Check for updates

Received: 8 February 2023 | Revised: 17 May 2023 | Accepted: 12 June 2023

DOI: 10.1002/mp.16589

RESEARCH ARTICLE

MEDICAL PHYSICS

A deep learning approach to estimate x-ray scatter in digital breast tomosynthesis: From phantom models to clinical applications

Marta C. Pinto[1] | **Franziska Mauter**[1,2] | **Koen Michielsen**[1] | **Ramyar Biniazan**[3] | **Steffen Kappler**[3] | **Ioannis Sechopoulos**[1,4,5]

[1]Dept. of Medical Imaging, Radboud University Medical Center, Nijmegen, The Netherlands

[2]Div. of Ionizing radiation, Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany

[3]Siemens Healthcare GmbH, Forchheim, Germany

[4]Dutch Expert Centre for Screening (LRCB), Nijmegen, The Netherlands

[5]Technical Medicine Centre, University of Twente, Enschede, The Netherlands

Correspondence
Ioannis Sechopoulos, Radboud University Medical Centrum, Postbus 9101, 6500 HB Nijmegen (766) Geert Grooteplein 10 (route 767), The Netherlands.
Email: ioannis.sechopoulos@radboudumc.nl

Abstract

Background: Digital breast tomosynthesis (DBT) has gained popularity as breast imaging modality due to its pseudo-3D reconstruction and improved accuracy compared to digital mammography. However, DBT faces challenges in image quality and quantitative accuracy due to scatter radiation. Recent advancements in deep learning (DL) have shown promise in using fast convolutional neural networks for scatter correction, achieving comparable results to Monte Carlo (MC) simulations.

Purpose: To predict the scatter radiation signal in DBT projections within clinically-acceptable times and using only clinically-available data, such as compressed breast thickness and acquisition angle.

Methods: MC simulations to obtain scatter estimates were generated from two types of digital breast phantoms. One set consisted of 600 realistically-shaped homogeneous breast phantoms for initial DL training. The other set was composed of 80 anthropomorphic phantoms, containing realistic internal tissue texture, aimed at fine tuning the DL model for clinical applications. The MC simulations generated scatter and primary maps per projection angle for a wide-angle DBT system. Both datasets were used to train (using 7680 projections from homogeneous phantoms), validate (using 960 and 192 projections from the homogeneous and anthropomorphic phantoms, respectively), and test (using 960 and 48 projections from the homogeneous and anthropomorphic phantoms, respectively) the DL model. The DL output was compared to the corresponding MC ground truth using both quantitative and qualitative metrics, such as mean relative and mean absolute relative differences (MRD and MARD), and to previously-published scatter-to-primary (SPR) ratios for similar breast phantoms. The scatter corrected DBT reconstructions were evaluated by analyzing the obtained linear attenuation values and by visual assessment of corrected projections in a clinical dataset. The time required for training and prediction per projection, as well as the time it takes to produce scatter-corrected projection images, were also tracked.

Results: The quantitative comparison between DL scatter predictions and MC simulations showed a median MRD of 0.05% (interquartile range (IQR), −0.04% to 0.13%) and a median MARD of 1.32% (IQR, 0.98% to 1.85%) for homogeneous phantom projections and a median MRD of −0.21% (IQR, −0.35% to

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

−0.07%) and a median MARD of 1.43% (IQR, 1.32% to 1.66%) for the anthropomorphic phantoms. The SPRs for different breast thicknesses and at different projection angles were within ± 15% of the previously-published ranges. The visual assessment showed good prediction capabilities of the DL model with a close match between MC and DL scatter estimates, as well as between DL-based scatter corrected and anti-scatter grid corrected cases. The scatter correction improved the accuracy of the reconstructed linear attenuation of adipose tissue, reducing the error from −16% and −11% to −2.3% and 4.4% for an anthropomorphic digital phantom and clinical case with similar breast thickness, respectively. The DL model training took 40 min and prediction of a single projection took less than 0.01 s. Generating scatter corrected images took 0.03 s per projection for clinical exams and 0.16 s for one entire projection set.
Conclusions: This DL-based method for estimating the scatter signal in DBT projections is fast and accurate, paving the way for future quantitative applications.

KEYWORDS
breast tomosynthesis, deep learning, mammography, Monte Carlo, x-ray scatter correction

1 | INTRODUCTION

Digital breast tomosynthesis (DBT) is a fast-growing technology that has seen its deployment spreading to diagnostic and screening settings for breast cancer care.[1–3] When compared to digital mammography (DM), its 2D counterpart, DBT has shown improvements in breast cancer detection, mainly due to a pseudo-3D reconstruction of the breast.[1,4,5] Through a series of low-dose projection images that are acquired at different x-ray source angles with respect to the breast, DBT images reduce the superimposition of different breast structures, offering a superior detection accuracy for breast cancer. On the other hand, this acquisition strategy also comes with intrinsic challenges that deteriorate its image quality. For instance, while in DM the number of scattered photons reaching the detector can be reduced by adding a physical anti-scatter grid between the breast and the detector, for DBT systems the use of such a method is impractical. Primarily, due to its commonly implemented acquisition geometry in which the detector remains static while the x-ray tube rotates in an arc. Therefore, scatter radiation is a known contributing factor to image quality degradation in DBT exams,[6–9] impacting lesion conspicuity, possibly contributing to the lower sensitivity of calcifications,[10] creating cupping artifacts, and reducing image contrast.[6]

For that reason, there has been an interest in developing scatter correction approaches for DBT. Most commonly, this involves the use of Monte Carlo (MC) simulations to characterize and define the scatter radiation in DBT projections. However, the computation times of MC-based techniques are, usually, too long to be clinically feasible. Different approaches have since been proposed by multiple studies showing alternatives to shorten these MC calculations.[8,11,12]

In the work from Diaz et al. a fast kernel-based method was developed that can compute scatter field radiation images in approximately 80 min, while taking into account the scatter radiation from the compression and breast support plate, and achieving a 10% error across most of the breast area when compared with direct MC simulations.[8] Feng et al. created a precomputed library of various x-ray scatter maps that can be used for the correction of clinical cases in a matter of seconds.[11] In the case of Lu et al., their method relies on interpolating the measured scatter radiation in a region outside of the breast into a scatter estimate within the breast.[12] Their approach is patient specific and can be obtained in about 5 min but can be accelerated to under a minute. However, either the long times required to obtain a scatter field image is not appropriate for clinical implementation, or the use of simplistic breast shape phantoms requires the deformation of the precomputed scatter maps in order for it to be used on a particular clinical case. In addition, sensitivity to distortions in the open field can yield larger errors in scatter estimates.

Current advances in deep-learning (DL) applied to medical imaging have shown that it is possible to overcome the trade-off between computation time and scatter estimate accuracy. Particularly, some studies have already shown promising results in the implementation of fast convolutional neural networks for scatter estimation and correction methods in x-ray breast imaging techniques,[13–15] achieving comparable performances to that of MC simulations. The present work focuses thus on the prediction of the scatter radiation signal, within clinically acceptable times and with only using clinically-available data (such as compressed breast thickness and acquisition angle), by means of a DL model that is trained on MC simulations with a realistic breast model. We demonstrate that the scatter signal correction can be applied to DBT projections by

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

testing our DL model on DBT images created from realistic heterogeneous digital breast phantoms, as well as on clinical DBT cases. The proposed scatter estimation method can naturally be applied to DM images acquired without anti-scatter grid as well.

2 | MATERIALS AND METHODS

To estimate the scatter signal for clinical DBT cases, first, a DL model was trained based on scatter estimates created through MC simulations. These simulations involved two types of digital phantoms: one generated with a realistic breast shape model for the inclusion of outer breast shape variations into initial DL training, and a second set of anthropomorphic phantoms, aimed at fine-tuning the DL model to include different tissue distributions for clinical application. Both datasets were pre-processed so that these DL inputs could closely match the clinical DBT cases and scaled to facilitate DL model training. Finally, the processed datasets were input into the DL model for training, validation, and testing and the output was compared with the corresponding MC ground truth. The final model was then applied to physical phantoms and clinical cases. A full description of the breast phantoms, MC simulations, and DL model is provided here, together with a description of the metrics used to evaluate the model's performance quantitatively and qualitatively.

2.1 | Datasets

2.1.1 | Digital breast phantoms

Realistically-shaped digital breast phantoms were generated based on a publicly available patient-based model obtained by scanning the external 3D shape of the breast during patient DBT CC-view acquisitions with structured light scanning technology.[16] This is a model that refines and extends previously published work,[17–19] and with which it is possible to generate digital phantoms with different compressed thicknesses, outer breast curvatures, and at different positions in the detector. Using this model, we created 600 randomly-shaped homogeneous phantoms, with compressed breast thicknesses uniformly distributed over 6 thickness groups of 10 mm range each, between 30 and 89 mm. The digital phantom voxel size was set to 1 mm in all directions.

Additionally, 80 anthropomorphic digital breast phantoms were obtained from the work of García et al.[20] to be used as an extra dataset of digital phantom cases with realistic internal glandular and adipose texture to further fine-tune our model to be applied to a clinical case. These digital phantoms were generated by virtual mechanical compression of patient breast computed tomography scans. The digital phantom voxel size in this case was 237 μm in all directions and the included breast thicknesses ranged from 35 to 82 mm in varying proportions among the 6 thickness groups previously described.

2.1.2 | Clinical dataset

To compare our proposed DL scatter estimation method to physical measurements, a breast phantom designed for quality control evaluation (Model 082, CIRS Inc., Norfolk, VA, USA) 50 and 80 mm thick was used. In addition, we retrieved 5 de-identified patient DBT scans, collected for a different study with ethics approval obtained from our institutional ethics committee, as an additional independent test set to evaluate our scatter estimation DL model for clinical cases. The DBT scans had compressed breast thicknesses of 30, 40, 47, 57, and 61 mm and were reviewed to ensure that they contained no visible artifacts. All images were obtained with a wide-angle DBT system (MAMMOMAT Inspiration, Siemens Healthcare, Forchheim, Germany).

2.2 | Monte Carlo simulations

Primary and scatter DBT projections of the created homogeneous and anthropomorphic digital breast phantoms were generated using a previously validated[7] MC simulation program based on the Geant4 toolkit. The simulations reflected the geometry of a wide-angle DBT system with 50° of scan angle, x-ray spectrum with tungsten anode and rhodium filtration,[21] typical and commercially used materials and thicknesses for the compression paddle and breast support table, detector housing, and an x-ray detector with an active selenium layer. The detector dimensions were set to 224 × 176 pixels with 1.36 mm pitch, corresponding to a 16 by 16 binning of the original 85 μm pixel pitch to reduce the computational cost. This down-sampling should not affect the scatter estimates since we aim to correct for the low-frequency component of the scatter signal.[7] The source to detector distance was set to 655 mm, including an air gap of 17 mm, with a point source emitting a beam collimated to the detector and including the heel effect. The simulated x-ray spectra matched those selected by the systems automatic exposure control (AEC) for the different breast thickness groups: tube voltages of 27, 28, 29, 30, 31, and 32 kV were used for the 30–39, 40–49, 50–59, 60–69, 70–79, and 80–89 mm thickness groups, respectively. The x-ray spectra were modeled based on the work of Hernandez et al.[21] to match the measured first half-value layer. The number of simulated x-rays was set to result in similar detector dose for all breasts by taking into account the thickness and glandular composition of each digital phantom case. The material composition for the randomly shaped digital phantoms was defined as a homogeneous mix of adipose and glandular tissue, varying randomly between 0% to 60% glandular tissue, by mass.

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

These simulations generated estimated scatter and primary maps per projection angle. The scatter image contains the energy deposited in the active layer by the photons that underwent any number of scatter interactions before reaching the detector. Any photons that might reach the active layer from below were excluded. The primary image contains the energy deposited in the active layer of the photons that reached the detector with no prior interactions. The set of projection angles simulated per digital phantom always included the 0° projection angle and additional projection angles randomly selected over a]−26°, 26°[range. For each homogeneous digital phantom, four zero-degree and twelve non-zero degree projections were simulated and one zero-degree and two non-zero degree projections were generated for each anthropomorphic digital phantom, since the latter were only used for fine-tuning the model. Therefore, the final homogeneous and anthropomorphic digital phantom datasets contained 9600 projections (600 phantoms × 16 projections) and 240 projections (80 phantoms × 3 projections), respectively.

2.3 | Data pre-processing

2.3.1 | Projection image

Currently, the reconstructed DBT data is pre-processed prior to reconstruction by performing a logarithmic transformation. Therefore, to optimize the performance of the DL model, it was trained to minimize the differences in the logarithmic domain.

For this reason, pre-processing of these images first involved flat-field correction of all the MC simulated data, as is commonly applied to clinical scans, followed by a further 2 by 2 binning to 112 × 88 pixel images to reduce the DL computational cost. Next, the pixel values of all projections were normalized to the median value measured in a region of interest (ROI) in the projection background (bg_{roi}) of the total (scatter + primary) signal, as illustrated in Figure 1. The bg_{roi} was obtained by first thresholding the breast projection image to obtain a mask, then using the inverse of such binary mask, excluding the corners of the image, and eroding it (one iteration using a kernel size of 29 × 29 pixels) to obtain an ROI large enough that it would not touch the breast area or the image limits, so that the region only includes background intensity values.

Each projection was then divided by the median of the obtained bg_{roi}, converted to the logarithmic domain, and divided by −10 to ensure most of the intensity range present in the image would be between 0 and 1, as per Equation (1):

$$I^{\theta}_{norm}(x,y) = \left[-\frac{1}{10}\ln\left(\frac{I^{\theta}(x,y)}{median\left(bg_{roi}\right)}\right)\right]_{+} \quad (1)$$

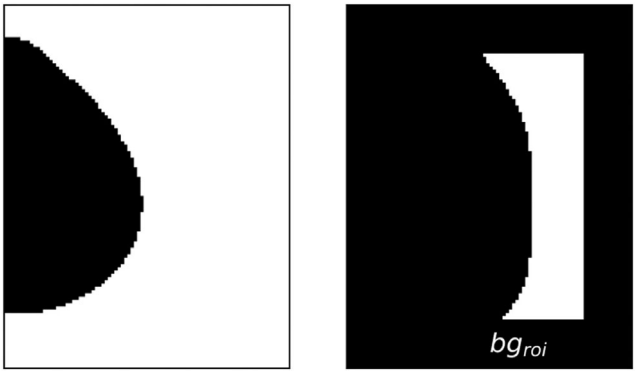

FIGURE 1 Binary representation of the breast projection at 15 degrees (left) and the selected ROI (bg_{roi}, right) used to obtain the median value in the projection background.

where θ is the projection angle, $I(x, y)$ is the DBT projection image and $median(bg_{roi})$ is the median value in the background ROI. Any zero (but not negative) values were set to 10^{-5}, indicated by the operation $[]_{+}$. Lastly, all images were zero-padded by 20 pixels before being given as input to the DL model.

The simulated and processed sets of scatter, primary and total (primary + scatter) projections from the homogeneous digital breast phantoms were split into training (480 phantoms, eight per mm of breast thickness, 7680 projections), validation (60 phantoms, one per mm of breast thickness, 960 projections) and test (60 phantoms, one per mm of breast thickness, 960 projections) sets. On the other hand, the anthropomorphic digital phantom projections were split into a fine-tuning (64 phantoms, 192 projections) and test set (16 phantoms, 48 projections).

2.3.2 | Thickness and angle maps

Since DBT projection images are also a function of the compressed breast thickness and the projection angle, in addition to the breast contents, thickness and angle maps were generated for each input into the DL model to provide additional information that can help improve the accuracy of our model to different breast cases, as demonstrated in another similar study.[15] We expected that this would create a more robust model that can better handle variations in breast shape and positioning. The maps, having the same size as the processed projection images, represent the pixel-by-pixel effective thickness travelled between compression paddle and breast support, and angle between the source and the corresponding pixel (see Figure 2). Provided the compressed breast thickness (th) value in millimeters, the thickness map would then consist of values in the range $[\mathrm{th}/100, \frac{th/100}{cosine(\alpha)}]$, where α is the largest angle found between source and one of the detector's correspon-

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

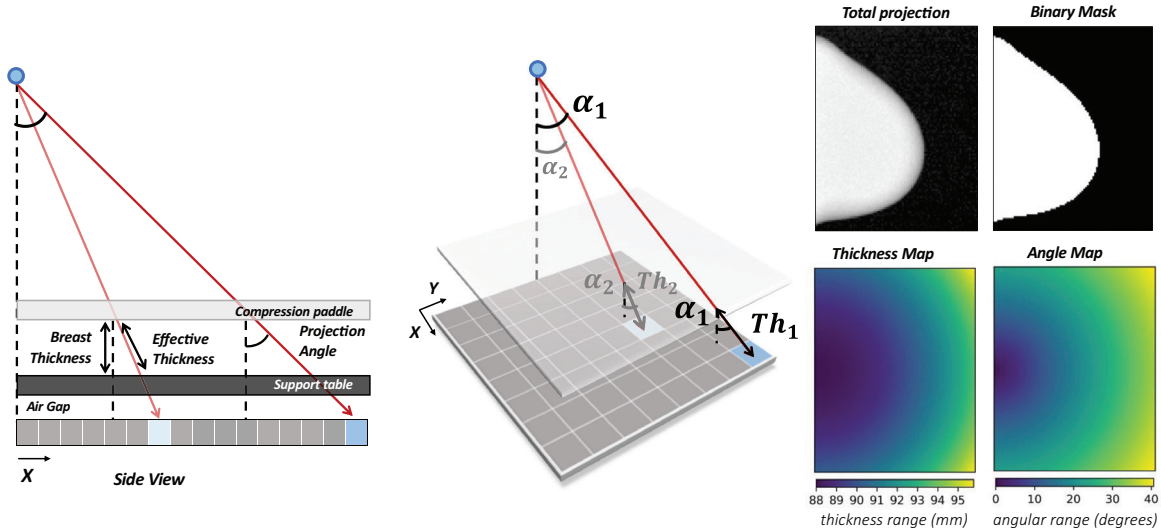

FIGURE 2 Schematic illustration to show how the projection angle (α) and effective thickness between compression paddle and breast support (Th) were calculated and corresponding final maps (with the range displayed in millimeters and degrees) for an 88 mm DBT homogeneous digital phantom projection at zero-degrees.

dent pixels in degrees, that is, the one found at the farthest, nipple-side, corner of the support table in a zero-degree projection, for instance. The angle map values were expressed in degrees and divided by 100 to scale them to below 1.

2.4 | Deep learning model for scatter estimation

For our DL model, we modified a U-Net architecture[22] where the compressed breast thickness map and projection angle map were input to the bottom layer and the pre-processed total projection image (i.e., the MC primary + scatter or later the acquired image) was input at the top layer (see Figure 3). The network was implemented using Python 3.8.2 and TensorFlow 2.4.0 backend with Keras (v. 2.4.3).[23]

The model's total loss function (Equation 2) was defined as the mean absolute relative difference between ground truth (I_{GT}) and predicted image (I_{pred}) and was calculated as an equally weighted sum between the losses from the area inside the breast (index *b*) and the background (index *bg*) area:

$$loss = \frac{\sum_{i=1}^{n} \left| \frac{I_{pred,\,i} - I_{GT,\,i}}{I_{GT,\,i}} \right|_{b}}{n_b} + \frac{\sum_{i=1}^{n} \left| \frac{I_{pred,\,i} - I_{GT,\,i}}{I_{GT,\,i}} \right|_{bg}}{n_{bg}} \quad (2)$$

where n_b and n_{bg} are, respectively, the total number of pixels inside and outside the breast segmented area of the projection image (meaning that, $n_b + n_{bg}$ = total number of pixels in the projection image) and *i* represents each projection image from a total of *n* projections. The mean absolute relative difference was used to avoid a bias toward thicker breast since these cases normally occupy larger areas of the projection image and would present a larger weight during loss minimization, particularly when compared with average to smaller thickness breasts. On the other hand, the equally weighted sum of the breast and background area was applied to help ensure that more accurate estimates of both areas were obtained. This is because, even though our primary interest is to ensure an accurate scatter prediction inside the breast segmented area, performing an estimation of the signal intensity in the whole image directly did not help to provide the accurate background signal intensity that is needed to correctly scale the DBT projections for reconstruction.

For the DL model training we initialized all kernel weights randomly. We set the batch size to 32 and used the Adam optimizer[24] and a learning rate of 5×10^{-4} that dropped in value by a factor of 0.1 after 5 epochs when the change in the model's total loss (monitored using the validation set for better generalization of our model after training) is smaller than 10^{-5}. The number of epochs was chosen experimentally and then controlled using early stopping to avoid overfitting of the training set.

We evaluated the stability of our model by applying a 5-fold cross-validation and selecting the best model. Fine-tuning of our model was performed by using the weights obtained from the best model achieved and further fitting it to the anthropomorphic digital phantom dataset to bring our scatter projection predictions closer to what would be found in a clinical DBT projection. Further details of our model, including parameter choices and illustration of our network's learning curve (Figure S1), has been added to the Supplementary Material of this paper.

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

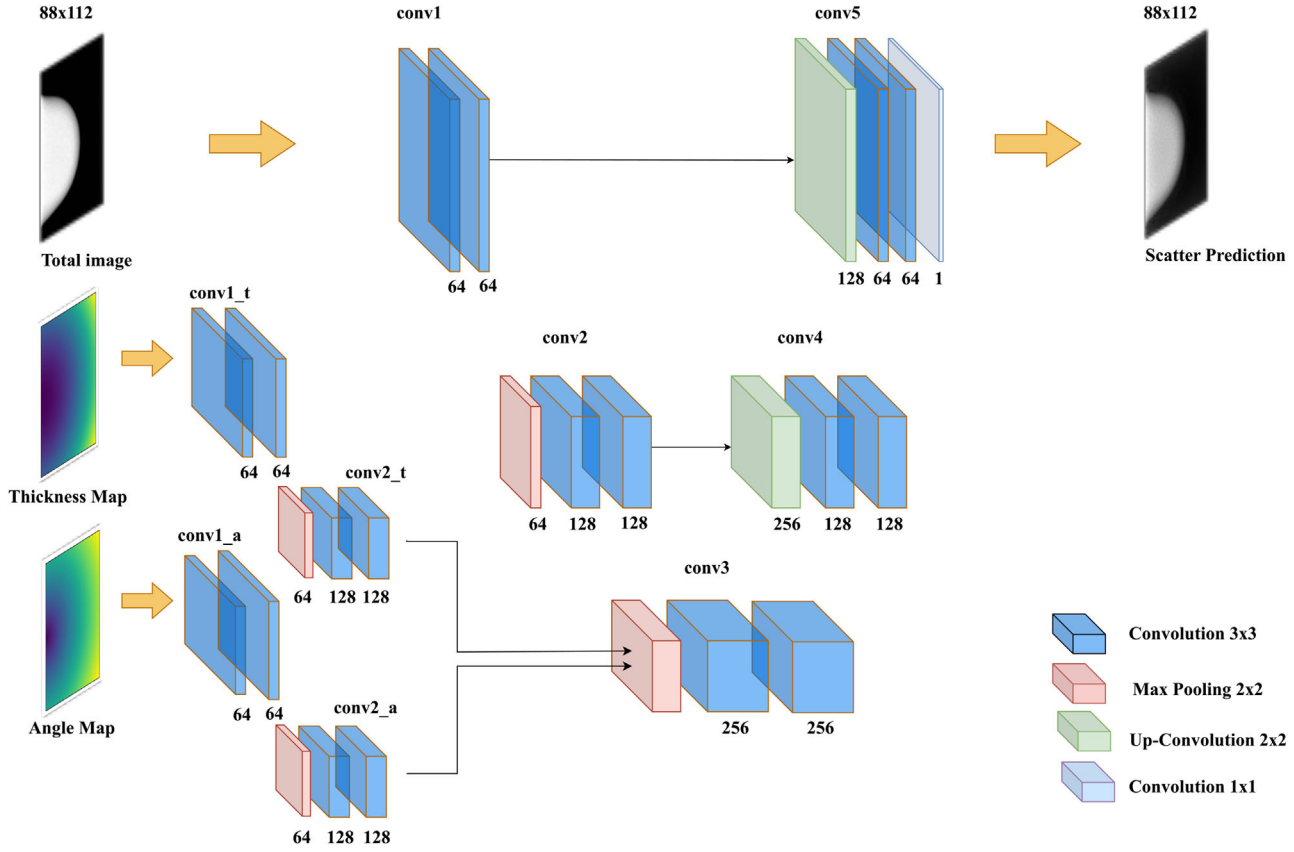

FIGURE 3 Modified U-Net architecture used to predict the scatter estimate based on a total projection input image, thickness map and angle map. Displayed values correspond to the number of feature channels for convolutional layers. Batch normalization layer and a leaky rectified linear unit activation layer followed each of the 3 × 3 convolution layers.

2.5 | Evaluation

The performance of our model was evaluated by comparing the scatter projections predicted by the DL model (S_{DL}) to the MC simulated scatter projections (S_{MC}) on our homogeneous digital test set, followed by a further assessment of its fine-tuned performance on the anthropomorphic digital phantom test set. The final signal intensity obtained by removing the scatter signal predicted by our DL method was also compared to physical scatter reduction methods, specifically the use of an anti-scatter grid in DM, performed with a physical homogeneous phantom. For the patient clinical DBT dataset, scatter-to-primary ratio (SPR) and reconstructed linear attenuation values were measured and compared to expected values from literature, and visual assessment of the original and corrected projections was performed.

2.5.1 | Mean relative differences

The mean relative difference (MRD, Equation 3) and the mean absolute relative difference (MARD, Equation 4) were measured inside the breast segmented area in the logarithmic domain for both the homogeneous and anthropomorphic test sets,

$$MRD\,(\%) = \frac{\sum_{i=1}^{n}\left[\left(\frac{S_{MC,i}-S_{DL,i}}{S_{MC,i}}\right)_{breast\ seg.\ area}\right]}{n} \times 100 \quad (3)$$

$$MARD\,(\%) = \frac{\sum_{i=1}^{n}\left|\frac{S_{MC,i}-S_{DL,i}}{S_{MC,i}}\right|_{breast\ seg.\ area}}{n} \times 100 \quad (4)$$

where n is once more the total number of pixels inside the breast segmented area in the projection image and i is the pixel index.

In this way, besides comparing S_{DL} to S_{MC}, we could evaluate, if possible, correlations or biases with respect to the digital phantom thickness or projection angle existed.

2.5.2 | Experimental evaluation

To assess the performance of our scatter correction method in physical phantoms of known ground truth, DM

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

acquisitions were obtained for the CIRS breast phantom. Images of the phantom were acquired without any scatter correction or reduction approach and acquired with the anti-scatter grid (DM + Grid), the commonly used physical scatter reduction method for DM. After acquisition, our DL scatter estimation method was applied to the full-field DM acquisitions where no scatter reduction approach was used.

For each of these conditions, the signal intensity was measured and compared by placing rectangular ROIs vertically (along the chest-wall) and horizontally (from chest-wall to nipple) at the mid-plane of both 50 and 80 mm homogeneous semi-circular CIRS phantoms. These measurements were performed at the center of the phantom and excluding its borders since our model was trained on realistic breast shaped phantoms that present differences in curvature at the border of the breast.

2.5.3 | Quantitative evaluation

While true scatter is unknown for clinical DBT data, we can test if the estimated scatter falls within the range expected from simulations and whether reconstructed linear attenuation of adipose tissue matches values reported in literature. Since the scatter signal has been shown to mainly depend on thickness, angle, and, to a minor extent, glandular fraction,[7] the SPRs from our clinical DBT data were calculated, using the best model trained on homogeneous digital phantoms, based on an ROI placed at the center of mass (COM) of the breast area. Thereafter, these SPR values were compared to theoretical estimates obtained from previously fitted equations that computed the SPR for breast thicknesses ranging from 2 to 8 cm.[7] The theoretical estimates were measured at the breasts' COM and for any tomosynthesis projection angle between ±30° of homogeneous simulated digital breast phantoms.

To evaluate the quantitative accuracy of the reconstruction after scatter correction, corrected and uncorrected projection data of a 46 mm anthropomorphic and a 47 mm clinical case with negligible amount of fibro-glandular structure, and using the fine-tuned model, were further corrected for beam hardening and then reconstructed using a model-based iterative reconstruction with spectral model,[25] generating a virtual monochromatic reconstruction at 20 keV. Linear attenuation values were then measured in an ROI located in the area of constant thickness through all the reconstructed image planes of the anthropomorphic and a clinical case and compared to values reported in literature.[26,27] Additionally, the processing steps mentioned were applied to another ground truth reference for further comparison: the MC simulated scatter-free (primary) signal from the anthropomorphic digital phantom case.

2.5.4 | Visual assessment and prediction run time

Additionally, a visual assessment was performed to identify possible artifacts not captured with the quantitative metrics. In the case of known ground truth (homogeneous and anthropomorphic digital breast phantoms), the scatter predictions from DL and MC were displayed together and an additional relative difference image was assessed to evaluate the DL scatter predictions inside the breast area, all in the logarithmic domain. In addition, vertical and horizontal profiles inside the breast were plotted for three similar cases, in terms of thickness and angular range, in each of the homogeneous, anthropomorphic and clinical DBT datasets.

The time taken to train and predict a single projection was also evaluated, together with the run times for generating scatter corrected projection images. The latter were evaluated as an average of the time taken to predict and correct a single and all 25 DBT projections for each of the five collected clinical cases. This was done on a workstation with 128 GB RAM, an AMD Ryzen Threadripper 1950 × 16-core processor, and an NVIDIA RTX A6000 48GB GPU.

3 | RESULTS

3.1 | Mean relative differences

The MRD between the DL scatter predictions and the MC simulations for all homogeneous test digital phantom projections had a median of 0.05% with an interquartile range (IQR) of −0.04% to 0.13%, while the MARD was 1.32% with an IQR of 0.98% to 1.85% (Figure 4, blue dots). The descending trend in the MARD results (Figure 4, bottom left corner) is due to the change in the number of photons used per MC simulation being set according to the thickness and glandular composition of each digital phantom case. However, we found that the pre-processing applied to normalize our projection images based on the median value measured in an ROI of the projection background was imprecise, resulting in lower noise in large thickness groups while it was intended to be constant. The MRD and MARD for the anthropomorphic digital phantoms, obtained an approximately similar MRD with a median of −0.21% and an IQR of −0.35% to −0.07%, and a MARD of 1.43% with an IQR of 1.32 to 1.66% (Figure 4, orange diamonds).

3.2 | Experimental evaluation

The signal intensity in the rectangular ROIs placed on the CIRS phantom DM images are plotted in Figure 5. It can be seen that our DL scatter estimation method

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

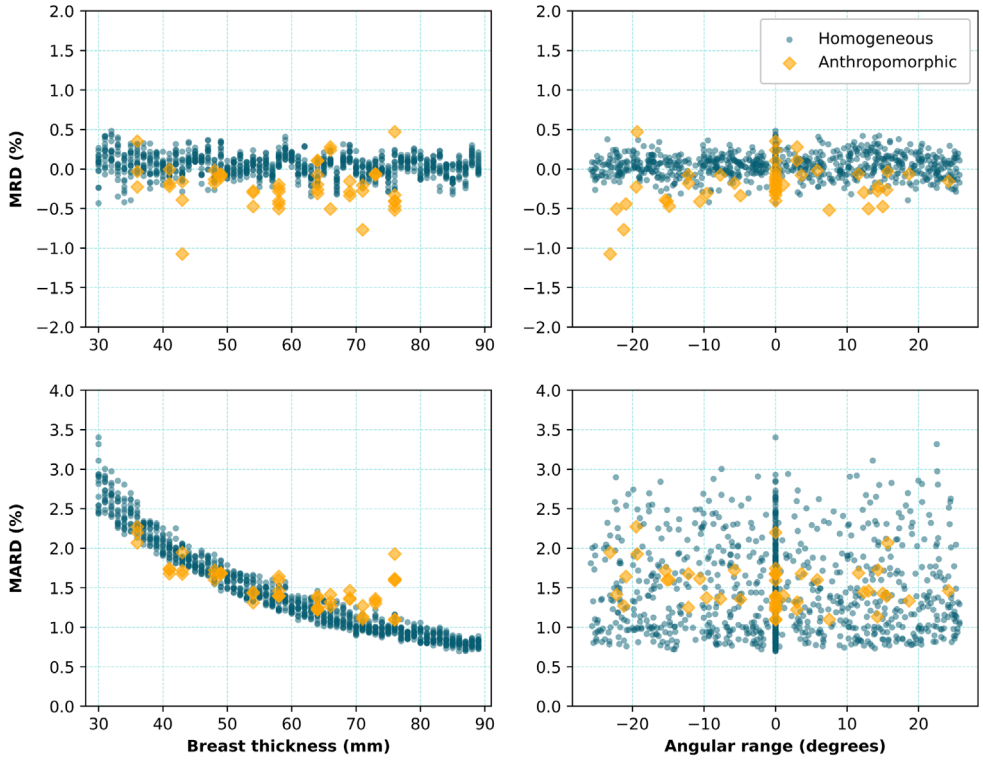

FIGURE 4 MRD and MARD measured in the logarithmic domain between MC and DL predictions for the 60 homogeneous and the 16 anthropomorphic digital breast phantoms of the respective test sets, plotted against their breast thickness (left column) and angular range (right column).

maintains the signal trend and intensity of the physical scatter reduction approach available, namely the use of an anti-scatter grid.

3.3 | Quantitative evaluation

The results of the SPR analysis of the five clinical cases are plotted in Figure 6. The theoretical SPR values for a given thickness and angle were calculated taking into account the possible extremes in breast glandularity (0% and 100%) that could be present in the ROI placed at the breast COM.

In the volume reconstructed from uncorrected projections, the linear attenuation of a 46 mm anthropomorphic digital phantom was found to be 0.409 cm^{-1}, compared to 0.474 cm^{-1} in the scatter corrected volume, bringing it closer to its simulated scatter-free reconstruction attenuation coefficient of 0.497 cm^{-1}. For the 47 mm-thick clinical case, the linear attenuation coefficient changed from 0.434 cm^{-1} in the uncorrected reconstruction to 0.511 cm^{-1} in the scatter corrected volume. The scatter correction brought the linear attenuation values in both cases closer to its theoretical value of 0.489 cm^{-1}, which is the mean from previously reported values in,[26,27] when compared to the 16% and 11% linear attenuation underestimation found in the uncorrected anthropomorphic and clinical cases, respectively. The comparison is shown in Figure 7.

3.4 | Visual assessment and prediction run time

Figure 8 shows MC and DL scatter predictions for three homogeneous and anthropomorphic digital phantoms, where the overall DL impact on scatter predictions inside the breast area in the logarithmic domain can be observed. It can be seen on the top and fourth row of Figure 8, how the thinner breasts result in higher positive and negative difference values (flatter histogram) when comparing the DL to the MC predictions, following what was observed in the MARD noise trend of the bottom left graphic in Figure 4. In addition, it can be observed how these values, which are all centered around zero, decrease their variance as the compressed breast thickness (and number of photons used in our simulations) increases (third and last row of Figure 8). It is also important to note that most cases do not present any particular pattern in the error by the DL predictions.

When plotting profiles of the scatter predicted by MC and by DL in Figure 9 for the two digital phantoms (left and middle), MC and DL scatter estimates

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

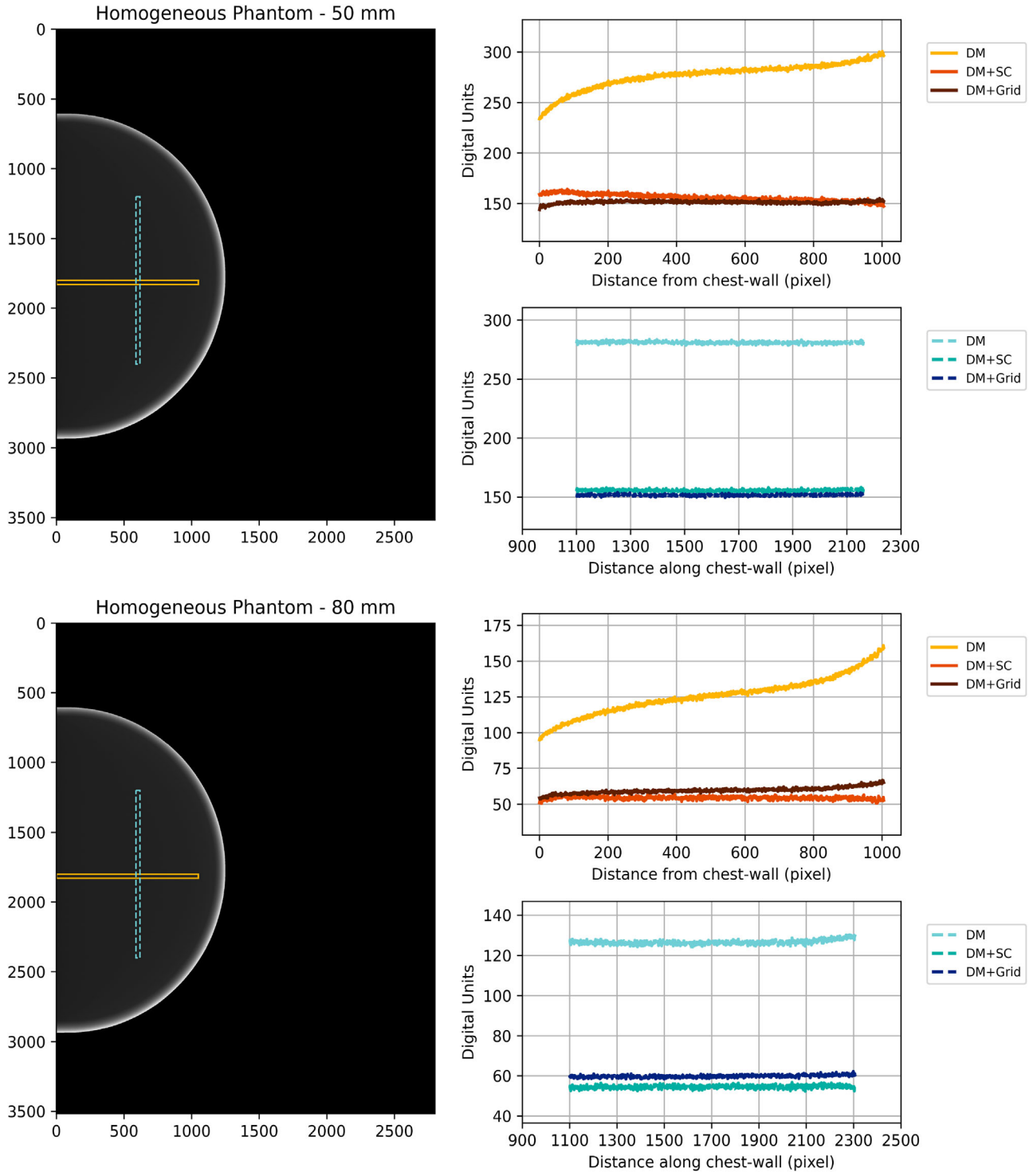

FIGURE 5 Physical measurements performed for DM acquisitions in 50 mm (top) and 80 mm (bottom) thick CIRS phantoms. Our proposed DL scatter correction (SC) method resulted in images comparable to those of the other scatter reduction method, the anti-scatter grid. Note that the scale of the y-axis in the graphs do not start at zero and that the digital units on the graphics y-axis are offset corrected.

are again closely matched. The trends followed by the total and primary signal in our MC simulations can also be observed in the clinical case (Figure 9, right side), where we present our scatter corrected signal (i.e., the primary signal) and the total signal inside of the breast.

Training of this model took approximately 40 min, while the prediction of a single projection takes less than 0.01 s. Generating the scatter corrected images took an average of 0.03 s per projection view for the clinical DBT exams, while for the entire DBT projection set it took on average 0.16 s.

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

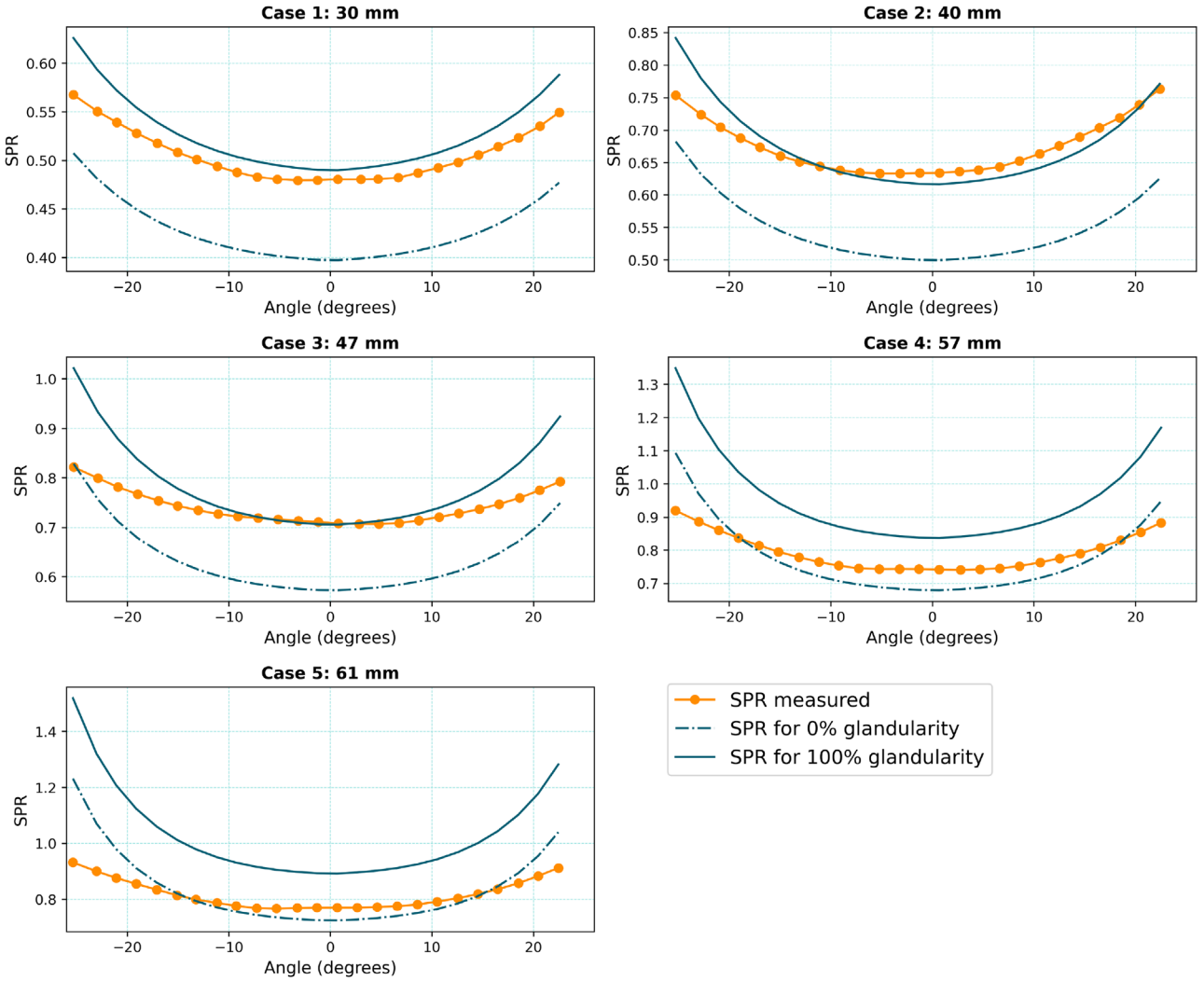

FIGURE 6 Graphical representation of SPR comparison between measured (dots) and theoretical (0% granularity—dashed lines; 100% granularity—solid lines) values for five clinical cases. Note that the y-axis scale varies per graph.

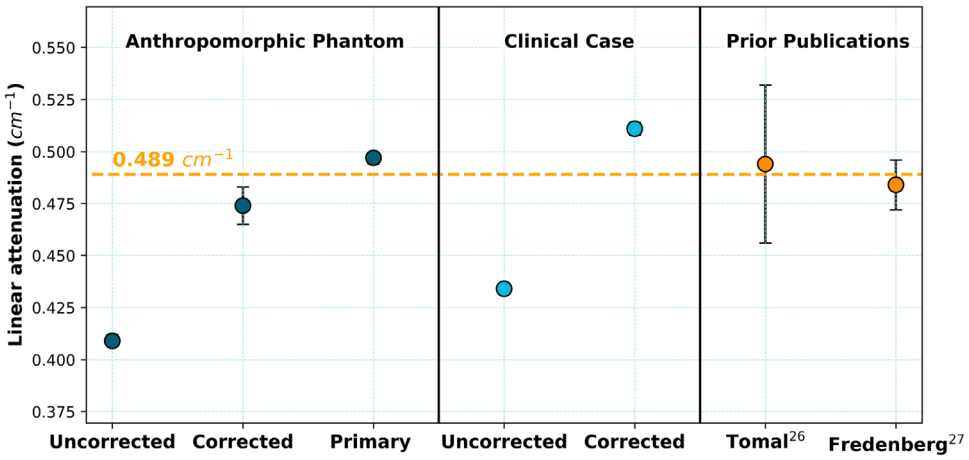

FIGURE 7 Measured reconstructed linear attenuation at 20 keV in a 46 mm anthropomorphic digital phantom (dark blue dots) and a 47 mm clinical case (light blue dots) compared to values reported in literature[26,27] (orange dots). Error bars represent one standard deviation, within a single case for our measurements, and between multiple measured cases for the values from literature. Dashed orange line represents the mean theoretical linear attenuation taken from the values reported in the literature.

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

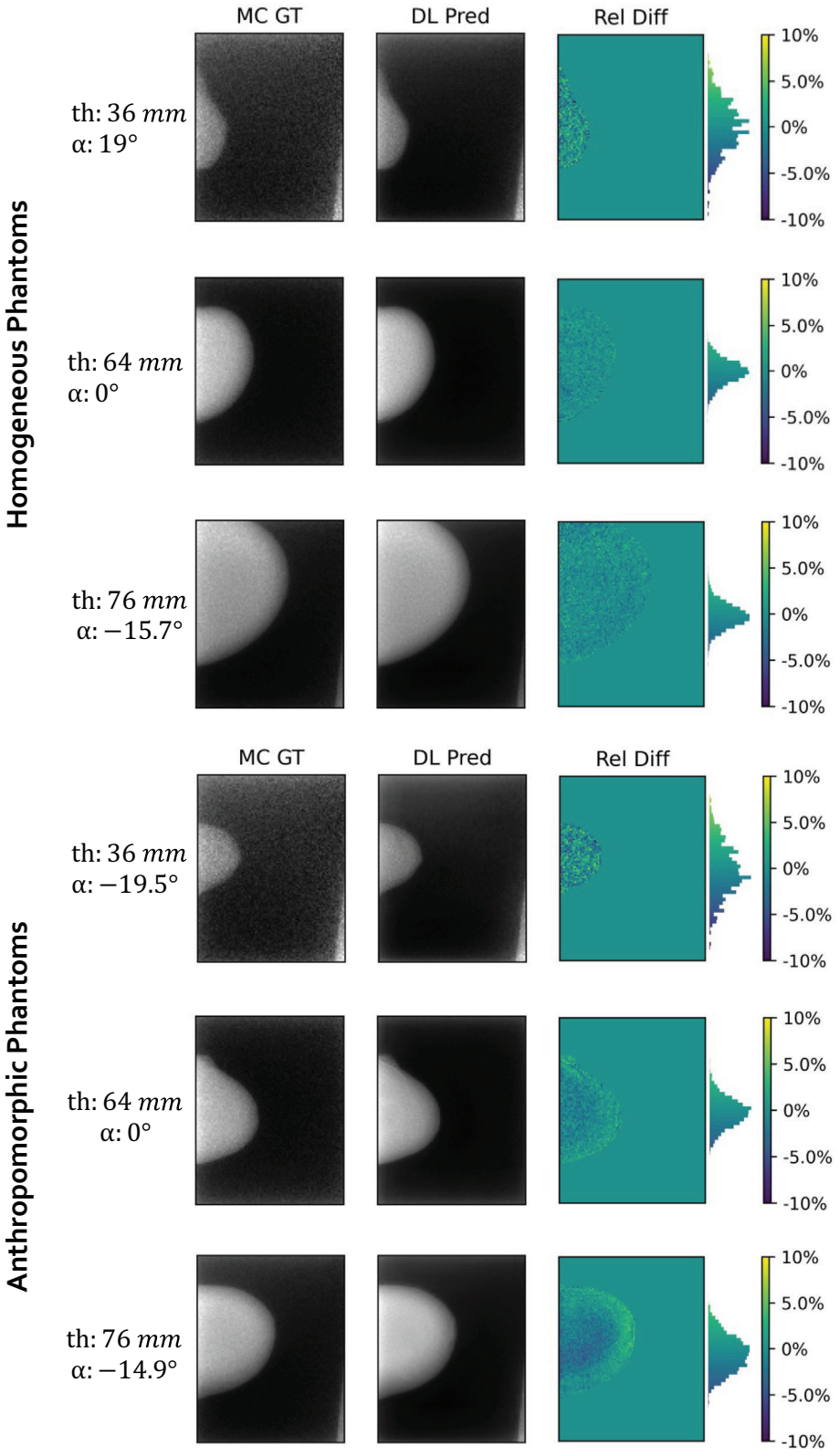

FIGURE 8 Illustration of scatter predictions from homogeneous (three first rows) and anthropomorphic (last three rows) digital phantoms for: MC (first column) and DL (second column) together with the relative difference image (third column) to observe the overall DL impact on scatter predictions in the logarithmic domain. The histograms plotted alongside the different digital breast phantom cases indicate the relative difference values found inside the breast segmented area only. Cases are shown from top to bottom going from lower to higher compressed breast thickness.

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

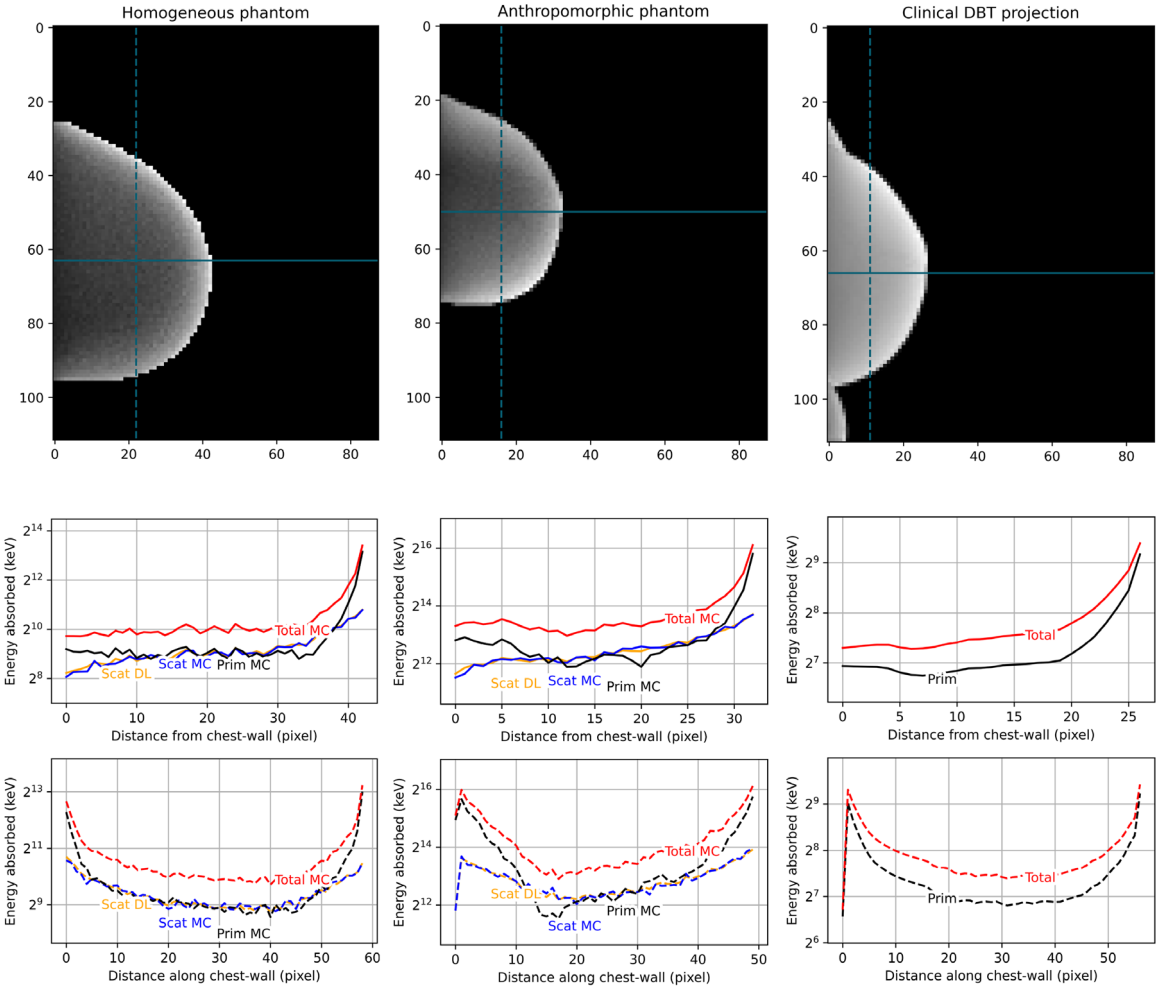

FIGURE 9 Vertical (dashed) and horizontal (full) profile lines inside of the breast for a 57 mm homogeneous digital phantom at a −21.8° angle (left), a 58 mm anthropomorphic digital phantom at a −20.9° angle (center) and a 57 mm clinical DBT projection at a −21.05° angle (right). In each graphic, the primary (black) and total (red) signal are plotted and for the digital phantom cases, the DL (yellow) and MC (blue) predicted scatter signal is also shown.

4 | DISCUSSION

Recent developments in breast cancer imaging led to the implementation and use of DBT as a screening and diagnostic modality. Nonetheless, DBT is still subject of ongoing research, for instance regarding its image quality enhancement. In this work, we present our contribution to this field by developing a DL model, trained on realistic MC simulations, that is capable of predicting the scatter signal from a DBT projection image without the need to provide other information than already clinically available, such as the compressed breast thickness and projection angle of the DBT projection.

Our model is able to provide a scatter signal prediction similar to those obtained through MC simulations, with a single prediction being on average 10^4 times faster than using MC simulations, comparable to what has been previously demonstrated by another study.[14] Additionally, unlike the work from Feng et al.,[11] where a lowpass filter needs to be applied to the original image to obtain the scatter signal, in our new approach the scatter signal is directly obtained and no extra empirical pre-processing step is needed.

A median MRD of approximately 0.05% (IQR, −0.04 to 0.13%) together with a median MARD of 1.32% (IQR, 0.98 to 1.85%) was obtained for the homogeneous digital phantom cases, while the model fine-tuned on anthropomorphic digital phantoms achieved similar MRD and MARD ranges of −0.21% (IQR, −0.35 to −0.07%) and 1.43% (IQR, 1.32 to 1.66%), respectively. The negative error observed for the MRD of the anthropomorphic digital phantoms indicates that there is a slight overestimation in the scatter signal by our DL model for these digital phantoms up to a maximum of approximately −1%. However, we believe this is a relatively small deviation from the ground truth scatter signal obtained from our MC simulations and thus it shows that the DL model can adapt well to reflect the scatter signal present in our MC simulations for different digital phantoms and for further use, even for other x-ray medical applications.

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The results presented in Figures 8 and 9 further exemplify the good prediction capabilities of the DL model, where the introduction of artifacts or major changes in the breast signal was avoided. The comparison between MC and DL predictions in Figure 8 shows that diverse breast segmented areas presented without any distinct pattern errors and with a low error range within ±10%. This indicates that the DL model correctly predicted the scatter signal in these digital phantoms without any specific bias, apart from some subtleties found near the skin line for some of the anthropomorphic digital phantoms. These biases, such as the one encountered near the skin line of the 76 mm anthropomorphic digital phantom (bottom row case in Figure 8), are likely due to a mismatch between the breast shape of this digital phantom and the breast shapes generated based on the patient-based model, which were used as input for training and prediction of the scatter signal. In Figure 9, it is shown that the intensity of the signal in our simulated input data was different when compared to the clinical cases used (Figure 9, y-axis values), but yet again the DL proved to be unaffected and provided a primary signal with the trend previously observed for the digital phantom cases, in the correct range of intensity values for the evaluated clinical case.

Further evaluation of the performance of our model included its application in the clinical setting. In Figure 5 it could be observed that our DL method provides an estimate of the scatter signal in the CIRS phantom comparable to the one obtained with physical reduction approaches, such as with an anti-scatter grid. The subsequent application of our model to patient clinical DBT exams presented comparable to improved results to previous studies in terms of reported SPR and reconstructed linear attenuation values, as well as the computational time required to scatter correct a DBT exam.

SPR values were measured for clinical cases with different breast thicknesses (as presented in Figure 6) and at different projection angles and presented values within ±15% of the ranges previously reported from a simulated framework,[7] except for projections angles higher than 21° in the 56 and 61 mm cases. This can be due to a lack of angle variability in the anthropomorphic digital phantoms used to create the fine-tuning model, but it can also be due to a difference between the system setup simulated in[7] compared to ours, since our validation did not show a bias for large angles (Figure 4).

Quantitative evaluation of linear attenuation values of adipose tissue (illustrated in Figure 7) for the DBT projection data of a 46 mm anthropomorphic digital phantom and a 47 mm clinical case showed an improvement from 16% and 11% underestimation to 2.3% underestimation and 4.4% overestimation after scatter correction, when compared to previously reported values from literature.

The run times obtained from predicting scatter maps through our DL model and correcting a DBT exam were improved regarding previous work. In our DL approach, the average time to scatter correct five entire DBT cases with 25 projections each was less than a second, with around 0.03 s per projection for prediction and correction using our previously specified GPU system, compared to the fastest reported time of 3 s taken to just predict scatter per projection using a GPU-based fast Monte Carlo simulation approach.[28] Of course, variations in hardware and future improvements in GPU-based Monte Carlo algorithms could reduce this difference.

Limitations in our work include the fact that our MC, and consequently DL, simulations were based on the geometry of one vendor and used physical parameters dependent on one specific system, which might be different for other systems depending on manufacturing tolerances. The DL model was tested as well on one DBT system, and so validation of the model on other systems would be needed to further strengthen our conclusions. Nonetheless, we believe that possible system deviations can be accounted for in the data pre-processing step, before applying the scatter estimation. The set of clinical cases used was limited because in the dataset available, for cases above 61 mm the background signal intensity was clipped to a constant value while our approach requires the background information to pre-process the image before the DL scatter prediction. Finally, we developed and applied this model for the CC view only and for practical application the MLO view should be included as well, since standard DM and DBT exams include both views. Therefore, future work includes demonstrating it clinically in very large breasts (70 and 80 mm thickness range), extending this model to the MLO view, and training and validating a fine-tuned model for the scatter correction of contrast-enhanced mammography images.

5 | CONCLUSION

We developed a fast and accurate method to estimate the scatter signal for breast tomosynthesis that will allow further development of quantitative applications.

ACKNOWLEDGMENTS

This work was supported by an academic-industrial collaboration with Siemens Healthineers, XP Division. The presented method is not commercially available and its future availability is not guaranteed. We would also like to thank the support from Nvidia Corporation with the donation of the GPU used in this research as part of the NVIDIA Academic Hardware Grant Program, and SURF (www.surf.nl) for the support in using the Dutch National Supercomputer Snellius.

CONFLICT OF INTEREST STATEMENT

The authors have the following conflicts of interest to disclose: I. Sechopoulos—research agreements: Canon Medical Systems, Siemens Healthcare, Sectra Benelux,

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ScreenPoint Medical, Volpara Health; R. Biniazan, and S. Kappler are employees of Siemens Healthcare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are not publicly available due to the constraints imposed by our academic-industrial collaboration.

REFERENCES

1. Houssami N, Zackrisson S, Blazek K, et al. Meta-analysis of prospective studies evaluating breast cancer detection and interval cancer rates for digital breast tomosynthesis versus mammography population screening. *Eur J Cancer*. 2021;148:14-23. doi:10.1016/j.ejca.2021.01.035
2. Peppard HR, Nicholson BE, Rochman CM, Merchant JK, Mayo RC, Harvey JA. Digital breast tomosynthesis in the diagnostic setting: indications and clinical applications. *Radiographics*. 2015;35(4):975-990. doi:10.1148/rg.2015140204
3. Butler R, Conant EF, Philpotts L. Digital breast tomosynthesis: what have we learned? *J Breast Imaging*. 2019;1:9-22. doi:10.1093/jbi/wby008
4. Bernardi D, Gentilini MA, De Nisi M, et al. Effect of implementing digital breast tomosynthesis (DBT) instead of mammography on population screening outcomes including interval cancer rates: results of the Trento DBT pilot evaluation. *Breast*. 2020;50:135-140. doi:10.1016/j.breast.2019.09.012
5. Georgian-Smith D, Obuchowski NA, Lo JY, et al. Can digital breast tomosynthesis replace full-field digital mammography? A multireader, multicase study of wide-angle tomosynthesis. *Am J Roentgenol*. 2019;212(6):1393-1399. doi:10.2214/AJR.18.20294
6. Wu G, Mainprize JG, Boone JM, Yaffe MJ. Evaluation of scatter effects on image quality for breast tomosynthesis. *Med Phys*. 2009;36(10):4425-4432. doi:10.1118/1.3215926
7. Sechopoulos I, Suryanarayanan S, Vedantham S, D'Orsi CJ, Karellas A. Scatter radiation in digital tomosynthesis of the breast. *Med Phys*. 2007;34(2):564-576. doi:10.1118/1.2428404
8. Diaz O, Dance DR, Young KC, Elangovan P, Bakic PR, Wells K. Estimation of scattered radiation in digital breast tomosynthesis. *Phys Med Biol*. 2014;59(15):4375-4390. doi:10.1088/0031-9155/59/15/4375
9. Bliznakova K. Power spectrum analysis of the x-ray scatter signal in mammography and breast tomosynthesis projections. *Med Phys*. 2013;40(10):1-7. doi:10.1118/1.4820442
10. Tagliafico A, Houssami N. Digital breast tomosynthesis might not be the optimal modality for detecting microcalcification. *Radiology*. 2015;275(2):618-619.
11. Feng SSJ, Sechopoulos I. A software-based x-ray scatter correction method for breast tomosynthesis. *Med Phys*. 2011;38(12):6643-6653. doi:10.1109/SCCC.1999.810170
12. Lu Y, Peng B, Lau BA, et al. A scatter correction method for contrast-enhanced dual-energy digital breast tomosynthesis. *Phys Med Biol*. 2015;60(16):6323-6354. doi:10.1088/0031-9155/60/16/6323
13. Duan X, Sahu P, Huang H, Zhao W. Scatter correction with deep learning approach for contrast enhanced digital breast tomosynthesis (CEDBT) in both cranio-caudal (CC) view and mediolateral oblique (MLO) view. Proc SPIE 11513, IWBI2020. 2020:99. doi:10.1117/12.2564358
14. Rodríguez-Ruiz A. Artificial intelligence and tomosynthesis for breast cancer detection. 2019.
15. Pautasso JJ, Caballo M, Mikerov M, Boone JM, Michielsen K, Sechopoulos I. Deep learning for x-ray scatter correction in dedicated breast CT. *Med Phys*. 2023;50:2022-2036. doi:10.1002/mp.16185
16. Pinto MC, Michielsen K, Biniazan R, Kappler S, Sechopoulos I. Generative compressed breast shape model for digital mammography and digital breast tomosynthesis. *Med Phys*. 2023;50:2928-2938. doi:10.1002/mp.16133
17. Rodríguez-Ruiz A, Agasthya GA, Sechopoulos I. The compressed breast during mammography and breast tomosynthesis: in vivo shape characterization and modeling. *Phys Med Biol*. 2017;62(17):6920-6937. doi:10.1088/1361-6560/aa7cd0
18. Rodríguez-Ruiz A, Feng SSJ, Mann JR, D'Orsi CJ, Sechopoulos I. Improvement of an objective model of compressed breasts undergoing mammography: generation and characterization of breast shapes. *Eur Congr Med Phys*. 2017;44(6):2161-2172.
19. Pinto M, Egging R, Michielsen K, Sechopoulos I. Compressed breast shape characterization and modelling during digital breast tomosynthesis using 3D stereoscopic surface cameras. In: Proc. SPIE 11513, IWBI2020. 1151307:2020. doi:10.1117/12.2563823
20. García E, Fedon C, Caballo M, Martí R. Sechopoulos I, Diaz O. Realistic compressed breast phantoms for medical physics applications. In: Proc SPIE 11513. 2020;1151304:73. doi:10.1117/12.2564273
21. Hernandez AM, Seibert JA, Nosratieh A, Boone JM. Generation and analysis of clinically relevant breast imaging x-ray spectra. *Med Phys*. 2017;44(6):2148-2160. doi:10.1002/mp.12222
22. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2015;9351:234-241. doi:10.1007/978-3-319-24574-4_28
23. Chollet F. Keras. 2015. https://keras.io
24. Kingma DP, Ba JL. Adam: a method for stochastic optimization. *Proc. of the 3rd International Conference on Learning Representations (ICLR2015)*. 2015.
25. Michielsen K, Fedon C, Nagy JG, Sechopoulos I. Dose reduction in breast CT by spectrum switching. In: Proc SPIE 10718, IWBI2018. 2018:65. doi:10.1117/12.2318425
26. Tomal A, Mazarro I, Kakuno EM, Poletti ME. Experimental determination of linear attenuation coefficient of normal, benign and malignant breast tissues. *Radiat Meas*. 2010;45(9):1055-1059. doi:10.1016/j.radmeas.2010.08.008
27. Fredenberg E, Willsher P, Moa E, Dance DR, Young KC, Wallis MG. Measurement of breast-tissue x-ray attenuation by spectral imaging: fresh and fixed normal and malignant tissue. *Phys Med Biol*. 2018;63(23). doi:10.1088/1361-6560/aaea83
28. Kim K, Lee T, Seong Y, et al. Fully iterative scatter corrected digital breast tomosynthesis using GPU-based fast Monte Carlo simulation and composition ratio update. *Med Phys*. 2015;42(9):5342-5355. doi:10.1118/1.4928139

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pinto MC, Mauter F, Michielsen K, Biniazan R, Kappler S, Sechopoulos I. A deep learning approach to estimate x-ray scatter in digital breast tomosynthesis: From phantom models to clinical applications. *Med Phys*. 2023;50:4744–4757. https://doi.org/10.1002/mp.16589

24734209, 2023, 8, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16589 by University Of Twente Finance Department, Wiley Online Library on [23/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License