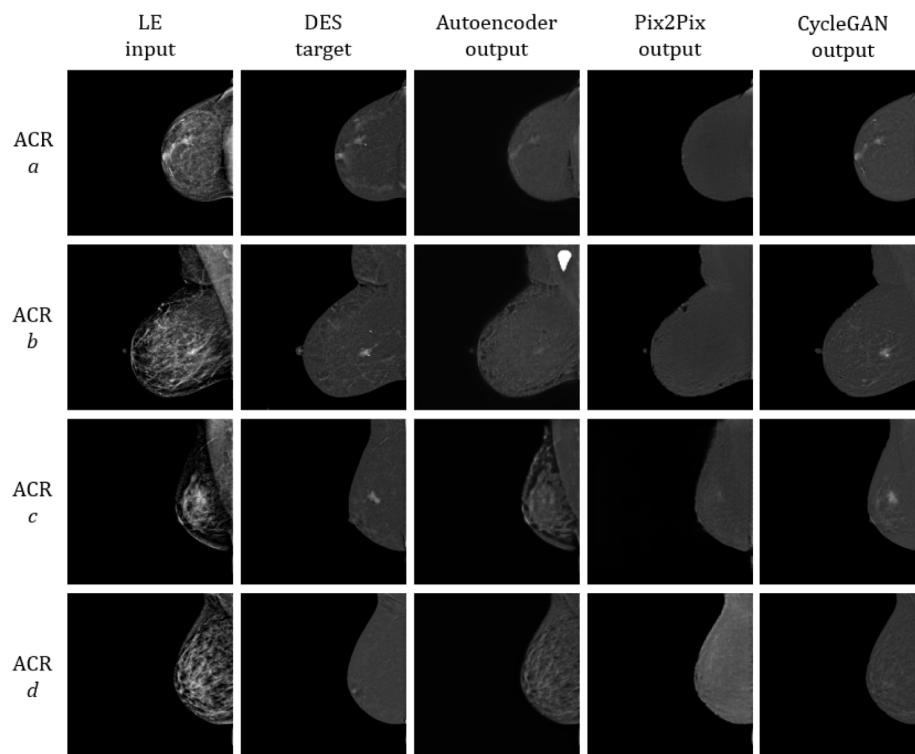
These results suggest that the synthetic DES images have the potential to be used for diagnosis, so it is possible to apply the VCE on CESM images by using AI techniques. Thereby, the advantages of CESM, such as better diagnostic accuracy compared to FFDM, especially in patients with dense breast tissue, could be achieved without the requirement of intravenous administration of contrast medium or the need for dual-energy acquisition.

Moreover, it is interesting to note that being able to generate CESM DES images when FFDM images are already available also has advantages in terms of energy consumption. Indeed, utilizing AI for CESM image generation is significantly more efficient compared to traditional CESM scans. To quantify the energy consumption involved in generating images using DL models, we consider the NVIDIA Tesla V100 as a representative GPU, which has a power usage of 250 W. Based on our experiments, each image translation operation takes 0.1 s. Therefore, the energy consumption per image can be calculated as  $250 \text{ W} \times 0.1 \text{ s} = 25 \text{ W-s} = 6.9 \times 10^{-6} \text{ kWh}$ . On the other hand, estimating the energy consumption of a CESM scan requires considering the power usage of the scanner, reported at 5 kW (as detailed in the datasheet of the Senographe Pristina mammography system). Assuming each CESM scan takes 2 min (Bhimani et al., 2017), the energy consumption per scan can be calculated as  $5000 \text{ W} \times 2 \text{ min} = 10000 \text{ W-min} = 1.7 \times 10^{-1} \text{ kWh}$ . This is five orders of magnitude larger than the energy consumption for DL image generation, and since energy consumption

**Table 5.3**

Quantitative analysis performed for Autoencoder, Pix2Pix and CycleGAN distinguishing the test set images according to the breast's ACR categories *a*, *b*, *c*, and *d*.

Model	ACR	MSE ↓	PSNR ↑	VIF ↑	SSIM ↑
Autoencoder	<i>a</i>	$0.0108 \pm 0.0048$	$21.6093 \pm 2.9097$	$0.1384 \pm 0.0270$	$0.5054 \pm 0.2055$
	<i>b</i>	$0.0114 \pm 0.0052$	$21.7321 \pm 2.9283$	$0.1676 \pm 0.0249$	$0.5426 \pm 0.2071$
	<i>c</i>	$0.0065 \pm 0.0031$	$23.8066 \pm 2.2557$	$0.1566 \pm 0.0247$	$0.5358 \pm 0.2110$
	<i>d</i>	<b><math>0.0053 \pm 0.0024</math></b>	<b><math>24.5332 \pm 2.7948</math></b>	$0.1612 \pm 0.0256$	$0.5274 \pm 0.2108$
Pix2Pix	<i>a</i>	$0.0034 \pm 0.0014$	$25.8957 \pm 1.4328$	$0.1505 \pm 0.0322$	$0.8170 \pm 0.0529$
	<i>b</i>	$0.0047 \pm 0.0026$	$26.0185 \pm 2.3233$	$0.1807 \pm 0.0173$	$0.8365 \pm 0.0207$
	<i>c</i>	<b><math>0.0029 \pm 0.0009</math></b>	<b><math>27.1512 \pm 1.4586</math></b>	$0.1738 \pm 0.0139$	<b><math>0.8546 \pm 0.0129</math></b>
	<i>d</i>	$0.0031 \pm 0.0013$	$27.0355 \pm 1.2134$	<b><math>0.1917 \pm 0.0310</math></b>	$0.8509 \pm 0.0424$
CycleGAN	<i>a</i>	$0.0055 \pm 0.0036$	$24.7935 \pm 1.8855$	$0.1582 \pm 0.0320$	$0.8189 \pm 0.0500$
	<i>b</i>	$0.0060 \pm 0.0038$	$25.9411 \pm 2.0808$	<b><math>0.1982 \pm 0.0147</math></b>	$0.8486 \pm 0.0196$
	<i>c</i>	$0.0039 \pm 0.0016$	$26.2804 \pm 1.1737$	$0.1827 \pm 0.0135$	<b><math>0.8564 \pm 0.0103</math></b>
	<i>d</i>	<b><math>0.0028 \pm 0.0007</math></b>	<b><math>26.9839 \pm 1.1798</math></b>	$0.1859 \pm 0.0332$	$0.8439 \pm 0.0497$



**Fig. 5.** Four representative cases of LE input images, DES target images, and output images of Autoencoder, Pix2Pix, and CycleGAN (from left to right) varying the ACR category (from top to bottom).

correlates with CO<sub>2</sub> emissions, the environmental impact of traditional CESM scans is considerably higher. It is important to note that these considerations are confined to the inference level, without considering model training, nor the production, delivery, installation, and testing of the mammography system.

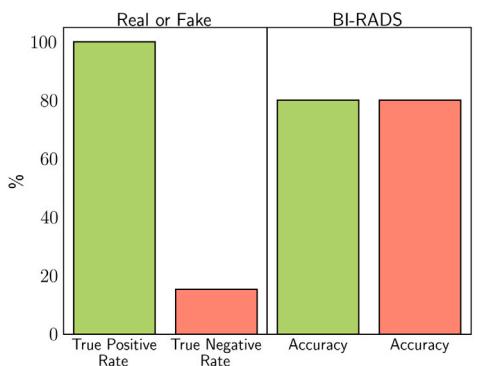
Furthermore, it is worth noting that while the proposed method assumes translation from images without contrast medium to images with contrast medium, the LE images used for the present study were actually acquired after the contrast medium was administered. Nevertheless, the assumption that the LE images are contrast-free images is justified because they were acquired in an energy range where the influence of iodine K-edge is considered to be negligible (Francescone et al., 2014).

## 6. Conclusions

CESM is a breast imaging technique that produces LE images, similar to FFDM, and DES images where a tumor lesion can be easily

seen thanks to an injected iodine contrast medium. In this context, our work offers two contributions. First, this study has demonstrated that VCE can be applied to CESM images avoiding the need for contrast medium administration and minimizing radiation exposure. This claim is supported by experiments using three well-established generative deep models, namely an autoencoder, Pix2Pix, and CycleGAN, which generate DES images solely from LE images. The quantitative analysis results, pertaining to the computation of evaluation metrics, highlighted the inferiority of the autoencoder's performance compared to Pix2Pix and CycleGAN, but not a clear difference between the two. Conversely, the qualitative analysis findings, derived from visual Turing tests administered to four expert radiologists, underscored that CycleGAN provides the best-quality synthetic images, which are often hard to distinguish from real images, even for radiologists. Furthermore, the qualitative results suggest that BI-RADS scores can be assigned from CycleGAN-generated images with an accuracy equal to using real images.

Second, we have presented and made publicly accessible a unique dataset of CESM images. We have taken great care to ensure the



**Fig. 6.** Results of the second visual Turing test. True Positive Rate and True Negative Rate are the percentages at which the radiologists correctly label real images  $y$  and synthetic images  $\hat{y}$ , respectively. The Accuracy is the percentage at which the radiologists correctly assign the BI-RADS score to  $y$  and  $\hat{y}$ .

dataset's value to the scientific community by including data extracted from the medical reports and from the biopsies, together with DICOM images. This offers image information available in the DICOM tags, which would not be accessible in JPEG format as is the case of the other CESM public dataset (Khaled et al., 2022). We hope that our dataset provides all the necessary information for conducting CESM analyses, and we extend an invitation to the scientific community to delve deeper into this area.

Despite these promising findings, the way to the clinical adoption of VCE technology in clinical practice needs further research and validation, including the development of custom deep architectures that optimize previous ones and the incorporation of loss functions that prioritize lesion reconstruction during training. We are working to enlarge the dataset and to set up external validation involving other centers. Furthermore, while here we have generated full-dose DES images from LE images that simulate images with zero-dose contrast medium, we deem that it would be interesting to implement models producing full-dose DES images from low-dose DES images as this approach holds the potential for even more improved outcomes.

#### CRediT authorship contribution statement

**Aurora Rofena:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Valerio Guarasci:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Marina Sarli:** Data curation, Resources. **Claudia Lucia Piccolo:** Conceptualization, Resources. **Matteo Sammarra:** Conceptualization, Resources. **Bruno Beomonte Zobel:** Conceptualization, Resources. **Paolo Soda:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgments

Aurora Rofena is a Ph.D. student enrolled in the National Ph.D. in Artificial Intelligence, XXXVIII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma.

This work was partially founded by: (i) Bando Accordo Innovazione DM 24/5/2017 (Ministero delle Imprese e del Made in Italy), under the project with CUP B89J23000580005; (ii) PNRR MUR project PE0000013-FAIR; (iii) University Campus Bio-Medico di Roma under the program “University Strategic Projects” within the project “AI-powered Digital Twin for next-generation lung cancEr cAre (IDEA)”.

Resources are provided by the National Academic Infrastructure for Supercomputing in Sweden (NAIIS) and the Swedish National Infrastructure for Computing (SNIC) at Alvis @ C3SE, partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973.

#### References

- Azarfar, Ghazal, Ko, Seok-Bum, Adams, Scott J, Babyn, Paul S, 2023. Applications of deep learning to reduce the need for iodinated contrast media for CT imaging: A systematic review. *Int. J. Comput. Assist. Radiol. Surg.* 1–12.
- Bhimani, Chandni, Matta, Danielle, Roth, Robyn G, Liao, Lydia, Tinney, Elizabeth, Brill, Kristin, Germaine, Pauline, 2017. Contrast-enhanced spectral mammography: technique, indications, and clinical applications. *Acad. Radiol.* 24 (1), 84–88.
- Choi, Jae Won, Cho, Yeon Jin, Ha, Ji Young, Lee, Seul Bi, Lee, Seunghyun, Choi, Young Hun, Cheon, Jung-Eun, Kim, Woo Sun, 2021. Generating synthetic contrast enhancement from non-contrast chest computed tomography using a generative adversarial network. *Sci. Rep.* 11 (1), 20403.
- D'orsi, CJ, Bassett, LWea, Berg, WA, Feig, SA, Jackson, VP, Kopans, DB, et al., 2003. Breast imaging reporting and data system: ACR BI-RADS-mammography. *Am. College Radiol. (ACR)*, Reston 230–234.
- Elkassas, Hebatalla, El-Maadawy, Samar M, Saad, Sherihan Mahmoud, Radwan, Amira H, Gareer, Sherihan WY, 2022. Role of contrast-enhanced spectral mammography in the assessment of residual disease following neoadjuvant chemotherapy in patients with breast cancer. *Egypt. J. Radiol. Nucl. Med.* 53 (1), 249.
- Francescone, Mark A, Jochelson, Maxine S, Dershaw, D David, Sung, Janice S, Hughes, Mary C, Zheng, Junting, Moskowitz, Chaya, Morris, Elizabeth A, 2014. Low energy mammogram obtained in contrast-enhanced digital mammography (CEDM) is comparable to routine full-field digital mammography (FFDM). *Eur. J. Radiol.* 83 (8), 1350–1355.
- Iotti, Valentina, Ravaioli, Sara, Vacondio, Rita, Coriani, Chiara, Caffari, Sabrina, Sghedoni, Roberto, Nitroso, Andrea, Ragazzi, Moira, Gasparini, Elisa, Masini, Cristina, et al., 2017. Contrast-enhanced spectral mammography in neoadjuvant chemotherapy monitoring: A comparison with breast magnetic resonance imaging. *Breast Cancer Res.* 19, 1–13.
- Isola, Phillip, Zhu, Jun-Yan, Zhou, Tinghui, Efros, Alexei A, 2017. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1125–1134.
- James, Jonathan, 2022. Contrast-enhanced spectral mammography (CESM)-guided breast biopsy as an alternative to MRI-guided biopsy. *Brit. J. Radiol.* 95 (1132), 20211287.
- James, J.J., Tennant, S.L., 2018. Contrast-enhanced spectral mammography (CESM). *Clin. Radiol.* 73 (8), 715–723.
- Jochelson, Maxine S., Lobbes, Marc B.I., 2021. Contrast-enhanced mammography: State of the art. *Radiology* 299 (1), 36–48.
- Jones, Jeremy, Hunter, Duncan, 1995. Qualitative research: Consensus methods for medical and health services research. *BMJ* 311 (7001), 376–380.
- Kaji, Shizuo, Kida, Satoshi, 2019. Overview of image-to-image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiol. Phys. Technol.* 12, 235–248.
- Khaled, Rana, Helal, Maha, Alfarghaly, Omar, Mokhtar, Omnia, Elkorany, Abeer, El Kasas, Hebatalla, Fahmy, Aly, 2022. Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. *Sci. Data* 9 (1), 1–10.
- Kleesiek, Jens, Morshuis, Jan Nikolas, Isensee, Fabian, Deike-Hofmann, Katerina, Paech, Daniel, Kickingereder, Philipp, Köthe, Ullrich, Rother, Carsten, Försting, Michael, Wick, Wolfgang, et al., 2019. Can virtual contrast enhancement in brain MRI replace gadolinium?: A feasibility study. *Invest. Radiol.* 54 (10), 653–660.
- Liu, Yanxia, Chen, Anni, Shi, Hongyu, Huang, Sijuan, Zheng, Wanjia, Liu, Zhiqiang, Zhang, Qin, Yang, Xin, 2021a. CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. *Comput. Med. Imaging Graph.* 91, 101953.
- Liu, Xiaofeng, Xing, Fangxu, El Fakhri, Georges, Woo, Jonghye, 2021b. A unified conditional disentanglement framework for multimodal brain mr image translation. In: *2021 IEEE 18th International Symposium on Biomedical Imaging*. ISBI, IEEE, pp. 10–14.

- Lorek, Andrzej, Steinhof-Radwańska, Katarzyna, Barczyk-Gutkowska, Anna, Zarębski, Wojciech, Paleń, Piotr, Szyłuk, Karol, Lorek, Joanna, Grażyńska, Anna, Niemiec, Paweł, Gisterek, Iwona, 2021. The usefulness of spectral mammography in surgical planning of breast cancer treatment—Analysis of 999 patients with primary operable breast cancer. *Curr. Oncol.* 28 (4), 2548–2559.
- Pasternak, Jeffrey J., Williamson, Eric E., 2012. Clinical pharmacology, uses, and adverse reactions of iodinated contrast agents: A primer for the non-radiologist. In: *Mayo Clinic Proceedings*. Vol. 87, Elsevier, pp. 390–402.
- Patel, Bhavika K., Lobbes, M.B.I., Lewin, John, 2018. Contrast enhanced spectral mammography: A review. In: *Seminars in Ultrasound, CT and MRI*. Vol. 39, Elsevier, pp. 70–79.
- Ponomarenko, N., Lukin, V., Egiazarian, K., Astola, J., Carli, M., Battisti, F., 2008. Color image database for evaluation of image quality metrics. In: 2008 IEEE 10th Workshop on Multimedia Signal Processing. pp. 403–408. <http://dx.doi.org/10.1109/MMSP.2008.4665112>.
- Sammarra, Matteo, Piccolo, Claudia Lucia, Sarli, Marina, Stefanucci, Rita, Tommasiello, Manuela, Orsaria, Paolo, Altomare, Vittorio, Beomonte Zobel, Bruno, 2024. Contrast-enhanced mammography-guided biopsy: Preliminary results of a single-center retrospective experience. *J. Clin. Med.* 13 (4), 933.
- Sheikh, Hamid R., Bovik, Alan C., 2006. Image information and visual quality. *IEEE Trans. Image Process.* 15 (2), 430–444.
- Umeshara, Kensuke, Ota, Junko, Ishida, Takayuki, 2018. Application of super-resolution convolutional neural network for enhancing image resolution in chest CT. *J. Digit. Imaging* 31, 441–450.
- Wang, Zhou, Bovik, Alan C., Sheikh, Hamid R., Simoncelli, Eero P., 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 13 (4), 600–612.
- Xie, Huiqiao, Lei, Yang, Wang, Tonghe, Patel, Pretesh, Curran, Walter J., Liu, Tian, Tang, Xiangyang, Yang, Xiaofeng, 2021. Generation of contrast-enhanced CT with residual cycle-consistent generative adversarial network (res-cyclegan). In: *Medical Imaging 2021: Physics of Medical Imaging*. Vol. 11595, SPIE, pp. 1042–1048.
- Yang, Qingsong, Yan, Pingkun, Kalra, Mannudeep K., Wang, Ge, 2017. CT image denoising with perceptive deep neural networks. *arXiv preprint arXiv:1702.07019*.
- Yoo, Jaechang, Eom, Heesong, Choi, Yong Suk, 2019. Image-to-image translation using a cross-domain auto-encoder and decoder. *Appl. Sci.* 9 (22), 4780.
- Zamora, Kathryn, Allen, Elizabeth, Hermezc, Brittany, 2021. Contrast mammography in clinical practice: Current uses and potential diagnostic dilemmas. *Clin. Imaging* 71, 126–135.
- Zhu, Jun-Yan, Park, Taesung, Isola, Phillip, Efros, Alexei A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 2223–2232.