Article

Automatic Classification of Simulated Breast Tomosynthesis Whole Images for the Presence of Microcalcification Clusters Using Deep CNNs

Ana M. Mota [1,*], Matthew J. Clarkson [2], Pedro Almeida [1] and Nuno Matela [1]

1   Instituto de Biofísica e Engenharia Biomédica, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal
2   Department of Medical Physics and Biomedical Engineering and the Centre for Medical Image Computing, University College London, London WC1E 6BT, UK
*   Correspondence: ammota@fc.ul.pt

Abstract: Microcalcification clusters (MCs) are among the most important biomarkers for breast cancer, especially in cases of nonpalpable lesions. The vast majority of deep learning studies on digital breast tomosynthesis (DBT) are focused on detecting and classifying lesions, especially soft-tissue lesions, in small regions of interest previously selected. Only about 25% of the studies are specific to MCs, and all of them are based on the classification of small preselected regions. Classifying the whole image according to the presence or absence of MCs is a difficult task due to the size of MCs and all the information present in an entire image. A completely automatic and direct classification, which receives the entire image, without prior identification of any regions, is crucial for the usefulness of these techniques in a real clinical and screening environment. The main purpose of this work is to implement and evaluate the performance of convolutional neural networks (CNNs) regarding an automatic classification of a complete DBT image for the presence or absence of MCs (without any prior identification of regions). In this work, four popular deep CNNs are trained and compared with a new architecture proposed by us. The main task of these trainings was the classification of DBT cases by absence or presence of MCs. A public database of realistic simulated data was used, and the whole DBT image was taken into account as input. DBT data were considered without and with preprocessing (to study the impact of noise reduction and contrast enhancement methods on the evaluation of MCs with CNNs). The area under the receiver operating characteristic curve (AUC) was used to evaluate the performance. Very promising results were achieved with a maximum AUC of 94.19% for the GoogLeNet. The second-best AUC value was obtained with a new implemented network, CNN-a, with 91.17%. This CNN had the particularity of also being the fastest, thus becoming a very interesting model to be considered in other studies. With this work, encouraging outcomes were achieved in this regard, obtaining similar results to other studies for the detection of larger lesions such as masses. Moreover, given the difficulty of visualizing the MCs, which are often spread over several slices, this work may have an important impact on the clinical analysis of DBT images.

Keywords: digital breast tomosynthesis; microcalcifications; deep-learning; convolutional neural network; virtual clinical trial

1. Introduction

Breast cancer is the most commonly diagnosed type of cancer worldwide [1]. Over the last three decades, mortality rates for breast cancer have dropped from their peak by 41%, likely reflecting advancements in treatment and earlier detection through increased screening programs [2]. However, in women, this disease is still the leading cause of cancer death [1].

Breast screening is crucial in identifying breast cancer at an early stage, when it can be better located and treated, thus reducing the breast cancer mortality. It is estimated

that women who chose to participate in an organized breast cancer screening programs have 60% lower risk of dying from breast cancer within 10 years after diagnosis [3]. Until recently, these screenings and breast cancer detection in general were mainly performed using digital mammography (DM). However, as a result of its 2D nature, DM presents two major limitations: low sensitivity in dense breasts with pathology and low specificity due to normal tissue superposition [4].

The use of digital breast tomosynthesis (DBT) has been confirming the potential of DBT to address these limitations. Initially, DBT was studied and approved in conjunction with DM, demonstrating an increase in breast cancer detection rates and a significant reduction in recall rates [4–9], particularly with dense breasts [6]. Currently, by including synthetic mammography (SM) generated from DBT data, DBT alone is approved as a stand-alone modality to replace DM [10–15].

One major drawback with DBT is its increase in interpretation time when compared to DM [16,17]. Computer-aided detection (CAD) systems with DBT were implemented and evaluated in an attempt to shorten the reading time while maintaining the radiologist performance. In fact, some results are very encouraging with reading time reductions between 14% and 29.2% without loss of diagnostic performance [18–20].

On the other hand, there are mixed observations with respect to DBT technology for the detection of microcalcification clusters (MCs). Some studies have revealed inferior image quality for visibility of MCs with DBT [21–23] while others have not [24–26]. As MCs are among the most important biomarkers for breast cancer [27,28], especially in cases of nonpalpable lesions, another CAD approach that has been extensively studied with DBT is the use of these conventional CAD systems to assist in the correct detection of MCs [29–37]. However, despite the efforts and improvements already achieved (such as decreasing the false negative rate), due to the high false positive rates and low specificity, these CAD systems have not reached a level of performance that can be translated into a true improvement in the real screening of breast cancer [38–41].

In recent years, the increase in computational power has allowed the development of deep learning artificial intelligence (AI) algorithms composed of multilayered convolutional neural networks (CNNs). These AI systems have emerged as a potential solution in the field of automated breast cancer detection in DM and DBT [41]. In fact, recently, there have been several published large-scale studies where the aim was to analyze the performance of AI systems alone, as well as the performance of breast radiologists with and without AI [20,42–49]. The AI systems under evaluation achieved a comparable or even improved cancer detection accuracy when compared with the human experts. With these promising results and the need for an automatic detection system for lesions in DBT and in screening, much research has been carried out in this regard. A brief summary of these studies is presented in Table 1.

Table 1. Summary of deep learning DBT studies (ROI: region of interest, AUC: area under the curve, pAUC: partial AUC).

Ref.	Classification Task	ROI/Patch/Image	Model	Best Metric
[50]	True MCs vs. false positives	ROI (16×16)	Own	AUC: 0.93
[51]	Presence/absence of masses and architectural distortions	Patch (256×256)	Based on AlexNet	Accuracy: 0.8640
[52]	Presence/absence of masses	ROI ($32 \times 32 \times 25$)	Own	AUC: 0.847
[53]	True masses vs. false positives	ROI (128×128)	Own	AUC: 0.90
[54]	True masses vs. false positives	ROI (64×64)	Based on VGG16	AUC: 0.919
[55]	Positive (malignant, benign masses) vs. negative images	Image (224×224)	Based on AlexNet	AUC: 0.6632
[56]	Malignant vs. benign masses	ROI (128×128)	Based on AlexNet	AUC: 0.90

Table 1. *Cont.*

Ref.	Classification Task	ROI/Patch/Image	Model	Best Metric
[57]	Malignant vs. benign masses	Image (256 × 256)	Own	AUC: 0.87
[58]	Presence/absence of MCs	Patch (29 × 29 × 9)	Based on [59]	pAUC: 0.880
[60]	Positive vs. negative volumes	Image (1024 × 1024)	Based on AlexNet, ResNet50, Xception	AUC: 0.854 (AlexNet)
[61]	Positive vs. negative volumes	Image (832 × 832)	Based on AlexNet, ResNet, DenseNet and SqueezeNet	AUC: 0.91 (DenseNet)
[62]	Benign vs. malignant lesions	ROI (224 × 224)	Based on VGG19	AUC (MCs): 0.97
[63]	Positive vs. negative patches	Patch (512 × 512)	Based on ResNet	AUC: 0.847
[64]	Malignant vs. benign vs. normal masses	ROI (256 × 256)	Based on VGG16	AUC: 0.917, 0.951, 0.993 (malignant, benign, normal)
[65]	Malignant vs. benign masses	ROI (224 × 224)	Based on DenseNet121	AUC: 0.8703
[66]	BIRADS 0 vs. BIRADS 1 vs. BIRADS 2	Image (2200 × 1600)	Based on ResNet50	AUC: 0.912 (BIRADS 0 vs. non-0)
[67]	Predict breast density	Image	Based on ResNet34	AUC: 0.952
[68]	True MCs vs. false positives	ROI (128 × 128)	Based on ResNet18	AUC: 0.9765
[69]	Malignant vs. benign vs. normal images	Image (150 × 150)	Own	AUC: 0.89
[70]	Malignant vs. benign MCs	Patch (224 × 224)	Ensemble CNN (2D ResNet34 and anisotropic 3D Resnet)	AUC: 0.8837
[71]	Malignant vs. benign vs. normal slices based on masses and architectural distortions	Image (input size of each CNN: 224 × 224, 227 × 227)	ResNet18, AlexNet, GoogLeNet, VGG16, MobileNetV2, DenseNet201, Mod_AlexNet	Accuracy: 0.9161 (Mod_AlexNet)

The vast majority of these studies focused on detecting and classifying soft-tissue lesions, such as masses [51–57,64,65]. In addition to the fact that these are important lesions for the characterization of breast cancer, in this type of lesion, it is possible to greatly reduce the data input size through interpolation, without losing the spatial resolution required to observe the lesion (the same does not occur with MCs). In this way, faster transfer learning solutions, useful when there is a lack of available training data (as in the case of DBT), can be used with very positive results [53–56,64,65]. Even in cases where only regions of interest (ROIs) and not full images are selected, such resizing is usually carried out. Furthermore, the vast majority of the works use ROIs or patches where objectively there is or is not a lesion [55,57,60,61,66,67,69,70], instead of using the whole image or volume. The use of the whole image or volume is important to contextualize the lesions but also to make the classification a useful and quick tool in screening, where an image/volume should ideally give some type of direct outcome.

One of the biggest challenges involving DBT in AI is the lack of a large, properly labeled public database. All studies mentioned in the Table 1, except one [71], used private databases, making generalization and a fair comparison between different studies impractical [72]. Recently, two publicly accessible annotated DBT datasets that will facilitate the evaluation and validation of AI algorithms were released. Buda et al. made publicly available a large-scale dataset of DBT data. It contains 5610 studies from 5060 patients: 5129 normal cases (no abnormal findings), 280 cases where additional imaging was needed but no biopsy was performed, 112 benign biopsied cases, and 89 cases with proven cancer.

This dataset includes masses and architectural distortions and was used to train and test a single-phase deep learning detection model that reached a baseline sensitivity of 65% at two false positives per DBT volume [73]. El-Shazli et al. used this database to propose a computer-aided multiclass diagnosis system for classifying DBT slices as benign, malignant, or normal considering masses and architectural distortions [71]. The other public dataset resulted from the advancement of in silico tools. The Virtual Imaging Clinical Trial for Regulatory Evaluation (VICTRE) project was created for the evaluation of the imaging performance of DBT as a replacement to DM for breast cancer screening. In VICTRE, the whole imaging chain was simulated with state-of-the-art tools, and a total of 2986 virtual realistic patients were generated and imaged with both modalities. The positive cohort (that comprises malignant spiculated masses and MCs) included 1944 and 1042 virtual patients with and without lesions, respectively [74].

In this paper, fully automatic methods based on deep learning were studied for classifying DBT data. The aim is to input a whole DBT image and have a direct answer about the absence or presence of MCs, without the need for prior identification of lesions in specific regions and, thus, completely automate the process of DBT classification. Four existing popular networks were considered and compared with a new network proposed by us for this purpose. In order to study the impact of some preprocessing methods in increasing the visibility of MCs, the input data were considered with and without preprocessing. The VICTRE public database was used. To the best of our knowledge, this is the first study of automatic classification specifically dedicated to the presence or absence of MCs in whole DBT images.

2. Materials and Methods

2.1. Database

This study was centered on the database created for the VICTRE trial [74]. Synthetic images of virtual patients were obtained using an in silico version of the Siemens Mammomat Inspiration DBT system using Monte Carlo X-ray simulations. These data are available to the public in the Cancer Imaging Archives [75]. Physical compression of left breasts was considered in the craniocaudal (CC) orientation. In this database, the cases are divided into the absence and presence of lesions, as well as according to the density of the breast (fatty, scattered, heterogeneous, and dense). The absent cases have no findings, and each case with lesions present contains four spiked masses with a 5 mm nominal diameter and mass density 2% higher than normal glandular tissue, and four MCs consisting of five calcified lesions modelled as 195, 179, and 171 μm of solid calcium oxalate. In this study, we included cases without ("absent") and with MCs ("present MCs").

Table 2 presents a detailed summary of the dataset selected for this work. The reconstructed cases had different dimension in x, y, and z, depending on breast density: $1624 \times 1324 \times 62$, $1421 \times 1024 \times 57$, $1148 \times 753 \times 47$, and $1130 \times 477 \times 38$ for fatty, scattered, heterogeneous, and dense breasts, respectively, with a voxel size of $0.085 \times 0.085 \times 1$ mm^3. For the absent category, five slices proportionally spaced between the first and the last slice were selected for each case (for example, as fatty cases have 62 slices: slices 1, 17, 33, 49, and 62 were selected; as dense cases have 38 slices: slices 1, 11, 21, 31, and 38 were chosen). On the other hand, for the presentMCs class, slices containing the center of the cluster were selected for each case (in some cases, two clusters had their center on the same slice). Numerically, we adopted the usual distribution of breast density in the population: 10% fatty, 40% scattered, 40% heterogeneous, and 10% dense, with an approximate balance between cases without and with lesions.

Table 2. Detailed summary of the VICTRE data selected for this study.

	Absent		Present MCs	
Density	Number of Cases	Number of Slices	Number of Cases	Number of Slices
Fatty	20	100	25	99
Scattered	80	400	100	386
Heterogeneous	80	400	100	371
Dense	20	100	25	93
Total		1000		949

2.2. Data PreProcessing

In the VICTRE database, the reconstructed data have signal contamination outside the breast region, i.e., in the background (BG). This information is worthless for training the networks and, when present, slows down the process, as pixels without any useful information end up contributing to the mathematical operations involved. In this way, through binarization and region-growing operations, binary masks that keep information belonging to the breast and make everything else zeros were created ("BG suppression"). This step was applied to the original data and after all the other types of processing.

The very-low-dose projections acquired within a limited angular range in a DBT examination result in low statistics (high noise level) in the reconstructed images and data insufficiency. For this reason, image denoising methods are very important in order to improve the image quality of DBT data. Total variation (TV) minimization algorithms have attracted considerable attention in the field because of their ability to smooth images while preserving the edges. Studies applying TV minimization to DBT data have shown excellent results so far [76–80]. This methodology was applied during the preprocessing step. Minimization of TV greatly improves the contrast-to-noise ratio by reducing the noise. In this way, in order to also increase the contrast, two other techniques were studied. The contrast-limited adaptive histogram equalization (CLAHE) technique was implemented to increase the contrast of all breast structures in general, and a simpler operation was applied to increase the contrast of structures with greater intensity, such as MCs, in particular. Since we wanted to study whether image noise reduction or contrast has any impact on CNN training, some combinations of these methods were made, resulting in six different preprocessing approaches (Figure 1), as described below.

Pre-processing 1:	noise reduction	→	BG supression		
Pre-processing 2:	contrast enhancement	→	BG supression		
Pre-processing 3:	noise reduction	→	contrast enhancement	→	BG supression
Pre-processing 4:	contrast enhancement	→	noise reduction	→	BG supression
Pre-processing 5:	normData^2	→	BG supression		
Pre-processing 6:	normData^2	→	noise reduction	→	BG supression

Figure 1. The six preprocessing methodologies implemented in order to reduce noise and amplify the visibility of the MCs (BG: background, normData: data normalized between 0 and 1).

PreProcessing 1: As DBT data are composed of a high level of noise resulting from the acquisition of low-dose projections, the application of a noise reduction filter was analyzed. This filter consists of minimizing the TV of the data, allowing the noise to be significantly reduced while preserving the edges and lesion resolution (which is a very important factor when the structures under analysis are small MCs). TV is a measure of pixel intensity variation in an image and increases significantly in the presence of noise. In each preprocessing that included this filter, several Lagrange multipliers were tested to study which allowed the minimum TV value [78], and 14 was the chosen value for the application of the filter in all cases.

PreProcessing 2: The CLAHE technique [81] was implemented using the MATLAB R2020a function adapthisteq [82] to enhance the contrast of the images and the MCs. With this technique, the contrast in homogeneous areas is limited to avoid the amplification of noise. The contrast transformation function is calculated in small regions of the image individually, rather than in the whole image, and neighboring regions are then combined through bilinear interpolation to eliminate artificially induced boundaries. The contrast enhancement limit was 0.01, and a uniform distribution of the histogram was used with a distribution parameter of 0.4.

PreProcessing 3 and 4: The techniques described for preprocessing 1 and 2 were combined and used together by varying the order in which each one was applied. These steps (3 and 4) were also included since techniques 1 and 2 could complement each other and, through preliminary studies, it was possible to conclude that their order of implementation showed differences in the appearance of the final image. In preprocessing 3, the TV minimization filter for noise reduction was first applied, followed by the contrast enhancement technique. For preprocessing 4, the application was in the opposite order, with contrast enhancement technique first and then noise reduction.

PreProcessing 5: The data intensity was first normalized between 0 and 1 and then squared to attenuate the lower values, highlighting the higher ones belonging to the MCs. With this filter, our aim was to specifically increase the contrast of regions of higher intensities.

PreProcessing 6: The method applied in preprocessing 5 was followed by the TV minimization filter, as described in preprocessing 1.

In order to homogenize the data, as well as to find a balance between training time/memory and the necessary spatial resolution for the visibility and conspicuity of MCs, all data were resized in x and y to 512×512. No adjustments were made in the z-direction since training was performed slice-by-slice. The images were converted into TIFF slices of 8 bits, and input data were normalized using the zero center method.

2.3. CNNs

Since it was crucial to maintain image spatial resolution under certain limits to allow the detection of the small MCs, it was not possible to reduce the image dimension to values such as 224×224 or 227×227, which are the most used in pretrained networks for transfer learning. Our approach was then to train from scratch four architectures that already exist: AlexNet [83], GoogLeNet [84], ResNet18 [85], and SqueezeNet [86]. The choice of these popular networks was based on the comparison of each model's speed and accuracy [87].

In addition, to alleviate some computational effort, one faster and lighter new architecture, based on AlexNet, is proposed by us: CNN-a (Figure 2).

In CNN-a, the channel-wise local response normalization layers were replaced by batch normalization layers ("norm") and a new max pooling layer with a stride of 2, padding of 0, and size of 3×3 was added between the two grouped convolutional layers. These modifications were the result of several empirical trial-and-error studies conducted by us during the experiment.
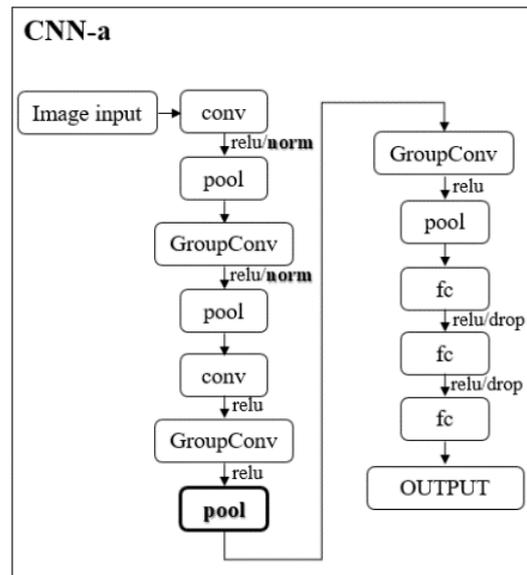
Figure 2. Illustration of CNN-a that resulted from the modifications made (bold) to the AlexNet architecture. Conv and GroupConv: convolutional and grouped convolutional layers, respectively; pool: max pooling layers; fc: fully connected layer; relu: rectified linear unit layer; norm: batch normalization layer; drop: dropout layer.

2.4. Methodology Pipeline

Figure 3 shows the pipeline followed in this work. Absent and presentMCs data samples were selected, and the described preprocessing techniques were applied. The training dataset was used to train the CNNs from scratch, and the testing dataset was used after training to evaluate the performance of the trained CNNs.
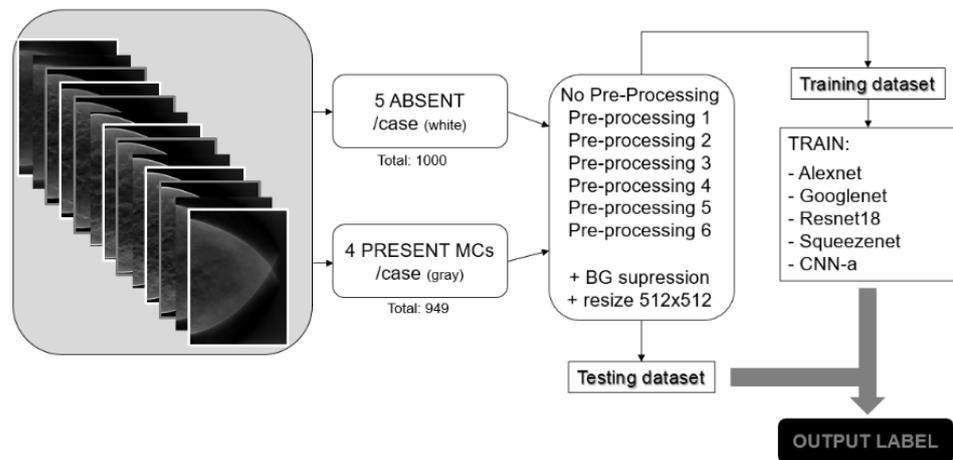
Figure 3. Summary of the methodological pipeline followed in this work.

2.5. Training Options

The *k*-fold technique was used as the cross-validation method to estimate the generalization error of the learning process. The dataset used was divided into $k = 3$ subsets, i.e., each network was trained and tested three times with different datasets, always according to the proportion of two-thirds of the cases for training and one-third for testing. Since the split was performed at the patient level, all the images of the same patient were in either the training set or the test set. Training data augmentation was used through random reflection in the left–right direction (to simulate the inclusion of examples of right breasts) and data rotation between $\pm 20°$.

The CNNs were trained using the stochastic gradient descent optimizer with momentum 0.9 to minimize the cross-entropy loss for classification. The maximum number of epochs was 200 with a mini-batch size of 32 and a learning rate of 1×10^{-3}. In addition to the threefold cross-validation, an L2 regularization term of 5×10^{-3} was introduced in the loss function to prevent overfitting.

2.6. Evaluation Metrics

Classification problems usually involve distinguishing between two classes. In the case of medical imaging, this distinction is usually made between the absence or presence of abnormalities or between benign/malignant lesions. In our work, the objective was to distinguish between the absence or presence of MCs. Sensitivity, specificity, accuracy, and area under the receiver operating characteristic (ROC) curve (AUC) were considered to evaluate the performance. The analysis of only the first three metrics can be limitative because they depend on the defined threshold to accept a case as presentMCs or absent. In this way, we used the AUC (positive class: presentMCs) as a summary tool that contains the space of all these possible thresholds.

Differences in the performance of each classifier were tested using a statistical *t*-test. A two-tailed *p*-value < 0.05 was considered to indicate a significant difference.

3. Results

3.1. Data Preprocessing

All the steps involved in the BG suppression are presented through an example case in Figure 4. The original data were first binarized (Figure 4b) by thresholding, the holes in the image were filled (Figure 4c), the largest resultant object was selected (Figure 4d), and the complete binary mask was achieved by performing region growing in (Figure 4e). The profile traced for the white ROI (lower right corner of (a) and (f)) shows the cleaning effect.
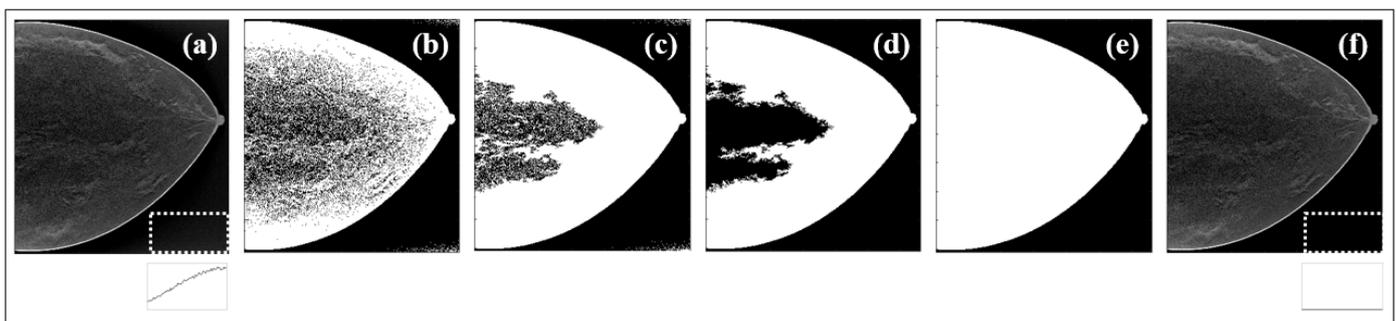
Figure 4. (**a**) Data with contaminated BG; (**b**) first binary image; (**c**) filled binary image; (**d**) largest object extracted from binary image; (**e**) result from region growing; (**f**) final image with BG corrected after binary mask from (**e**) applied to (**a**).

This methodology was included in all preprocessing approaches, as mentioned in Section 2.2. Zooming in on one MC (Figure 5), we can see the different results achieved in this type of lesions with each preprocessing method.
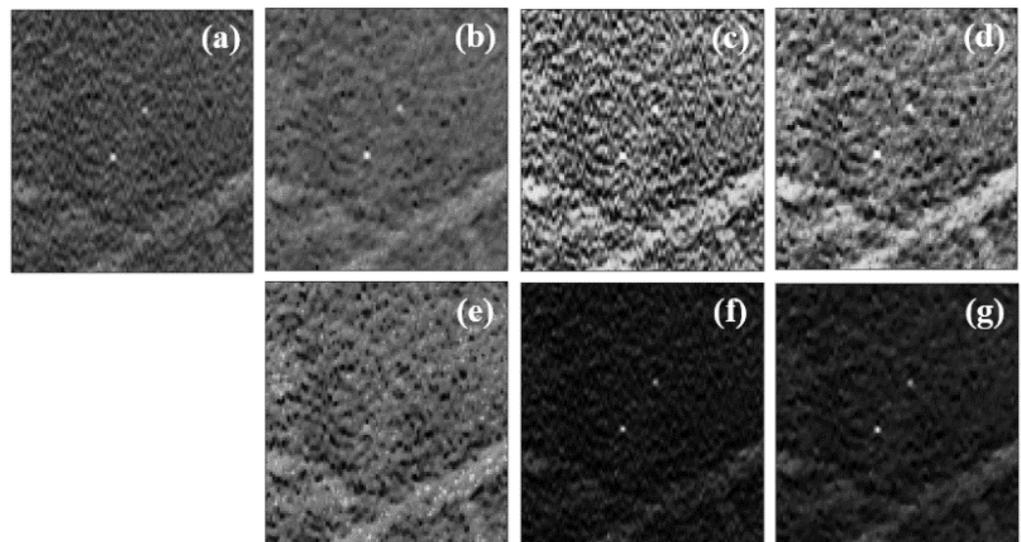
Figure 5. (**a**) Original data without preprocessing; (**b**) preprocessing 1 (minimization of TV); (**c**) preprocessing 2 (CLAHE); (**d**) preprocessing 3 (minTV + CLAHE); (**e**) preprocessing 4 (CLAHE + minTV); (**f**) preprocessing 5 (dataNorm2); (**g**) preprocessing 6 (dataNorm2 + minTV).

3.2. Performance Analysis

Our research was guided by the AUC results obtained for the different architectures and preprocessing methods. As mentioned above, the training and testing were repeated three times (threefold cross-validation) using three distinct datasets. The averaged performances and standard deviation values found over the three folds are shown in Table 3.

Table 3. Performance results of CNNs trained with original data and with data resulting from the preprocessing methodologies, in terms of mean AUC.

|	AUC (%): Mean $\pm$ SD				
	AlexNet	**GoogLeNet**	**ResNet18**	**SqueezeNet**	**CNN-a**
Original data	87.92 ± 2.01	90.14 ± 0.38	86.84 ± 2.62	87.43 ± 0.78	89.79 ± 1.23
Preprocessing 1	87.35 ± 1.63	88.38 ± 1.12	87.96 ± 0.96	88.78 ± 0.99	90.66 ± 0.15
Preprocessing 2	87.29 ± 0.78	93.02 ± 3.59	86.42 ± 3.26	86.84 ± 3.82	86.95 ± 0.97
Preprocessing 3	88.61 ± 0.43	94.19 ± 1.12	86.33 ± 1.46	82.15 ± 1.51	85.80 ± 1.73
Preprocessing 4	90.82 ± 1.29	94.15 ± 1.54	90.13 ± 0.32	86.33 ± 6.31	89.07 ± 1.62
Preprocessing 5	87.62 ± 0.35	88.65 ± 4.27	90.44 ± 0.41	85.18 ± 2.78	89.54 ± 2.63
Preprocessing 6	87.47 ± 1.13	89.76 ± 1.76	89.00 ± 1.33	84.09 ± 3.13	91.17 ± 0.07

In Table 4 presents the *p*-values calculated to study the measurable statistical differences between the best mean AUCs obtained in Table 3.

Table 4. Levels of significance (*p*-values) obtained from the statistical analysis of the difference between the best mean AUCs found.

p-Value	GoogLeNet PreProc3 (94.19 ± 1.12)	ResNet18 PreProc5 (90.44 ± 0.41)	SqueezeNet PreProc1 (88.78 ± 0.99)	CNN-a PreProc6 (91.17 ± 0.07)
AlexNet preProc4 (90.82 ± 1.29)	**0.027** (AlexNet < GoogLeNet)	0.654	0.095	0.662
GoogLeNet preProc3 (94.19 ± 1.12)		**0.006** (GoogLeNet > ResNet18)	**0.003** (GoogLeNet > SqueezeNet)	**0.010** (GoogLeNet > CNN-a)
ResNet18 preProc5 (90.44 ± 0.41)			0.055	**0.038** (ResNet18 < CNN-a)
SqueezeNet preProc1 (88.78 ± 0.99)				**0.014** (SqueezeNet < CNN-a)

p-Values <0.05 (in bold) indicate a significant difference; preProc—preProcessing.

Considering only the best results obtained for averaged AUC, Figure 6 shows the ROC curves of the CNN network trained with the respective data. These curves were obtained by averaging between the ROC curves of each fold. Additionally, Figure 7 analyzes the values of the respective sensitivities, specificities, and accuracy in detecting the cases with MCs.
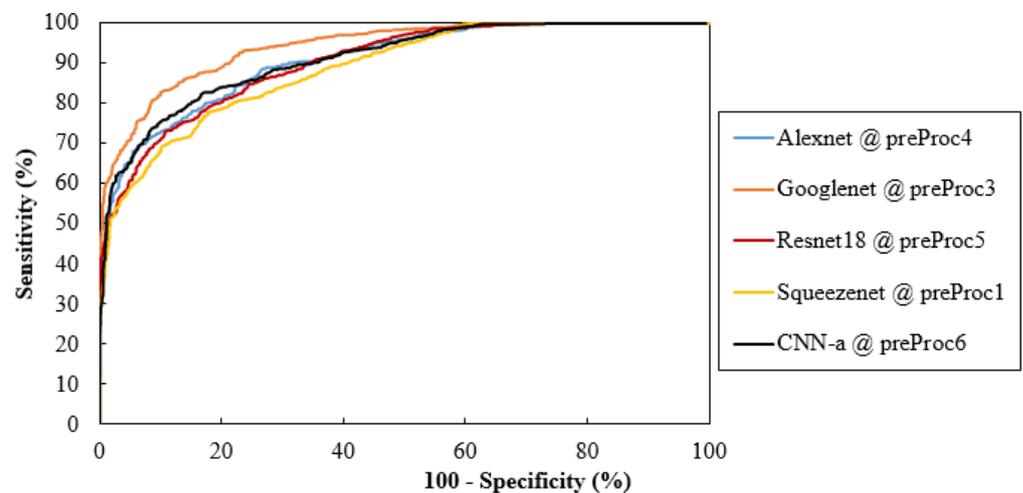

Figure 6. Comparisons of ROC curves for the CNNs and training data with the best AUC values; preProc—preProcessing.
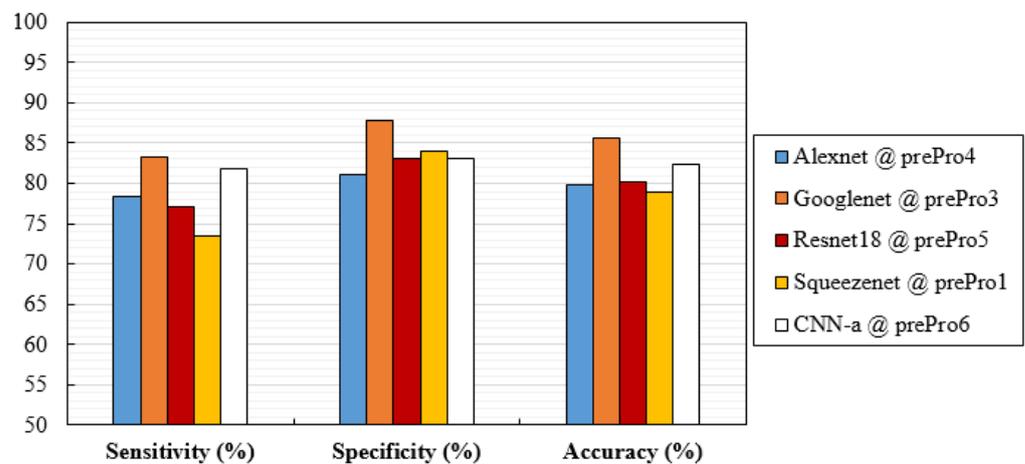
Figure 7. Values of sensitivity, specificity, and accuracy obtained with the architectures trained with preprocessed data that achieved the best mean AUC.

3.3. Influence of Breast Density on Classification

Breast density interferes with the detection of lesions [88]. In this way, it was important to explore the influence of density on the specific detection of MCs with these CNNs trained by these datasets. For this purpose, the training dataset were not changed, i.e., the CNNs were trained including all breast densities, but they were tested separately with specific datasets for each breast density. The results, in form of AUC values, are shown in Figure 8.
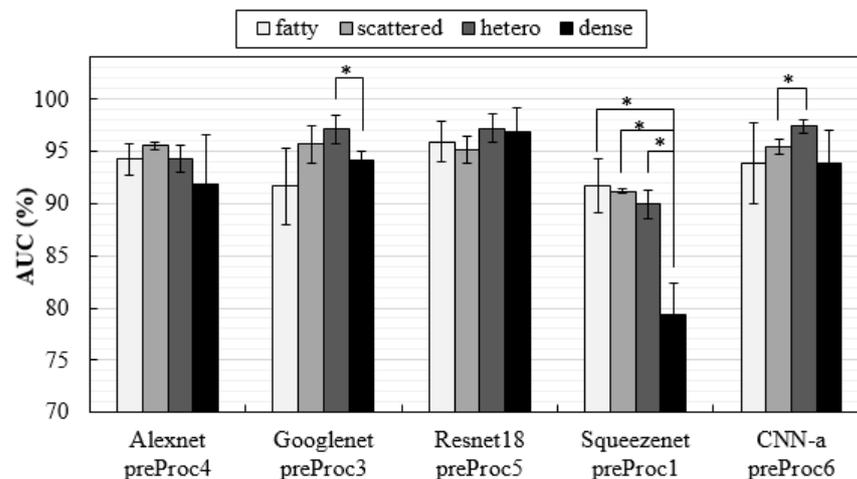
Figure 8. AUC values obtained with test datasets composed by the four different breast densities separately (* $p < 0.05$ indicates a significant difference between groups).

The training that provided the best performance (GoogLeNet @ preprocessing 3) required a training time of approximately 9 h for all three folds (using an NVIDIA Quadro P4000 GPU). On the other hand, the fastest training and second-best performance were obtained, simultaneously, for our CNN-a with data from preprocessing 6. Table 5 shows the training and inference times for all CNNs.

Table 5. Training times, in hours, needed for each CNN after threefold cross-validation and mean inference time (in seconds) needed to classify each image.

	Training Time (h)	Inference Time/Slice (s)
CNN-a	2.4	0.0057
AlexNet	4.1	0.0062
SqueezeNet	4.4	0.0083
ResNet18	7.8	0.0143
GoogLeNet	8.9	0.0158

4. Discussion

In this work, the training from scratch of four popular CNNs and a new architecture proposed by us was investigated. Given the whole DBT image (and not only some specific ROIs) as input, the classification of cases by absence or presence of MCs was the main task of these trainings. Original data and data resulting from preprocessing methods (to increase MCs visibility) were considered. The DBT dataset used for training and testing are from the public database available at The Cancer Imaging Archive website [75].

In order to avoid useless complex mathematical operations, all the information outside the breast region was eliminated. In four steps, an automatic methodology that creates a binary image where only the information inside the breast is considered was implemented. The comparison between the contaminated data and the data with complete suppression of BG signal can be observed through the profiles of the yellow regions in Figures 4a and 4f, respectively. This operation represented a difference of about 5% in training times, without performance losses, and it is usually applied in this type of CNN training.

Data preprocessing can be very useful when training CNNs from scratch to facilitate the detection and classification processes. In this work, both original data and data resulting from different preprocessing methods were considered as input. A comprehensive study of different methods to make the MCs more visible to the algorithms was carried out.

In original data, the MCs showed reasonable contrast to the naked eye (Figure 5a). This highlight can be compromised due to their size, the presence of noise, and other structures that can make them less visible. Both preprocessing 1 and preprocessing 2, had a great influence on MCs data. Preprocessing 1 smoothed the region around the MCs, preserving its edges (Figure 5b), while preprocessing 2 contributed to an increase in contrast between all structures, whether they were MCs or not (Figure 5c). We thought it might be interesting to combine a technique that is essentially for noise reduction (TV minimization) with a CLAHE technique; in this way, preprocessing 3 and preprocessing 4 corresponding to Figures 5d and 5e, respectively, were implemented. While, visually, the MCs stand out from the surrounding noise in Figure 5d, in Figure 5e, where the contrast enhancement was applied first and the noise reduction latter, the MCs appear to fade. Additionally, for its simplicity, another method based on squared normalized data was also studied (preprocessing 5). This operation worked quite well when it comes to highlighting high-intensity structures (Figure 5f). The application of the TV minimization filter to these data (preprocessing 6) also resulted in a reduction in anatomical noise that allowed for greater differentiation of the MCs, as can be seen in Figure 5g.

This descriptive analysis is in line with the numerical results obtained for the trained CNNs. From Table 3, it can be seen that the results were affected not only by the type of input data, but also the CNN architecture itself. In fact, the best AUC value of each CNN was achieved with different input data. GoogLeNet showed the best AUC with data processed using method 3 (94.19%), CNN-a showed the best AUC with data processed using method 6 (91.17%), AlexNet showed the best AUC with data processed using method 4 (90.82%), ResNet18 showed the best AUC with data processed using method 5 (90.44%), and SqueezeNet showed the best AUC with data processed using method 1 (88.78%). CNNs trained with original data did not generate a maximum AUC. However, all the AUC values were higher than 86%, showing that, even without any preprocessing, this could be an option. As shown in Figure 9a, for cases where the MCs were in a region with less noise

and were more evident, all the CNNs achieved a correct classification in the original data. On the other hand, despite the efforts to reduce noise and increase contrast, some cases such as the one in Figure 9b were incorrectly classified as negative by all CNNs, even when varying the pre-processing. Although preprocessing 2 did not contribute to a maximum either, it resulted in the third-best AUC for GoogLeNet. From Table 3, it is also possible to conclude that GoogLeNet was the most sensitive CNN to data contrast since its best results of AUC were obtained with methods where the contrast enhancement operation was performed. In the example of a case where MCs were in a region with other structures also of greater contrast (Figure 9c,d), GoogLeNet took advantage of preprocessing 3 and was the only CNN to correctly classify this case. As a matter of fact, the GoogLeNet trained with data processed using method 3 presented significantly higher values in the detection of cases with MCs (p-value < 0.05, Table 4). This superiority is quite visible in the isolated ROC curve in Figure 6. The second-best performance corresponded to CNN-a trained with data from preprocessing 6, with this superiority significant in relation to ResNet18 and SqueezeNet (Table 4). In Figure 9e there is a case of a MCs that were masked and only detected by CNN-a after preprocessing 6 (Figure 9f). Thus, in agreement with the results in Table 3, we can assume that it is the combination of both factors (data type and CNN) that determines the result of a correct classification.
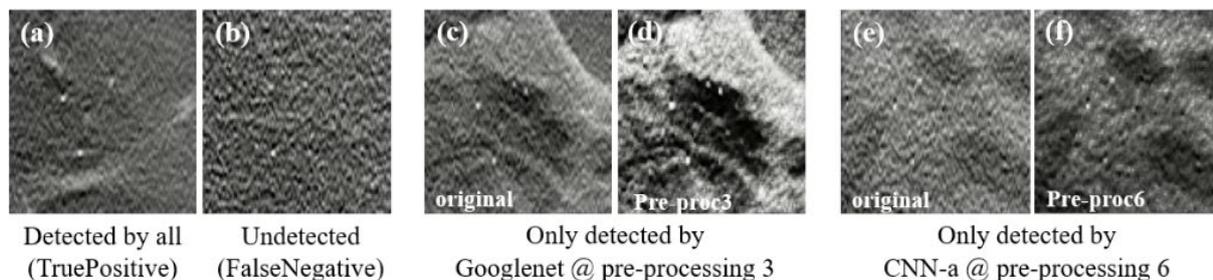
Figure 9. Some examples of MCs in the DBT data used. (**a**) True positive (case correctly classified as positive by all CNNs, even in the original image); (**b**) false negative (case incorrectly classified as negative by all CNNs, even when varying the preprocessing); (**c**) original case classified as negative and that was only detected by GoogLeNet when preprocessed with method 3 (**d**); (**e**) original case classified as negative that was only detected by CNN-a when preprocessed with method 6 (**f**).

The variations and differences in AUC values obtained for each situation were, in general, in agreement with the specificity, sensitivity, and accuracy values obtained in Figure 7. Although specificity values were higher than sensitivity in most cases, these differences were not significant (p-value > 0.05 in all cases). As for accuracy, GoogLeNet and CNN-a presented the best values of 85.68% and 82.45%, respectively.

In the VICTRE database, it is possible to separate the cases by breast density, and a study was published where a model observer was trained separately for detecting lesions in each of the four breast density types and then tested on the same density type to obtain the individual AUC for each density [89]. As a conclusion of this study, Zeng et al. believed it would be appropriate to train the model observer with mixed breast density images. This was exactly what we did with the deep learning architectures proposed in this work. However, in order to understand whether the presented methodologies were influenced or not by breast density, the same CNNs were tested separately for classifying the DBT data about the presence of MCs in each of the four breast density types (fatty, scattered, heterogeneous, and dense), and the results were analyzed in terms of AUC. As seen in Figure 8, only SqueezeNet was especially sensitive to density, showing significant differences in detection among the three density types. The correct classification of cases with MCs in dense breasts with SqueezeNet was significantly lower compared to the other densities. In general, due to the lower anatomical background, fatty breasts allowed good classifications of cases with MCs. GoogLeNet was the exception, with fatty breasts corresponding to the lowest AUC value ($p > 0.05$).

Training and inference times of Table 5 are purely indicative as they vary depending on the computation power available. However, in relative terms, the already existent networks (GoogLeNet, ResNet18, SqueezeNet, and AlexNet) led with the four longest times. On the other hand, although the CNN with the best AUC (GoogLeNet) showed the longest time, the second best (CNN-a) was the faster network. As inference time is the key when the models are used in clinic, it should be noted that, with CNN-a, it was possible to classify an image never seen by the model before about three times faster than with GoogLeNet. From our point of view, this fact makes this architecture adapted from AlexNet very interesting for future studies that involve more complex and longer trainings, such as object detection with state-of-the-art faster region-based CNNs. One of the most determining factors in the training/testing time of these CNNs is the feature extraction network that is used as the basis. Thus, a faster model such as CNN-a, which presents good results regarding the classification of cases with MCs, should be an option to be studied in the future.

In two published studies (2D and 3D), where a prescreening stage generates possible MCs and the proposed CNNs differentiate between true MCs and false positives, AUC values of 93% [50] and 97.65% [68] were reported. Both studies used ROIs instead of the whole image/volume. Some regions do not have any lesions or relevant information, while others contain only the lesions. On the other hand, in a study where the main objective was to compare the detection of MCs in images reconstructed with two different reconstruction algorithms (EMPIRE and filtered back projection), small 3D patches were used as input, and the best result obtained in terms of partial AUC was 88.0% [58].

In another study, an ROI was selected for each lesion on a DBT key slice, features were extracted using a pretrained CNN and served as input to a support vector machine classifier trained in the task of predicting likelihood of malignancy [62]. The AUC result obtained in CC view for MCs detection was 82%. Other views were included, and, considering MLO (mediolateral oblique) in addition to CC view, AUC improved to 97%, showing the importance of having both views available.

Xiao et al. proposed an interesting ensemble CNN to classify benign and malignant MCs in DBT. This classification was made on smaller patches (300 × 300) containing only the MCs. The AUC and accuracy using a decision-level ensemble strategy were 0.8837 and 0.82, respectively [70].

The only work that took the whole image information into account used 2D synthetic mammographic images obtained from DBT exams to train a multi-view deep CNN to classify screening images into BI RADS classes (0: further evaluation is required due to a suspicious abnormality; 1: the mammogram is negative; 2: the mammogram is benign). The AUC values obtained were as follows: BI-RADS 0 vs. others, 91.2%; BI-RADS 1 vs. others, 90.5%; BIRADS 2 vs. others, 90.0% [66].

A direct comparison between literature values and those obtained in this work is not fair due to several reasons. The first is that different databases were used (those of the studies mentioned were all private databases). The second is that the training data have quite different characteristics due to different detection tasks. Some used only small parts of the data, and those which used the entire image did not refer to DBT slices but rather to synthetic mammograms obtained with DBT. Nevertheless, it is possible to confirm that the results obtained by our study (maximum value of AUC achieved: 94.19%) are quite competitive when compared to those available in the literature.

There were some limitations in this study. The first is that the available dataset is limited to the CC view and one manufacturer. The second is that only one type of lesion (MCs) was considered, and, within the available data, there may be some similarities between lesions. We tried to overcome this fact through data augmentation with reflection and rotation. The third is that, despite being very realistic, the data are simulated and, therefore, do not correspond to real patients. Lastly, since DBT is a 3D technique, the fact that we consider information in 2D slices can limit the advantage provided by the depth information. Furthermore, the true clinical value lies in the classification of a volume, because this is what radiologists do every day in clinical practice. We believe that this work

is a starting point and can serve as a basis for the implementation of a 3D training with all volume and 3D architectures, considering real data volumes and not just some slices. In addition, it will also be important to diversify the lesions, including data obtained from other views (MLO), manufacturers, and reconstruction algorithms. As for the training of the CNNs themselves, other optimizers that have been producing good results (such as Adam optimizer), as well as different mini-batch sizes and learning rates, should be tested and evaluated.

5. Conclusions

Deep learning AI algorithms composed of multilayered CNNs have been growing over the past 5 years and have shown very promising results in supporting the detection of breast cancer. One of the great difficulties in training these algorithms is the lack of labeled DBT databases. Furthermore, all published studies refer to private databases, thus limiting the comparison and improvement of the studies carried out.

In this study, a public DBT dataset was used to train from scratch four popular CNNs and a new CNN model proposed by us. The main task of our algorithms was to classify a DBT case for the presence or absence of MCs, given the whole DBT image as input. In addition to the original data, six different preprocessing methodologies, the main purpose of which was to highlight MCs, were implemented to generate different input datasets.

Classifying the whole image according to the presence or absence of MCs is a difficult task due to the size of MCs and all the information present in an entire image. With this work, we were able to achieve encouraging outcomes in this regard, obtaining similar results to other studies for the detection of larger lesions such as masses. The classification of cases with/without MCs was greatly influenced by the type of input data, and our new model achieved the second-best performance in the shortest time, thus becoming a very interesting model to be considered in future studies.

Author Contributions: All the authors substantially contributed to this paper. Formal analysis, A.M.M.; methodology, A.M.M.; supervision, M.J.C., P.A. and N.M.; writing—original draft, A.M.M.; writing—review and editing, M.J.C. and N.M. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Ethical review and approval were waived for this study since data was obtained from publicly accessible repository.

Informed Consent Statement: Patient consent was waived since data was obtained from publicly accessible repository.

Data Availability Statement: Data from VICTRE project was used. Available at https://wiki.cancerimagingarchive.net/display/Public/The+VICTRE+Trial%3A+Open-Source%2C+In-Silico+Clinical+Trial+For+Evaluating+Digital+Breast+Tomosynthesis.

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

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2021. *CA Cancer J. Clin.* **2021**, *71*, 7–33. [CrossRef] [PubMed]
3. Tabár, L.; Dean, P.B.; Chen, T.H.H.; Yen, A.M.F.; Chen, S.L.S.; Fann, J.C.Y.; Chiu, S.Y.H.; Ku, M.M.S.; Wu, W.Y.Y.; Hsu, C.Y.; et al. The incidence of fatal breast cancer measures the increased effectiveness of therapy in women participating in mammography screening. *Cancer* **2019**, *125*, 515–523. [CrossRef] [PubMed]
4. Skaane, P.; Sebuødegård, S.; Bandos, A.I.; Gur, D.; Østerås, B.H.; Gullien, R.; Hofvind, S. Performance of breast cancer screening using digital breast tomosynthesis: Results from the prospective population-based Oslo Tomosynthesis Screening Trial. *Breast Cancer Res. Treat.* **2018**, *169*, 489–496. [CrossRef] [PubMed]

5. Ciatto, S.; Houssami, N.; Bernardi, D.; Caumo, F.; Pellegrini, M.; Brunelli, S.; Tuttobene, P.; Bricolo, P.; Fantò, C.; Valentini, M.; et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): A prospective comparison study. *Lancet Oncol.* **2013**, *14*, 583–589. [CrossRef]

6. Haas, B.M.; Kalra, V.; Geisel, J.; Raghu, M.; Durand, M.; Philpotts, L.E. Comparison of Tomosynthesis Plus Digital Mammography and Digital Mammography Alone for Breast Cancer Screening. *Radiology* **2013**, *269*, 694–700. [CrossRef]

7. Rose, S.L.; Tidwell, A.L.; Bujnoch, L.J.; Kushwaha, A.C.; Nordmann, A.S.; Sexton Jr, R. Implementation of Breast Tomosynthesis in a Routine Screening Practice: An Observational Study. *Am. J. Roentgenol.* **2013**, *200*, 1401–1408. [CrossRef]

8. Greenberg, J.S.; Javitt, M.C.; Katzen, J.; Michael, S.; Holland, A.E. Clinical Performance Metrics of 3D Digital Breast Tomosynthesis Compared With 2D Digital Mammography for Breast Cancer Screening in Community Practice. *Am. J. Roentgenol.* **2014**, *203*, 687–693. [CrossRef]

9. McDonald, E.S.; Oustimov, A.; Weinstein, S.P.; Synnestvedt, M.B.; Schnall, M.; Conant, E.F. Effectiveness of Digital Breast Tomosynthesis Compared With Digital Mammography: Outcomes Analysis From 3 Years of Breast Cancer Screening. *JAMA Oncol.* **2016**, *2*, 737–743. [CrossRef]

10. Zackrisson, S.; Lång, K.; Rosso, A.; Johnson, K.; Dustler, M.; Förnvik, D.; Förnvik, H.; Sartor, H.; Timberg, P.; Tingberg, A.; et al. One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): A prospective, population-based, diagnostic accuracy study. *Lancet Oncol.* **2018**, *19*, 1493–1503. [CrossRef]

11. Bernardi, D.; Macaskill, P.; Pellegrini, M.; Valentini, M.; Fantò, C.; Ostillio, L.; Tuttobene, P.; Luparia, A.; Houssami, N. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): A population-based prospective study. *Lancet Oncol.* **2016**, *17*, 1105–1113. [CrossRef]

12. Lång, K.; Andersson, I.; Rosso, A.; Tingberg, A.; Timberg, P.; Zackrisson, S. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: Results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study. *Eur. Radiol.* **2016**, *26*, 184–190. [CrossRef] [PubMed]

13. Gilbert, F.J.; Tucker, L.; Gillan, M.G.; Willsher, P.; Cooke, J.; Duncan, K.A.; Michell, M.J.; Dobson, H.M.; Lim, Y.Y.; Suaris, T.; et al. Accuracy of Digital Breast Tomosynthesis for Depicting Breast Cancer Subgroups in a UK Retrospective Reading Study (TOMMY Trial). *Radiology* **2015**, *277*, 697–706. [CrossRef] [PubMed]

14. Hofvind, S.; Hovda, T.; Holen, Å.S.; Lee, C.I.; Albertsen, J.; Bjørndal, H.; Brandal, S.H.; Gullien, R.; Lømo, J.; Park, D.; et al. Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Digital Mammography: Evaluation in a Population-based Screening Program. *Radiology* **2018**, *287*, 787–794. [CrossRef]

15. Freer, P.E.; Riegert, J.; Eisenmenger, L.; Ose, D.; Winkler, N.; Stein, M.A.; Stoddard, G.J.; Hess, R. Clinical implementation of synthesized mammography with digital breast tomosynthesis in a routine clinical practice. *Breast Cancer Res. Treat.* **2017**, *166*, 501–509. [CrossRef]

16. Skaane, P.; Bandos, A.I.; Gullien, R.; Eben, E.B.; Ekseth, U.; Haakenaasen, U.; Izadi, M.; Jebsen, I.N.; Jahr, G.; Krager, M.; et al. Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosynthesis in a Population-based Screening Program. *Radiology* **2013**, *267*, 47–56. [CrossRef]

17. Tagliafico, A.S.; Calabrese, M.; Bignotti, B.; Signori, A.; Fisci, E.; Rossi, F.; Valdora, F.; Houssami, N. Accuracy and reading time for six strategies using digital breast tomosynthesis in women with mammographically negative dense breasts. *Eur. Radiol.* **2017**, *27*, 5179–5184. [CrossRef]

18. Balleyguier, C.; Arfi-Rouche, J.; Levy, L.; Toubiana, P.R.; Cohen-Scali, F.; Toledano, A.Y.; Boyer, B. Improving digital breast tomosynthesis reading time: A pilot multi-reader, multi-case study using concurrent Computer-Aided Detection (CAD). *Eur. J. Radiol.* **2017**, *97*, 83–89. [CrossRef]

19. Benedikt, R.A.; Boatsman, J.E.; Swann, C.A.; Kirkpatrick, A.D.; Toledano, A.Y. Concurrent Computer-Aided Detection Improves Reading Time of Digital Breast Tomosynthesis and Maintains Interpretation Performance in a Multireader Multicase Study. *Am. J. Roentgenol.* **2017**, *210*, 685–694. [CrossRef]

20. Chae, E.Y.; Kim, H.H.; Jeong, J.W.; Chae, S.H.; Lee, S.; Choi, Y.W. Decrease in interpretation time for both novice and experienced readers using a concurrent computer-aided detection system for digital breast tomosynthesis. *Eur. Radiol.* **2019**, *29*, 2518–2525. [CrossRef]

21. Poplack, S.P.; Tosteson, T.D.; Kogel, C.A.; Nagy, H.M. Digital breast tomosynthesis: Initial experience in 98 women with abnormal digital screening mammography. *AJR Am. J. Roentgenol.* **2007**, *189*, 616–623. [CrossRef] [PubMed]

22. Andersson, I.; Ikeda, D.M.; Zackrisson, S.; Ruschin, M.; Svahn, T.; Timberg, P.; Tingberg, A. Breast tomosynthesis and digital mammography: A comparison of breast cancer visibility and BIRADS classification in a population of cancers with subtle mammographic findings. *Eur. Radiol.* **2008**, *18*, 2817–2825. [CrossRef] [PubMed]

23. Spangler, M.L.; Zuley, M.L.; Sumkin, J.H.; Abrams, G.; Ganott, M.A.; Hakim, C.; Perrin, R.; Chough, D.M.; Shah, R.; Gur, D. Detection and Classification of Calcifications on Digital Breast Tomosynthesis and 2D Digital Mammography: A Comparison. *Am. J. Roentgenol.* **2011**, *196*, 320–324. [CrossRef] [PubMed]

24. Kopans, D.; Gavenonis, S.; Halpern, E.; Moore, R. Calcifications in the breast and digital breast tomosynthesis. *Breast. J.* **2011**, *17*, 638–644. [CrossRef] [PubMed]

25. Svane, G.; Azavedo, E.; Lindman, K.; Urech, M.; Nilsson, J.; Weber, N.; Lindqvist, L.; Ullberg, C. Clinical experience of photon counting breast tomosynthesis: Comparison with traditional mammography. *Acta Radiol.* **2011**, *52*, 134–142. [CrossRef] [PubMed]

26. Wallis, M.G.; Moa, E.; Zanca, F.; Leifland, K.; Danielsson, M. Two-View and Single-View Tomosynthesis versus Full-Field Digital Mammography: High-Resolution X-Ray Imaging Observer Study. *Radiology* **2012**, *262*, 788–796. [CrossRef]

27. Nyante, S.J.; Lee, S.S.; Benefield, T.S.; Hoots, T.N.; Henderson, L.M. The association between mammographic calcifications and breast cancer prognostic factors in a population-based registry cohort. *Cancer* **2017**, *123*, 219–227. [CrossRef]

28. D'Orsi, C.J. *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*; American College of Radiology: Reston, VA, USA, 2013.

29. Samala, R.K.; Chan, H.P.; Lu, Y.; Hadjiiski, L.M.; Wei, J.; Helvie, M.A. Digital breast tomosynthesis: Computer-aided detection of clustered microcalcifications on planar projection images. *Phys. Med. Biol.* **2014**, *59*, 7457–7477. [CrossRef]

30. Samala, R.K.; Chan, H.P.; Hadjiiski, L.M.; Helvie, M.A. Analysis of computer-aided detection techniques and signal characteristics for clustered microcalcifications on digital mammography and digital breast tomosynthesis. *Phys. Med. Biol.* **2016**, *61*, 7092–7112. [CrossRef]

31. Park, S.C.; Zheng, B.; Wang, X.H.; Gur, D. Applying a 2D Based CAD Scheme for Detecting Micro-Calcification Clusters Using Digital Breast Tomosynthesis Images: An Assessment. In *Medical Imaging 2008: Computer-Aided Diagnosis*; Medical Imaging: San Diego, CA, USA, 2008; Volume 6915, pp. 70–77.

32. Reiser, I.; Nishikawa, R.M.; Edwards, A.V.; Kopans, D.B.; Schmidt, R.A.; Papaioannou, J.; Moore, R.H. Automated detection of microcalcification clusters for digital breast tomosynthesis using projection data only: A preliminary study. *Med. Phys.* **2008**, *35*, 1486–1493. [CrossRef]

33. Bernard, S.; Muller, S.; Onativia, J. Computer-Aided Microcalcification Detection on Digital Breast Tomosynthesis Data: A Preliminary Evaluation. In *Digital Mammography: 9th International Workshop*; Krupinski, E.A., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 151–157.

34. Sahiner, B.; Chan, H.P.; Hadjiiski, L.M.; Helvie, M.A.; Wei, J.; Zhou, C.; Lu, Y. Computer-aided detection of clustered microcalcifications in digital breast tomosynthesis: A 3D approach. *Med. Phys.* **2012**, *39*, 28–39. [CrossRef] [PubMed]

35. Samala, R.K.; Chan, H.P.; Lu, Y.; Hadjiiski, L.; Wei, J.; Sahiner, B.; Helvie, M.A. Computer-aided detection of clustered microcalcifications in multiscale bilateral filtering regularized reconstructed digital breast tomosynthesis volume. *Med. Phys.* **2014**, *41*, 021901. [CrossRef] [PubMed]

36. Wei, J.; Chan, H.P.; Hadjiiski, L.M.; Helvie, M.A.; Lu, Y.; Zhou, C.; Samala, R. Multichannel response analysis on 2D projection views for detection of clustered microcalcifications in digital breast tomosynthesis. *Med. Phys.* **2014**, *41*, 041913. [CrossRef] [PubMed]

37. Samala, R.K.; Chan, H.P.; Lu, Y.; Hadjiiski, L.M.; Wei, J.; Helvie, M.A. Computer-aided detection system for clustered microcalcifications in digital breast tomosynthesis using joint information from volumetric and planar projection images. *Phys. Med. Biol.* **2015**, *60*, 8457–8479. [CrossRef]

38. Fenton, J.J.; Taplin, S.H.; Carney, P.A.; Abraham, L.; Sickles, E.A.; D'Orsi, C.; Berns, E.A.; Cutter, G.; Hendrick, R.E.; Barlow, W.E.; et al. Influence of Computer-Aided Detection on Performance of Screening Mammography. *N. Engl. J. Med.* **2007**, *356*, 1399–1409. [CrossRef]

39. Lehman, C.D.; Wellman, R.D.; Buist, D.S.M.; Kerlikowske, K.; Tosteson, A.N.A.; Miglioretti, D.L.; Breast Cancer Surveillance Consortium. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Intern. Med.* **2015**, *175*, 1828–1837. [CrossRef]

40. Katzen, J.; Dodelzon, K. A review of computer aided detection in mammography. *Clin. Imaging.* **2018**, *52*, 305–309. [CrossRef]

41. Sechopoulos, I.; Teuwen, J.; Mann, R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Semin. Cancer Biol.* **2020**, *72*, 214–225. [CrossRef]

42. Rodriguez-Ruiz, A.; Wellman, R.D.; Buist, D.S.; Kerlikowske, K.; Tosteson, A.N.; Miglioretti, D.L.; Breast Cancer Surveillance Consortium. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *J. Natl. Cancer Inst.* **2019**, *111*, 916–922. [CrossRef]

43. McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.S.; Darzi, A.; et al. International evaluation of an AI system for breast cancer screening. *Nature* **2020**, *577*, 89–94. [CrossRef]

44. Kim, H.-E.; Kim, H.H.; Han, B.K.; Kim, K.H.; Han, K.; Nam, H.; Lee, E.H.; Kim, E.K. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: A retrospective, multireader study. *Lancet Digit. Health* **2020**, *2*, e138–e148. [CrossRef]

45. Wang, X.; Liang, G.; Zhang, Y.; Blanton, H.; Bessinger, Z.; Jacobs, N. Inconsistent Performance of Deep Learning Models on Mammogram Classification. *J. Am. Coll. Radiol.* **2020**, *17*, 796–803. [CrossRef]

46. Schaffter, T.; Buist, D.S.; Lee, C.I.; Nikulin, Y.; Ribli, D.; Guan, Y.; Lotter, W.; Jie, Z.; Du, H.; Wang, S.; et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Netw. Open* **2020**, *3*, e200265. [CrossRef] [PubMed]

47. Rodríguez-Ruiz, A.; Krupinski, E.; Mordang, J.J.; Schilling, K.; Heywang-Köbrunner, S.H.; Sechopoulos, I.; Mann, R.M. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. *Radiology* **2019**, *290*, 305–314. [CrossRef] [PubMed]

48. Conant, E.F.; Toledano, A.Y.; Periaswamy, S.; Fotin, S.V.; Go, J.; Boatsman, J.E.; Hoffmeister, J.W. Improving Accuracy and Efficiency with Concurrent Use of Artificial Intelligence for Digital Breast Tomosynthesis. *Radiol. Artif. Intell.* **2019**, *1*, e180096. [CrossRef]

49. van Winkel, S.L.; Rodríguez-Ruiz, A.; Appelman, L.; Gubern-Mérida, A.; Karssemeijer, N.; Teuwen, J.; Wanders, A.J.; Sechopoulos, I.; Mann, R.M. Impact of artificial intelligence support on accuracy and reading time in breast tomosynthesis image interpretation: A multi-reader multi-case study. *Eur. Radiol.* **2021**, *31*, 8682–8691. [CrossRef]

50. Samala, R.; Chan, H.P.; Hadjiiski, L.M.; Cha, K.; Helvie, M.A. Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis. In *Medical Imaging 2016: Computer-Aided Diagnosis*; SPIE Medical Imaging: San Diego, CA, USA, 2016; Volume 9785.

51. Fotin, S.; Yin, Y.; Haldankar, H.; Hoffmeister, J.W.; Periaswamy, S. Detection of soft tissue densities from digital breast tomosynthesis: Comparison of conventional and deep learning approaches. In *Medical Imaging 2016: Computer-Aided Diagnosis*; SPIE Medical Imaging: San Diego, CA, USA, 2016; Volume 9785.

52. Kim, D.H.; Kim, S.T.; Ro, Y.M. Latent feature representation with 3-D multi-view deep convolutional neural network for bilateral analysis in digital breast tomosynthesis. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016.

53. Samala, R.K.; Chan, H.P.; Hadjiiski, L.; Helvie, M.A.; Wei, J.; Cha, K. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Med. Phys.* **2016**, *43*, 6654. [CrossRef]

54. Kim, D.H.; Kim, S.T.; Chang, J.M.; Ro, Y.M. Latent feature representation with depth directional long-term recurrent learning for breast masses in digital breast tomosynthesis. *Phys. Med Biol.* **2017**, *62*, 1009–1031. [CrossRef]

55. Zhang, X.; Zhang, Y.; Han, E.Y.; Jacobs, N.; Han, Q.; Wang, X.; Liu, J. Classification of Whole Mammogram and Tomosynthesis Images Using Deep Convolutional Neural Networks. *IEEE Trans. NanoBioscience* **2018**, *17*, 237–242. [CrossRef]

56. Samala, R.K.; Chan, H.P.; Hadjiiski, L.M.; Helvie, M.A.; Richter, C.; Cha, K. Evolutionary pruning of transfer learned deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis. *Phys. Med. Biol.* **2018**, *63*, 095005. [CrossRef]

57. Yousefi, M.; Krzyżak, A.; Suen, C.Y. Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. *Comput. Biol. Med.* **2018**, *96*, 283–293. [CrossRef] [PubMed]

58. Rodriguez-Ruiz, A.; Teuwen, J.; Vreemann, S.; Bouwman, R.W.; van Engen, R.E.; Karssemeijer, N.; Mann, R.M.; Gubern-Merida, A.; Sechopoulos, I. New reconstruction algorithm for digital breast tomosynthesis: Better image quality for humans and computers. *Acta Radiol.* **2018**, *59*, 1051–1059. [CrossRef] [PubMed]

59. Mordang, J.J.; Janssen, T.; Bria, A.; Kooi, T.; Gubern-Mérida, A.; Karssemeijer, N. Automatic Microcalcification Detection in Multi-vendor Mammography Using Convolutional Neural Networks. In *International Workshop on Breast Imaging*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016.

60. Zhang, Y.; Wang, X.; Blanton, H.; Liang, G.; Xing, X.; Jacobs, N. 2D Convolutional Neural Networks for 3D Digital Breast Tomosynthesis Classification. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019.

61. Liang, G.; Wang, X.; Zhang, Y.; Xing, X.; Blanton, H.; Salem, T.; Jacobs, N. Joint 2D-3D Breast Cancer Classification. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019.

62. Mendel, K.; Li, H.; Sheth, D.; Giger, M. Transfer Learning From Convolutional Neural Networks for Computer-Aided Diagnosis: A Comparison of Digital Breast Tomosynthesis and Full-Field Digital Mammography. *Acad. Radiol.* **2019**, *26*, 735–743. [CrossRef] [PubMed]

63. Singh, S.; Matthews, T.P.; Shah, M.; Mombourquette, B.; Tsue, T.; Long, A.; Almohsen, R.; Pedemonte, S.; Su, J. Adaptation of a deep learning malignancy model from full-field digital mammography to digital breast tomosynthesis. *arXiv* **2020**, arXiv:2001.08381.

64. Li, X.; Qin, G.; He, Q.; Sun, L.; Zeng, H.; He, Z.; Chen, W.; Zhen, X.; Zhou, L. Digital breast tomosynthesis versus digital mammography: Integration of image modalities enhances deep learning-based breast mass classification. *Eur. Radiol.* **2020**, *30*, 778–788. [CrossRef]

65. Wang, L.; Zheng, C.; Chen, W.; He, Q.; Li, X.; Zhang, S.; Qin, G.; Chen, W.; Wei, J.; Xie, P.; et al. Multi-path synergic fusion deep neural network framework for breast mass classification using digital breast tomosynthesis. *Phys. Med. Biol.* **2020**, *65*, 235045. [CrossRef]

66. Seyyedi, S.; Wong, M.J.; Ikeda, D.M.; Langlotz, C.P. SCREENet: A Multi-view Deep Convolutional Neural Network for Classification of High-resolution Synthetic Mammographic Screening Scans. *arXiv* **2020**, arXiv:abs/2009.08563.

67. Matthews, T.P.; Singh, S.; Mombourquette, B.; Su, J.; Shah, M.P.; Pedemonte, S.; Long, A.; Maffit, D.; Gurney, J.; Hoil, R.M.; et al. A Multisite Study of a Breast Density Deep Learning Model for Full-Field Digital Mammography and Synthetic Mammography. *Radiol. Artif. Intell.* **2021**, *3*, e200015. [CrossRef]

68. Zheng, J.; Sun, H.; Wu, S.; Jiang, K.; Peng, Y.; Yang, X.; Zhang, F.; Li, M. 3D Context-Aware Convolutional Neural Network for False Positive Reduction in Clustered Microcalcifications Detection. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 764–773. [CrossRef]

69. Aswiga, R.V.; Shanthi, A.P. Augmenting Transfer Learning with Feature Extraction Techniques for Limited Breast Imaging Datasets. *J. Digit. Imaging* **2021**, *34*, 618–629.

70. Xiao, B.; Sun, H.; Meng, Y.; Peng, Y.; Yang, X.; Chen, S.; Yan, Z.; Zheng, J. Classification of microcalcification clusters in digital breast tomosynthesis using ensemble convolutional neural network. *Biomed. Eng. Online* **2021**, *20*, 71. [CrossRef] [PubMed]

71. El-Shazli, A.M.A.; Youssef, S.M.; Soliman, A.H. Intelligent Computer-Aided Model for Efficient Diagnosis of Digital Breast Tomosynthesis 3D Imaging Using Deep Learning. *Appl. Sci.* **2022**, *12*, 5736. [CrossRef]

72. Bai, J.; Posner, R.; Wang, T.; Yang, C.; Nabavi, S. Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review. *Med. Image Anal.* **2021**, *71*, 102049. [CrossRef] [PubMed]

73. Buda, M.; Saha, A.; Walsh, R.; Ghate, S.; Li, N.; Święcicki, A.; Lo, J.Y.; Mazurowski, M.A. Detection of masses and architectural distortions in digital breast tomosynthesis: A publicly available dataset of 5060 patients and a deep learning model. *arXiv* **2020**, arXiv:2011.07995.

74. Badano, A.; Graff, C.G.; Badal, A.; Sharma, D.; Zeng, R.; Samuelson, F.W.; Glick, S.J.; Myers, K.J. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. *JAMA Netw. Open* **2018**, *1*, e185474. [CrossRef]

75. VICTRE. The VICTRE Trial: Open-Source, In-Silico Clinical Trial For Evaluating Digital Breast Tomosynthesis. 2018. Available online: https://wiki.cancerimagingarchive.net/display/Public/The+VICTRE+Trial%3A+Open-Source%2C+In-Silico+Clinical+Trial+For+Evaluating+Digital+Breast+Tomosynthesis (accessed on 1 November 2021).

76. Sidky, E.Y.; Pan, X.; Reiser, I.S.; Nishikawa, R.M.; Moore, R.H.; Kopans, D.B. Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms. *Med. Phys.* **2009**, *36*, 4920–4932. [CrossRef]

77. Lu, Y.; Chan, H.P.; Wei, J.; Hadjiiski, L.M. Selective-diffusion regularization for enhancement of microcalcifications in digital breast tomosynthesis reconstruction. *Med. Phys.* **2010**, *37*, 6003–6014. [CrossRef]

78. Mota, A.M.; Matela, N.; Oliveira, N.; Almeida, P. Total variation minimization filter for DBT imaging. *Med. Phys.* **2015**, *42*, 2827–2836. [CrossRef]

79. Michielsen, K.; Nuyts, J.; Cockmartin, L.; Marshall, N.; Bosmans, H. Design of a model observer to evaluate calcification detectability in breast tomosynthesis and application to smoothing prior optimization. *Med. Phys.* **2016**, *43*, 6577–6587. [CrossRef]

80. Mota, A.M.; Clarkson, M.J.; Almeida, P.; Matela, N. An Enhanced Visualization of DBT Imaging Using Blind Deconvolution and Total Variation Minimization Regularization. *IEEE Trans. Med. Imaging* **2020**, *39*, 4094–4101. [CrossRef]

81. Zuiderveld, K. Contrast limited adaptive histogram equalization. In *Graphics Gems IV1994*; Academic Press Professional, Inc.: San Diego, CA, USA, 1994; pp. 474–485.

82. MathWorks. MATLAB Adapthisteq Function. [Cited 2021 May]. 2021. Available online: https://www.mathworks.com/help/images/ref/adapthisteq.html (accessed on 1 November 2021).

83. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Processing Syst.* **2012**, *25*, 1097–1105. [CrossRef]

84. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.

85. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

86. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.

87. MathWorks. Tranfer Learning. [Cited May 2021]. 2021. Available online: https://www.mathworks.com/discovery/transfer-learning.html (accessed on 1 November 2021).

88. Vourtsis, A.; Berg, W.A. Breast density implications and supplemental screening. *Eur. Radiol.* **2019**, *29*, 1762–1777. [CrossRef] [PubMed]

89. Zeng, R.; Samuelson, F.W.; Sharma, D.; Badal, A.; Christian, G.G.; Glick, S.J.; Myers, K.J.; Badano, A.G. Computational reader design and statistical performance evaluation of an in-silico imaging clinical trial comparing digital breast tomosynthesis with full-field digital mammography. *J. Med. Imaging* **2020**, *7*, 042802. [CrossRef] [PubMed]