Original paper

Breast glandularity and mean glandular dose assessment using a deep learning framework: Virtual patients study

Rodrigo T. Massera, Alessandra Tomal^{*}

Institute of Physics “Gleb Wataghin”, University of Campinas, Campinas, Brazil

ARTICLE INFO

Keywords:

Mammography
Mean Glandular Dose
breast glandularity
deep learning

ABSTRACT

Purpose: Breast dosimetry in mammography is an important aspect of radioprotection since women are exposed periodically to ionizing radiation due to breast cancer screening programs. Mean glandular dose (*MGD*) is the standard quantity employed for the establishment of dose reference levels in retrospective population studies. However, *MGD* calculations requires breast glandularity estimation. This work proposes a deep learning framework for volume glandular fraction (*VGF*) estimations based on mammography images, which in turn are converted to glandularity values for *MGD* calculations.

Methods: 208 virtual breast phantoms were generated and compressed computationally. The mammography images were obtained with Monte Carlo simulations (MC-GPU code) and a ray-tracing algorithm was employed for labeling the training data. The architectures of the neural networks are based on the XNet and multilayer perceptron, adapted for each task. The network predictions were compared with the ground truth using the coefficient of determination (r^2).

Results: The results have shown a good agreement for inner breast segmentation ($r^2 = 0.999$), breast volume prediction ($r^2 = 0.982$) and *VGF* prediction ($r^2 = 0.935$). Moreover, the *DgN* coefficients using the predicted *VGF* for the virtual population differ on average 1.3% from the ground truth values. Afterwards with the obtained *DgN* coefficients, the *MGD* values were estimated from exposure factors extracted from the DICOM header of a clinical cohort, with median(75 percentile) values of 1.91(2.45) mGy.

Conclusion: We successfully implemented a deep learning framework for *VGF* and *MGD* calculations for virtual breast phantoms.

1. Introduction

Digital mammography is an imaging technique recommended in several countries for breast cancer screening, being associated with a reduction in mortality rate and successful treatment of breast cancer [1]. Mean glandular dose (*MGD*) is currently the adopted quantity used for dosimetry in mammography, since glandular tissue is the most prone to radiation-induced mutations. Historically, *MGD* values have been estimated considering simplified breast models [2], which are not, however, representative of mean population glandular tissue content and distribution [3], and can results in significant differences [4,5].

With the development of new breast imaging techniques (i.e. digital breast tomosynthesis, DBT, and breast-CT), the quantity and vertical location of glandular tissue within the breast could be better evaluated to generate heterogeneous breast models [6–13]. Breast-CT provides more accurate breast models, since it allows to characterize in details the

3D distribution of the breast tissues [14] and could be used for patient-specific dose estimation. However, this technique is not available worldwide and it is limited to a small cohorts or number of images. Thus, breast models based on breast-CT images are not viable to cover the large populations variability where this technique is not implemented yet. Although the vertical location of glandular tissue within the breast is necessary for patient-specific dose estimation, homogeneous models based on more realistic and specific breast characteristics (i.e. dimensions and composition) could be used to obtain a more accurate radiation dose estimates. This allows to establish more accurate Dose Reference Levels [15] for large populations compared to a model that adopts the same glandularity for the entire population. In addition, the large number of images from mammography screening acquired each year around the world could be an important data source for the development of population-based breast models.

For a more accurate evaluation of *MGD* in a large population-based

^{*} Corresponding author.

E-mail address: atomal@ifi.unicamp.br (A. Tomal).

<https://doi.org/10.1016/j.ejmp.2021.03.007>

Received 30 November 2020; Received in revised form 27 February 2021; Accepted 2 March 2021

Available online 10 May 2021

1120-1797/© 2021 Associazione Italiana di Fisica Medica. Published by Elsevier Ltd. All rights reserved.

Fig. 1. Flowchart of the whole pipeline of this study to calculate the breast glandularity and mean glandular dose based on virtual patients.

screening mammography, it is desirable to use an estimator to predict the breast glandular content from the images for each patient breast. Quantitative assessment of breast density (BD) based on mammography images has been largely performed using automated or semi-automated imaging systems, such as Cumulus (University of Toronto, Canada), Volpara (Volpara Solutions Ltd., New Zealand), Quantra (Hologic Inc., Bedford, MA, USA) and LIBRA (University of Pennsylvania, USA), among others. Most of them are commercially available and used for clinical applications. Although the accuracy and robustness of these systems has been extensively explored and they have excellent or moderate reliability for repeated breast density measures [16,17], the results are validated against other 3D breast imaging modalities results or using physical breast phantoms. However, such validations can be challenging since the ground truth from real patient images are always unknown and the target breast density are indirectly estimated from 3D imaging techniques, resulting in uncertainties due to tissue segmentation. Moreover, when physical breast phantoms (simplified or anthropomorphic) are used, a limitation of availability and variability between them in terms of materials, costs and infrastructure appears. There is also the distinction between breast density based on area and the volumetric breast density (VBD), where the latter considers the variation of the breast thickness during compression and enables an estimation of the breast composition [18]. Recently, also deep learning techniques have been employed for breast density assessments in mammography [19–23]. Although with promising results, a significant number of labeled data is required to train these models, which are usually acquired via other breast density algorithms or categorized by radiologists. On top of the difficulty to obtain the training data, the labeled results acquired indirectly do not represent the ground truth values, but predictions (which could carry potential errors and biases). Therefore, another method to acquire labeled data to train deep learning models would be useful. Considering this scenario, a virtual clinical study could generate the necessary data, where the ground truth is known in order to train deep learning models. In addition, an independently trained deep learning model freely available could be a comparative/complementary result to other VBD algorithms.

This work presents the development of a deep learning framework for calculating the breast volume and the volume glandular fraction (VGF, consequently the VBD) from simulated mammography images obtained with anthropomorphic virtual phantoms. In this preliminary study, we show the feasibility and performance of training three neural networks, one for each specific task (segmentation, height prediction and relative glandular height prediction) with a labeling system based on a ray-tracing algorithm in order to obtain the breast glandularity. Afterwards, in an application topic, the MGD is estimated for the virtual breast phantoms using a homogeneous model approximation with the predicted glandularity and exposure parameters based on a clinical cohort.

Table 1

Characteristics of the breast phantoms. The radius is calculated by approximating the breast as a semicylinder. The values are expressed as median(25–75 percentiles).

# Phantoms	Thickness (cm)	Volume (cm ³)	Radius (cm)	VGF
208	5.8 (4.6–6.8)	420 (325–580)	6.7 (5.9–7.9)	0.23 (0.13–0.51)

2. Materials and Methods

The methodology used in this study is summarized in Fig. 1. Specific details related to each step of the flowchart are described in the following sections.

2.1. Computational breast phantoms

Anthropomorphic 3D breast phantoms were generated with the BreastPhantom software [24], based in two steps. First, the input parameters were selected in order to generate breasts with different sizes, shapes and glandular content, based on breast-CT data, to address the variability present in the population [6,9,25]. The phantoms consisted of 0.25 mm cubic voxels, and a 1.5 mm skin thickness was considered [26]. The virtual breasts are composed of skin, adipose, glandular, blood (for arteries), muscle and connective tissues. Second, the breast was compressed (BreastCompress program) [27] with a finite-element software (FEBio v. 2.9) [28] and cropped to remove uncompressed tissues, mainly muscle (BreastCrop program) [27]. The whole process for generating each phantom took less than 30 min (depending on the breast thickness) using a standard computer (Ryzen 2700 3.2 GHz, 8 cores 16 threads, 16 GB RAM). A Python (v.3) script was written to automate the process to generate a total of 208 phantoms, whose characteristics are summarized in Table 1.

2.2. Data generation: Monte Carlo simulations

The image acquisition for a mammography cranio-caudal (CC) examination was simulated with the MC-GPU Monte Carlo code [29]. This code was chosen because it was already used successfully in a virtual clinical trial for mammography studies [30,31]. Moreover, the code runs on graphical processing units (GPUs), which offers high performance to simulate the image acquisition in a reasonable time, considering the available hardware.

The simulated geometry consists of a point source, the compression plate (2.75 mm thick of polycarbonate), the support plate (2 mm of carbon fiber) and the compressed breast. The effects of a linear grid, whose grid parameters were based on Cunha et al. [32], were accounted using the transmission factors calculated using the algorithm proposed by Day and Dance [33]. An amorphous-selenium detector (200 μm thick), comprised of 1024 × 832 pixels, is located, respectively, 70 cm and 2.5 cm below the X-ray source and the bottom of the breast. The

Table 2

X-ray spectra used for image simulation as a function of the compressed breast thickness range.

Thickness range (mm)	Filter material	Tube potential (kV)
20–25	Rhodium	25
25–35		26
35–45		27
45–50		28
50–55		29
55–60		30
60–65		31
65–70	Silver	30
70–80		32
80–85		33
85–90		34

number of pixels was chosen based on real mammography detectors [34] with a 4×4 binning. The detector and the compression/support plates have an area equal to $29 \times 24 \text{ cm}^2$, while the X-ray field irradiated an area of approximately $27.6 \times 17.3 \text{ cm}^2$ of the detector (sufficient to irradiate all the surface of the breast phantoms). The detector is an ideal energy integrating device, with parameters set to default: electronic signal of 5200 electrons (mean), Swank factor of 0.99 and detector gain (W) equals to 50 eV per detected charge. Therefore, the signal is given in electrons/cm², and no dynamic range scaling was performed. After the image is generated, a mask is applied to modify the pixel intensity values to compensate for inhomogeneities introduced by the X-ray beam divergence geometry (i.e. inverse square law and angle of incidence). No further processing was applied, thus the images are equivalent as “for processing”.

The number of photon histories varied between approximately 10^{11} and 10^{12} with a speed of 10^8 histories/s on a Tesla P100 (NVIDIA, USA) GPU. The corresponding average and standard deviation of the incident air kerma values, including all cases, were 1.5 and 0.7 mGy, respectively.

The absorption energy for photons is 1 keV, while the electrons are locally absorbed [35]. The photon cross sections were calculated using PENELOPE (v. 2018) [36]. The composition of breast tissues (adipose, glandular and skin) were obtained from Hammerstein et al. [37], while for all other breast tissues were obtained from Woodard and White [38]. The X-ray spectra were generated using the TASMICS software for a tungsten target [39]. Table 2 shows the filter material and tube potential for the X-ray spectra used for each breast thickness interval, considering tungsten as the anode material. In some cases (when the breast thickness is near the limit for an interval), the tube potential selected was 1 kV above or below from those displayed from Table 2 in order to evaluate possible variations from the automatic exposure control (AEC).

2.3. Breast models masks: segmentation, relative height and relative glandular height

The breast phantoms segmentation to separate different tissues was performed by a simplified ray-tracing algorithm (a generalist and complete ray-tracing algorithm implementation is explained in the reference [40]). First, for each material, a numerical matrix (M_i) filled with zeros is defined in the same place as the X-ray detector with each matrix element representing the detector pixels. Afterwards, for each pixel, a dummy particle is generated and travels backwards, in the direction of the X-ray source (in this case a point source). For each step “s” traveled by the particle (defined by the user), the algorithm checks the current voxel material and adds one to the counter. After the mapping is finished for all matrix elements, the counts are returned for the material. This process is repeated for all materials, thus several matrices are generated ($M_1 \dots M_N$). The value s was empirically determined to be equal to the voxel side length.

After this process is performed and with the M_N matrices, three

masks were build in order to train the neural networks. Each mask belongs to a specific task, explained in the following.

The first mask (task I) classified the image in three regions: (i) background, (ii) skin contour plus nipple and (iii) inner tissues. For this purpose, the following criteria were used: (i) is defined as the elements of the matrices that did not contain any breast tissues; (ii) the matrices elements that contain nipple or a certain threshold of skin content (this value was determined as the minimum fraction to form a continuous contour around the breast); (iii) the elements that did not fit the criteria (i) or (ii) were selected as region (iii) using the flood fill technique inside the contour formed by (ii). A binary erosion technique [41] was used to remove isolated pixels from (ii) that could be present inside (iii).

The second mask (task II) describes the relative breast height (h) for each pixel located in the inner breast (obtained from the first mask). It is calculated by summing all the matrices elements in depth and normalizing them by the compressed breast thickness. Due to the beam divergence, the relative height was corrected to yield values between 0 and 1. It was noticed that during the breast compression routine, small air gaps were present between the breast and the compression or support plates (generally less than 3% the compressed breast thickness), shifting the maximum height in some regions.

The third mask (task III) quantifies the relative glandular height (g) of each projected pixel (i.e. the ratio between the height of glandular tissue and total height on the breast). For this specific case, the mask is divided in patches of 4×4 pixels, and the average glandular height is calculated for each patch.

The inner volume of the breast (Vol) (without skin and nipple) was estimated from masks I and II, as:

$$Vol = \left(\frac{SDD - AG - t/2}{SDD} \right)^2 \times \sum_{i=0}^{i=N} m_i \times h_i \times \left(t - 2 \times ST \right) \times A \quad (1)$$

where SDD is the source-to-detector distance, AG is the distance between the bottom of the breast and the detector, t is the compressed breast thickness, N the number of patches, m_i is the i th element of mask I (the value is 1 for inner breast, 0 otherwise), h_i the relative height from mask II for the element i, ST is the skin thickness and A the pixel area. This approximation considers that the projected breast image corresponds to the area at half of the breast thickness.

The volume glandular fraction (VGF, i.e. ratio between the glandular volume and the total volume, excluding skin) [6] was estimated by combining masks I, II and III, by performing a sum over the 4×4 patches

$$VGF = \frac{\sum_{i=0}^{i=N} \bar{A}_i \times \bar{h}_i \times g_i}{\sum_{i=0}^{i=N} \bar{A}_i \times \left(\bar{h}_i - h_s \right)} \quad (2)$$

where N the number of patches, $\bar{A}_i$ is the relative area of the patch belonging to the inner breast (from 0 to 1), g_i the glandular height (from 0 to 1), $\bar{h}_i$ the relative average height of the patch (with skin), h_s the relative skin thickness.

The calculated volumes were compared with the nominal volumes for 178 phantoms with average and maximum differences of 0.9% and 3.6%, respectively. We also studied the relation between the ground truth breast volume with and without skin, and found a linear behavior ($r^2=0.999$). These results are summarized in A. The calculated VGF using the masks were compared with the nominal values for 178 phantoms, showing differences up to 10% due the constant skin thickness approximation. A third order polynomial fit was adjusted to convert mask VGF to ground truth VGF to remove potential bias with this approximation, as explained in A. All VGF results further on have this correction applied.

Fig. 2. XNet [42] adapted architecture implemented in tissues segmentation and height quantification tasks. Each box corresponds to a layer, and the number below represents the number of channels.

Fig. 3. Illustration of glandular fraction estimation and the implemented architecture, which consists of five trained multi-layer, followed up by two fully connected layers (dense).

2.4. Neural networks: deep learning framework

A total of three neural networks architectures were employed, one for each segmentation task (I to III) described in Section 2.3. The first two are based on an adapted version of XNet [42], a convolutional neural network architecture previously used for X-ray imaging classification. From now in this work, the terms network and neural network are used interchangeably. The network implementation was made in Python (v. 3.6.9) using the PyTorch framework (v. 1.6.0) [43], while the training was performed on a NVIDIA GTX 1060 (6 GB of VRAM, CUDA v. 11.0). The network hyperparameters, if not explicitly mentioned, were obtained empirically by testing exhaustively the values that yielded better training and validation results. 178 phantoms were used for training and validation (80%:20% proportion), while 30 phantoms were selected for testing.

Fig. 2 depicts the configuration used for the skin segmentation and tissue height projection (masks I and II, respectively). The channel sizes and layers types are indicated with the concatenation operations (which skips certain layers indicated by the arrows). In this work, a transpose convolution layer substitutes a linear upsampling operation in the original architecture. The input layer for task I is the image (1024×832 pixels), normalized by the pixels from (487–537, 25–75), corresponding to a fixed small region near the breast phantoms center. For task II, the input is the same from task I, but only regions segmented as inner breast

are selected (the other values are zeroed). The output layer differs between tasks: for task I is a three channel image, with the probabilities for skin, inner breast and background, while for task II is a one channel image with the relative local height (the output layer is a sigmoidal function which limits the values from 0 to 1). Due to memory constraints, a batch of size 1 was used during training and consequently, the batch norm layers were removed from the original architecture. For both networks the ReLU activation function was used, and a weight decay (L2 penalization) of 5×10^{-5} .

For task III, an ensemble of five multi-layer perceptrons was adopted. The network consists of two fully connected layers of 200 neurons each and a linear output, as shown in Fig. 3. The procedure adopted started from the original image, where a box is drawn around the inner breast tissue segmented from mask I. Afterwards, this box is divided in images patches of 4×4 pixels. The pixel values are normalized by the number of incident photons. In total, 11 features are included: the X-ray spectra (a label varying between 0 to 9), spectrum HVL (in millimeters of aluminum) and mean energy (in keV), the breast thickness (in mm), the average relative local height, the distance between the patch and the mask's center of mass, the area covered by the breast patch on image detector, the average and standard deviation pixel intensity, and two first order statistics (kurtosis and skewness)[41]. All the features are normalized between 0 and 1. The network predicts the relative glandular height (i.e. the product of the glandular height by the relative

Table 3

Summary of the networks used in this work.

Task	Architecture	Input	Output	Loss function
I – Skin segmentation	Fig. 2	Mammography image, pixels binned 4×4 . Pixels normalized by those from a fixed region, with the average pixel value subtracted and divided by the standard deviation.	A tensor with three channels with probabilities for background, skin and inner breast.	Cross entropy
II – Relative height	Fig. 2	Mammography image, pixels binned 4×4 . Pixels normalized by the pixels closer to the center of the breast.	Matrix with the relative breast height by pixel (h , values between 0 and 1).	Mean squared error
III – Relative glandular height	Fig. 3	11 features. Pixels normalized by number of histories.	Relative glandular height (glandular height times relative height, $g \times h$).	Mean squared error

local height, $g \times h$) in that image patch, and the output values were manually limited between 0 to 1 (values predicted outside this range were rounded to the nearest acceptable value). For stability purposes, we only considered patches that have an area 100% covered by breast.

Table 3 summarizes the networks used in this work. To account for uncertainties derived from the relative height predicted by task II, we implemented a fivefold cross validation scheme (i.e. the training data is divided in 5 non-overlapping parts). Then, five networks were trained in which 4 parts are for training and one is for validation (each network has a distinct training combination validation part). For each batch interaction during training, the relative heights (h) were multiplied by a factor sampled with a normal distribution $\mathcal{N}(C, \sigma = 0.05)$, where C is specific for each each network: 0.90, 0.95, 1.00, 1.05, 1.10. The final predicted relative glandular height is calculated by averaging the output from the five networks and the variation is calculated by the standard deviation. The ReLU activation function was used, and a weight decay (L2 penalization) of 10^{-4} .

During the training of the network for task I, the skin and inner breast had a weight of 6 and 3, respectively, to compensate imbalanced dataset distributions. Moreover, the training data was augmented for tasks I and II, by flipping the image over the horizontal axis and translating the image by a random number between 0 and 50 in upwards and downwards directions, this procedure is applied for each epoch. The performance of the network from task I, besides the loss function used during training, is also benchmarked by comparing the inner breast and skin areas to the ground truth for validation and test samples. The task II is benchmarked by comparing the ground truth breast phantom volume to the predicted breast volume (excluding skin) calculated from Eq. 1. The task III performance was benchmarked by comparing the predicted (Eq. 2) and the ground truth VGF values.

For the training and validation parts, the networks are evaluated separately (i.e. compared to each mask), while for the test step, the full framework is implemented and compared to the VGF expected values. The coefficient of determination (r^2) [44] is calculated when comparing the ground truth (T) and the predicted values by the networks (P) by adjusting a linear equation: $P = a \times T + b$, and forcing $a = 1$ and $b = 0$. The best linear fit was also adjusted by determining a and setting $b = 0$. The absolute differences (Δ) and the relative differences (Δ_r) are calculated as:

$$\Delta = P - T, \Delta_r(\%) = 100 \times (P - T)/T \quad (3)$$

In this work, the result representing the output of the networks, or those calculated from these outputs will be defined as “predicted”.

Further information regarding comparison of our predicted VGF values and breast density values estimated with LIBRA tool [45] for a small phantom dataset is available in the Supplementary material.

2.5. Estimation of glandularity

After the volume glandular fraction (VGF) of the breast was estimated, the breast glandularity (G) (i.e. the percentage by mass of glandular tissue in the breast, excluding skin) was calculated as:

$$G = 100 \times \frac{VGF \times \rho_g}{(1 - VGF) \times \rho_a + VGF \times \rho_g} \quad \left(\frac{\%}{\%} \right) \quad (4)$$

where ρ_a and ρ_g are the densities of adipose and glandular tissues (0.93 g/cm^3 and 1.04 g/cm^3), respectively [37]. In our approximation, tissues that were neither glandular nor skin were classified as adipose tissue.

2.6. Dosimetry and dose levels in mammography

The mean glandular dose (MGD) for the breast phantoms was estimated as:

$$MGD = DgN \times K_{\text{air}} \quad (5)$$

where K_{air} is the incident air kerma, and DgN is the normalized glandular dose (a conversion coefficient) [2].

The DgN values were obtained using neural networks from our previous work [46] based on the following input parameters: X-ray beam (anode, filter, tube potential and HVL), breast radius and thickness, glandularity, skin thickness and compression-plate ionization chamber distance. Summarizing the process, the parameters were fed through an ensemble of multi-layer perceptrons (MLP) as it performed the regression operations and returned the DgN for each case.

The X-ray beam parameters were obtained from the input files used to generated the images (Table 2). The breast thickness is known, while the radius is calculated by approximating the breast as a semicylinder (the radius value is limited from 6 cm to 12 cm). The glandularity was obtained from Eq. 4 and the skin model was set to 1.5 mm skin. The compression-plate ionization chamber distance is equal to 40 cm (the maximum allowed distance). The reported DgN values follow the homogeneous adipose-glandular tissue distribution. More details are contained in the original paper [46].

2.7. Clinical case selection

This study employed data from a retrospective analysis of patient mammography images, acquired on routinely collected anonymous data, ethical-board approved (CAAE: 47878315. 2.0000.5404). All mammograms were acquired using AEC with the Selenia Dimensions system (Hologic, Danbury, CT, USA) which is installed at the Institute of Radiology (InRad) in the Faculty of Medicine, University of São Paulo. The data was filtered to only consider CC images. A total of 2134 clinical images were used. The HVL values in the DICOM header were matched from those used in the DgN calculations by adding 3 mm additional filtration of PMMA. The K_{air} is extracted from the DICOM header, which is believed to be a good approximation from measured values obtained with an ionization chamber [47]. We also extracted the MGD values reported by the Organ Doses tag for further comparisons.

The following procedure was adopted. For each phantom in this study, we filtered the DICOM headers that present the same X-ray spectra and the compressed breast thickness (with 2 mm tolerance).

Fig. 4. Incident air kerma values as function of breast thickness reported in the DICOM header and the respective percentile.

Afterwards, we established K_{air} intervals based on percentiles 10%, 25%, 50%, 75% and 90%, which were selected based on the breast phantom glandularity (respective intervals: $G \leq 5\%$, $5\% < G \leq 15\%$, $15\% < G \leq 25\%$, $25\% < G \leq 40\%$, $G > 40\%$), as shown in Fig. 4. Finally, K_{air} values returned for each case and the MGD was estimated from Eq. 5. From the 208 original virtual phantoms, the MGD was estimated for 132 cases, the others 76 failed due to insufficient patient data for comparison.

3. Results

3.1. Breast segmentation training and validation

The training and validation loss values as a function of the interaction number for the three networks: image segmentation, relative height and relative glandular height predictions are shown in Figs. 5 (a), (c), and (e), respectively. Figs. 5 (b), (d) and (f) show the learning rate as a function of the interaction number (in the same order as described above). As expected, the training and validation losses decrease with the learning rate, until a plateau is reached, triggering the early training stop. The validation values are more stable because they represent the average validation values for all cases on each epoch (calculated when the training iterated over all the training samples), besides they are similar to the training trend (calculated using a moving average).

The skin and inner breast segmented relative areas (ground truth $\times$ predicted) for the validation and test images are shown in Fig. 6(a) and 6(b), respectively. In this case, to facilitate the comparison, the results were normalized by the maximum value on each ground truth sample. The area marked for skin plus nipple is approximately two orders of magnitude lower than the inner breast area (respectively 7.0×10^3 pixels versus 2.5×10^5 , on average), thus a higher fluctuation is observed between the mask and the predicted values. For both validation and test, the predicted skin + nipple areas are usually higher than the ground truth, as seen by the angular coefficients a values (1.20 and 1.08, respectively) and r^2 (-0.638 and 0.284, respectively). This behavior is further explained in Fig. 7. An excellent agreement is observed regarding the inner breast area for both validation and test cases ($r^2 > 0.999$ for both), with an average relative error below 1%.

The image segmentation performed for the four randomly selected breast models from the test data are exemplified in Fig. 7. For each case,

Fig. 5. Training/validation results for: (a)/(b) breast segmentation; (c)/(d) relative height; (e)/(f) relative glandular height. A total of 60, 70 and 60 epochs are shown for (a), (c), (e), respectively. The step number is obtained by multiplying the value in the x-axis with the offset number indicated on each graph. C is the offset factor for each network of the ensemble for task III regarding the noise sampling for the relative height (Figs. (e) and (f)).

Fig. 6. (a) Validation and (b) test results for the image segmentation task. The areas are normalized by the maximum values for each region: (a) 13204 pixels for skin + nipple, 483592 pixels for inner breast; and (b) 13114 pixels for skin + nipple, 391670 pixels for inner breast. In both (a) and (b), the identity line is represented by the dashed gray line.

Fig. 7. Four randomly selected cases (with labels from (a) to (d)) illustrating the image segmentation process for the test sample.

the original simulated mammography image is compared with the ground truth mask and prediction. An interesting behavior observed is that the skin contour for the ground truth masks is not smooth, which contrasts with the predicted segmentation where the contours are smoothed out and continuous. Although the nipple is correctly identified in the four images, in some cases, we observe an over-classification inwards the breast. The validation and training results (images not shown here) also present a few cases where the nipple is only partially identified, and the other region is classified as inner breast tissue. The smoothing and the over-classification could be an explanation of the

skin + nipple area discrepancies showed in Fig. 6 and consequently low r^2 values. Nevertheless, for our application, due to the inner breast area being orders of magnitude higher than the skin + nipple region, the impact of this effect on the other results can be neglected.

The relative height prediction for the four selected breast phantoms (the same cases from Fig. 7) are shown in Fig. 8. The vertical and horizontal profile views are also compared. Cases (a) and (b) illustrate standard acquisitions with full breast compression and a smooth variation in the glandular distribution across the breast, where an excellent agreement between the mask and the predicted relative height map is

Fig. 8. Relative height prediction from the cases showed in Fig. 7. The blue and red patches represent cross sections in the horizontal and vertical directions. The respective profile views are displayed in the third row.

Fig. 9. (a) Validation and test results for volume predicted compared to the ground truth volumes derived from breast phantoms. The identity line is represented by the dashed gray line. (b) Relative differences Δ_r between the values. The dashed orange line and the shaded area indicate, respectively, the average relative difference for the test cases and one standard deviation.

observed. The cases (c) and (d) correspond to unusual virtual phantoms to assess the performance of the network. For the case (c), when a drastic variation of glandular content is present within a large region of the breast, the network interprets as a variation of the relative thickness, as shown by the discontinuity around the 0.8 relative x-axis location. The case (d) shows an example of the breast not being properly compressed, where the relative horizontal height drops smoothly as function of the pixel coordinates. For all selected cases, the nipple artifact discussed in

Fig. 7 is present in the images, since the breast was previously segmented with the first network.

The breast volume obtained with the relative height predicted maps versus the ground truth is shown in Fig. 9(a). Both validation and test results presented a good agreement with the ground truth volumes ($r^2 > 0.994$ and 0.982 , respectively). The relative differences (Δ_r) between the predicted and ground truth values are displayed in Fig. 9(b). The dashed red line and the shaded area represent, respectively, the average relative

Fig. 10. (a) Validation results for relative glandular height prediction compared to the ground truth values from the masks. The identity line is represented by the dashed gray line. (b) Distribution of the differences Δ between prediction and the ground truth values.

Fig. 11. Relative glandular height prediction from the cases shown in Fig. 7. The predicted patches are assembled to match the original mask shape.

difference and one standard deviation for test cases. The average relative difference and standard deviation are in the order of 4% for both validation and test data. The predicted breast volume is systematically underestimated for very thinner breasts (thickness near 2 cm and volumes below 250 cm³) due to the approximation of the skin layer being constant (1.5 mm thick).

Fig. 10(a) shows the relative glandular height prediction compared to the ground truth for the validation data, with the identity as the dashed gray

line. The results of relative glandular height present a good agreement, with a $r^2=0.986$. The differences Δ are quantified in Fig. 10(b), with an absolute average difference of 0.02 in the relative glandular height.

The ground truth relative glandular height, the patches calculated with the 4×4 binning mask and the network predictions for four test breasts are shown in Fig. 11. A good overall agreement was observed between the predicted and ground truth values for cases (a), (b) and (d). The incorrect height prediction showed in Fig. 8(c) causes the network to predict a slightly

Fig. 12. (a) Reconstructed volume glandular fraction (VGF) for the training and test data as function of the ground truth extracted from the breast phantoms. (b) Differences between predicted and the ground truth values. The shaded area indicates one standard deviation for test data. Bars: standard deviation of the ensemble prediction.

Fig. 13. (a) DgN relative differences for all 208 breasts calculated from the predicted glandularity and those using the median glandularity of the population, compared to the ground truth. (b) Boxplot of MGD distribution for three cases: (i) predicted by the network with 1.5 mm skin, (ii) estimated using 4 mm skin and 50% glandularity (iii) extracted from the DICOM header. Centerline: median, lower and upper box limits: 25 and 75 percentiles, whiskers: 10 and 90 percentiles, circles: outliers.

higher glandular height than the ground truth in case (c).

With the three networks trained, we implemented the full pipeline to estimate the volume glandular fraction (VGF) for the test data. For sake of completeness, we also predicted the VGF for the phantoms that were used for training since the full pipeline was not used beforehand. The results shown in Fig. 12(a) indicates a good correlation between the ground truth VGF and the predicted values, with r^2 coefficients of 0.959 and 0.935 for training and test data, respectively. The average absolute differences Δ (standard deviation) for training and test data are 0.03(4) and 0.04(5), respectively, as illustrated in Fig. 12(b). It is important to notice that the distribution of VGF, in a real population, is more concentrated towards low values, and values higher than 0.5 represent the minority of cases [3].

3.2. Breast dosimetry

The DgN conversion coefficients were estimated for all 208 phantoms after the glandularity values were calculated based on the VGF obtained with the ground truth and the ones predicted with the network. The DgN coefficients obtained with fixed glandularity given by the median glandularity of all phantoms (approximately 23%) were also calculated. Fig. 13(a) shows the relative DgN differences calculated from the predicted glandularity and those using the median glandularity compared to the ground truth. It is observed that using a fixed glandularity for the entire population introduces a systematic error that, in general, overestimates the DgN coefficients calculated compared the ground truth glandularity, on average 8.5%. Meanwhile, the DgN coefficients obtained with the glandularity calculated with the pipeline resulted in an average error of 1.3% compared to the ground truth.

Fig. 14. (a) *MGD* dose distribution from Fig. 13(b) as function of breast thickness. (b) Predicted *MGD* and Organ Dose from DICOM header trends as function of breast thickness for three intervals, bars indicate one standard deviation.

Fig. 13(b) shows the predicted *MGD* distribution for 132 test phantoms calculated using the pipeline, as described in Section 2.6, and the incident air kerma extracted from the DICOM header for clinical cases. For sake of completeness, we also included values estimated with a 4 mm skin thickness (model used for organ dose by the Hologic system [48]) and 50% glandularity for the test phantoms (Estimated) and the Organ Dose reported in the DICOM header for the entire clinical dataset. Simulated cases that did not presented a real counterpart were ignored in this process. The median (75 percentile) *MGD* values are 1.91(2.45) mGy, while the estimated and DICOM values are 1.60(2.10) mGy and 1.62(2.00) mGy, respectively. From those distributions, the dose reference levels could be extracted based on the 75 percentile values.

The predicted *MGD* as a function of the breast thickness is shown in Fig. 14(a). The apparent discontinuity for breasts thickness closer to 4 cm is due to the combination of variations on average breast compositions and incident air kerma values used in this work, considering the glandularity percentiles. Fig. 14(b) compares the *MGD* between the predicted and DICOM values for three breast thickness groups. The trend of increasing the *MGD* for thicker breasts becomes evident. It is observed a significant variation for each interval, represented by the bars (one standard deviation). This behavior can be explained because a 2 cm interval width is considered, covering a range of *MGD* values, also there is a variation of glandularity within the patients for each breast thickness interval. Nevertheless, the average *MGD* predicted values for each interval are higher than the respective Organ Dose DICOM values for all thickness range, due to the dosimetry models employed in each case.

4. Discussion

The segmentation of mammography images including edge detection and nipple removal was previously implemented by other authors using different algorithms [49,50]. In our implementation, the ground truth masks were created by establishing threshold values (only for task I) for the projected skin and nipple, which resulted in some skin and nipple boundaries to be not smooth and in some regions presenting slight discontinuities. This pattern could explain the discrepancies between the ground truth skin plus nipple areas and the predicted areas by the network. However, for the application proposed in this work, an

excellent agreement was found between the ground truth inner breast areas and the predicted ones. The performance of breast segmentation can still be improved since in a few cases, regions of the nipple were incorrectly classified as inner breast. Moreover, the breasts are imaged in a CC view, thus the pectoral muscle is not present and consequently, the inclusion of medium lateral oblique projections would require the additional segmentation of muscle tissue [51]. Since the exact shape of the breast cannot be reconstructed from a single 2D image, a model that approximates the breast with a 1.5 mm skin layer was adopted. This approximation according to our comparisons, induced systematic errors specially for thinner breasts that were compensated with correction factors using a polynomial fit. From the results, it was showed that the overall prediction performance is better for firmly compressed breasts and with phantoms with a more homogeneous glandular distribution, as those achieved in high quality clinical mammography images for real patients. Although the relative glandular height maps presented a lower resolution than a native mammography image, it did not interfere significantly on the overall VGF predictions. The highest errors were observed for unusual cases where the compression was partially incomplete or the phantoms presented an extremely heterogeneous glandular tissue distribution. We are studying, for future versions, to include the compression force as an input parameter to address some of these issues, since the correct measure of breast thickness is an important factor for an accurate breast density estimations [52]. We successfully trained a network to predict the relative glandular height within the breast, which showed a good performance for the validation and test samples. An interesting aspect observed in this work was the limitation of the technique when tissues different from glandular and adipose are present in the phantoms. The predicted values tend to overestimates the glandular height since these other tissues have higher attenuation coefficients than adipose tissue and more similar to glandular tissue. The framework is purposely divided in three parts to facilitate future implementations that requires one specific task. Although this work focused primarily on images generated from Monte Carlo simulations, preliminary tests (not included) showed that the deep learning algorithm performed equally under different quantum noise levels. For a relative dose varying from 0.1 to 10 times the original dose, the predicted VGD varied less than 0.01 for a nominal 0.12 VGF.

Moreover, the MC code tries to simulate a real mammography image, as previously stated in other works [31,30]. A future work is planned to apply this algorithm for real mammography breast images.

From the volume glandular fraction values predicted by our framework, it was possible to calculate the glandularity and finally the DgN factors using neural networks previously trained [46] for the 208 virtual breast phantoms generated in this work. We found 1.3% of difference, on average, between the DgN values from the predicted glandularity and the ground truth DgN , with a maximum below 8%. We also compared the case where the median glandularity (23%) of the phantoms is considered, which resulted on average error of 8.5%, and a maximum error of 35%. Naturally, these errors are propagated to the MGD calculations and could induce bias on the dose reference levels. In terms of radioprotection, it is desirable to minimize potential biases and errors for the dose estimation. With incidente air kerma values extracted from DICOM mammography images, we estimated the MGD for 132 phantoms and found a median (75 percentile) values of 1.89(2.44) mGy. On the other hand, by employing a traditional model of 4 mm thick skin used for Organ Dose estimation by the Hologic system [48] and 50% (constant) glandularity, the reported values are 1.60(2.10) mGy, which are closer to the Organ Dose tag values 1.62(2.00) mGy. This highlights the importance of knowing which model is being considered for the MGD estimations when comparing dose levels, since some models could result in systematically higher or lower doses than others [8]. It is important to notice that the model used for calculating the MGD value reported by the Organ Dose tag within the DICOM header can vary between manufactures [48].

The results presented in this work are based on anthropomorphic virtual phantoms and Monte Carlo simulations, thus they are based on the characteristics of the virtual breasts generated by the software and the limitations of the simulations. It offers the advantage of knowing the ground truth values and, to easily label the training data, which is a good environment for testing the concept of the deep learning proposed in this work. However, it is still an approximation and does not reflect completely a clinical situation, requiring further studies and refinements before applying for clinical cases. Therefore, this approach of determining the breast density with deep learning models trained on virtual phantoms shows promising preliminary results and it could be a complementary method over the existing ones, since it uses a different methodology of determining breast density with virtual breast phantoms, and not relying on physical models developed by other algorithms. Additionally, the framework could be adapted for other tasks and improved over time, with different neural network architectures or expanding the training dataset. Since the images were generated from

virtual patients, there is the advantage of openly sharing and distributing the database and the algorithm.

5. Conclusions

In this work we introduce a deep learning framework to estimate VGF values from 208 virtual breast phantoms, with promising preliminary results. From the predicted VGF values, the DgN coefficients were calculated by using another neural network. Afterwards, the mean glandular dose values are obtained from air kerma extracted from DICOM header for a clinical cohort. This approach enables to establish dose reference levels more accurately for a given population than using an average glandularity model. This framework could be adapted for other applications (e.g. image processing and segmentation) by employing transfer learning techniques, especially in cases where the training dataset is small. Moreover, future studies can explore this implementation for real mammography units and clinical images, and compare the results with other glandularity estimation algorithms. Finally, application for deriving patient-specific doses need different approaches, since the vertical location of glandular tissue cannot be determined by using single view mammography images. The database and the deep learning algorithm used in this study are freely available by request to the corresponding author (atomal@ifi.unicamp.br).

Acknowledgments

This work was supported by the Brazilian agencies: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, project number 2015/21873–8); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ, project number 140155/2019–8); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, finance code 001). The authors thank Dr. JM Fernández-Varea, Dr. PR Costa and MSc. HR Mendes for the assistance on the edition of this manuscript.

Appendix A. Calibration for breast volume and VGF

Fig. A.15(a) shows the ground truth breast volume compared to the mask volume calculated from Eq. (1), while Fig. A.15(b) compares the breast volumes with and without skin. Fig. A.16(a) shows the relation between the ground truth VGF calculated directly from the phantoms and the reconstructed VGF using the masks calculated from Eq. (2). Fig. A.16(b) quantifies the ratio between the ground truth and the

Fig. A.15. Comparison between: (a) the ground truth breast volume and the ones calculated using the height mask, and (b) the ground truth breast volumes with and without skin.

Fig. A.16. (a) Comparison between the ground truth *VGF* and the one calculated using the relative glandular height mask. (b) Ratio between the ground truth and the mask *VGF* (*rVGF*) as function of breast thickness (*t*, in cm), fitted with a third order polynomial (dashed line). $rVGF = 3.51 \times 10^{-4} \times t^3 - 8.96 \times 10^{-3} \times t^2 + 7.58 \times 10^{-2} \times t + 0.79$.

reconstructed *VGF* obtained from the masks as function of breast thickness. The results are for the training sample (178 phantoms).

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ejmp.2021.03.007>.

References

- [1] Sardanelli F, Aase HS, Álvarez M, Azavedo E, Baarslag HJ, Balleyguier C, et al. Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France. *G. Eur Radiol* 2017;27(7):2737–43. <https://doi.org/10.1007/s00330-016-4612-z>.
- [2] Dance DR, Sechopoulos I. Dosimetry in x-ray-based breast imaging. *Phys Med Biol* 2016;61(19):R271–304. <https://doi.org/10.1088/0031-9155/61/19/R271>.
- [3] Yaffe MJ, Boone JM, Packard N, Alonzo-Proulx O, Huang SY, Peressotti CL, et al. The myth of the 50–50 breast. *Med Phys* 2009;36(12):5437–43. <https://doi.org/10.1118/1.3250863>.
- [4] Hernandez AM, Seibert JA, Boone JM. Breast dose in mammography is about 30% lower when realistic heterogeneous glandular distributions are considered. *Med Phys* 2015;42(11):6337–48. <https://doi.org/10.1118/1.4931966>.
- [5] Sechopoulos I, Bliznakova K, Qin X, Fei B, Feng SSJ. Characterization of the homogeneous tissue mixture approximation in breast imaging dosimetry. *Med Phys* 2012;39(8):5050–9. <https://doi.org/10.1118/1.4737025>.
- [6] Huang SY, Boone JM, Yang K, Packard NJ, McKenney SE, Prionas ND, et al. The characterization of breast anatomical metrics using dedicated breast CT. *Med Phys* 2011;38(4):2180–91. <https://doi.org/10.1118/1.3567147>.
- [7] Caballo M, Boone JM, Mann R, Sechopoulos I. An unsupervised automatic segmentation algorithm for breast tissue classification of dedicated breast computed tomography images. *Med Phys* 2018;45(6):2542–59. <https://doi.org/10.1002/mp.12920>.
- [8] Sarno A, Mettievier G, Di Lillo F, Bliznakova K, Sechopoulos I, Russo P. Homogeneous vs. patient specific breast models for Monte Carlo evaluation of mean glandular dose in mammography. *Phys Med* 2018. <https://doi.org/10.1016/j.ejmp.2018.04.392>. 51:56–63.
- [9] Arana Peña LM, Fedon C, García E, Díaz O, Longo R, Dance DR, et al. Monte Carlo dose evaluation of different fibroglandular tissue distribution in breast imaging. In: Van Ongeval C, Marshall N, Bosmans H, editors. 15th International Workshop on Breast Imaging (IWBI2020); vol. 11513. SPIE; 2020, p. 76. <https://dx.doi.org/10.1117/12.2564278>.
- [10] Geeraert N. Quantitative evaluation of fibroglandular tissue for estimation of tissue differentiated absorbed energy in breast tomosynthesis. Ph.D. thesis; Télécom ParisTech/TSI; 2014.
- [11] Teuwen J, Moriakov N, Fedon C, Caballo M, Reiser I, Bakic P, et al. Deep learning reconstruction of digital breast tomosynthesis images for accurate breast density and patient-specific radiation dose estimation. *Med Image Anal.* 2021;102061. <https://doi.org/10.1016/j.media.2021.102061>.
- [12] di Franco F, Sarno A, Mettievier G, Hernandez AM, Bliznakova K, Boone JM, et al. GEANT4 Monte Carlo simulations for virtual clinical trials in breast X-ray imaging: Proof of concept. *Phys Med* 2020;74:133–42. <https://doi.org/10.1016/j.ejmp.2020.05.007>.
- [13] Fedon C, Caballo M, García E, Díaz O, Boone JM, Dance DR, et al. Fibroglandular tissue distribution in the breast during mammography and tomosynthesis based on breast ct data: A patient-based characterization of the breast parenchyma. *Med Phys* 2021;48(3):1436–47. <https://doi.org/10.1002/mp.14716>.
- [14] Sechopoulos I, Boone JM, Dance D, van Engen R, Russo P, Young KC. Mammography dose estimates do not reflect any specific patient's breast dose. *Eur J Radiol* 2020;131. <https://doi.org/10.1016/j.ejrad.2020.109216>.
- [15] Power G, Manley M, Baldelli P, Keavey E, Phelan N. Breast thickness based DRLs in screening mammography. *Phys Med* 2019;67:203. <https://doi.org/10.1016/j.ejmp.2019.09.204>.
- [16] Alonzo-Proulx O, Mawdsley GE, Patrie JT, Yaffe MJ, Harvey JA. Reliability of automated breast density measurements. *Radiology* 2015;275(2):366–76. <https://doi.org/10.1148/radiol.15141686>.
- [17] Ekpo EU, Hogg P, Highnam R, McEntee MF. Breast composition: Measurement and clinical use. *Radiography* 2015;21(4):324–33. <https://doi.org/10.1016/j.radi.2015.06.006>.
- [18] Ng KH, Lau S. Vision 20/20: Mammographic breast density and its clinical applications. *Med Phys* 2015;42(12):7059–77. <https://doi.org/10.1118/1.4935141>.
- [19] Lee J, Nishikawa RM. Automated mammographic breast density estimation using a fully convolutional network. *Med Phys* 2018;45(3):1178–90. <https://doi.org/10.1002/mp.12763>.
- [20] Lizzi F, Laruina F, Oliva P., Retico A., Fantacci M.E. Residual convolutional neural networks to automatically extract significant breast density features. In: Vento M., Percannella G., editors. *Computer Analysis of Images and Patterns (CAIP 2019)*; vol. 1089. 2019, p. 28–35.
- [21] Ionescu GV, Fergie M, Berks M, Harkness EF, Hulleman J, Brentnall AR, et al. Prediction of reader estimates of mammographic density using convolutional neural networks. *J Med Imag* 2019;6(03):1. <https://doi.org/10.1117/1.jmi.6.3.031405>.
- [22] Warren LM, Harris P, Gomes S, Trumble M, Halling-Brown MD, Dance DR, et al. Deep learning to calculate breast density from processed mammography images. In: Van Ongeval C, Marshall N, Bosmans H, editors. 15th International Workshop on Breast Imaging (IWBI2020); vol. 11513. SPIE; 2020, p. 24. <https://dx.doi.org/10.1117/12.2561278>.
- [23] Maghsoudi O.H., Gastouniotti A., Scott C., Pantalone L., Wu F.F., Cohen E.A., et al. Deep-libra: Artificial intelligence method for robust quantification of breast density with independent validation in breast cancer risk assessment. 2020. arXiv: 2011.08001.
- [24] Graff CG. In: Kontos D, Flohr TG, Lo JY, editors. *Medical Imaging 2016: Physics of Medical Imaging*, 9783. SPIE; 2016. p. 978309. <https://dx.doi.org/10.1117/1.2.2216312>.
- [25] Hernandez AM, Becker AE, Boone JM. Updated breast CT dose coefficients (DgN CT) using patient-derived breast shapes and heterogeneous fibroglandular distributions. *Med Phys* 2019;46(3):1455–66. <https://doi.org/10.1002/mp.13391>.
- [26] Huang SY, Boone JM, Yang K, Kwan ALC, Packard NJ. The effect of skin thickness determined using breast CT on mammographic dosimetry. *Med Phys* 2008;35(4):1199–206. <https://doi.org/10.1118/1.2841938>.
- [27] Sharma D, Graff CG, Badal A, Zeng R, Sawant P, Sengupta A, et al. Technical Note: In silico imaging tools from the VICTRE clinical trial. *Med Phys* 2019;46(9):3924–8. <https://doi.org/10.1002/mp.13674>.
- [28] Maas SA, Ellis BJ, Ateshian GA, Weiss JA. FEBio: Finite elements for biomechanics. *J Biomech Eng* 2012;134(1). <https://doi.org/10.1115/1.4005694>.

- [29] Badal A, Badano A. Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit. *Med Phys* 2009;36(11):4878–80. <https://doi.org/10.1118/1.3231824>.
- [30] Badano A, Graff CG, Badal A, Sharma D, Zeng R, Samuelson FW, et al. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. *JAMA Network Open* 2018;1(7):e185474. <https://doi.org/10.1001/jamanetworkopen.2018.5474>.
- [31] Badal A, Sharma D, Graff CG, Zeng R, Badano A. Mammography and breast tomosynthesis simulator for virtual clinical trials. *Comput Phys Commun* 2020;:107779. <https://doi.org/10.1016/j.cpc.2020.107779>.
- [32] Cunha DM, Tomal A, Poletti ME. Evaluation of scatter-to-primary ratio, grid performance and normalized average glandular dose in mammography by Monte Carlo simulation including interference and energy broadening effects. *Phys Med Biol* 2010;55(15):4335–59. <https://doi.org/10.1088/0031-9155/55/15/010>.
- [33] Day GJ, Dance DR. X-ray transmission formula for antiscatter grids. *Phys Med Biol* 1983;28(12):1429–33. <https://doi.org/10.1088/0031-9155/28/12/008>.
- [34] Bick U, Diekmann F. Digital Mammography. Medical Radiology. Berlin, Heidelberg: Springer; 2010. <https://doi.org/10.1007/978-3-540-78450-0>.
- [35] Sechopoulos I, Ali ESM, Badal A, Badano A, Boone JM, Kyprianou IS, et al. Monte Carlo reference data sets for imaging research: Executive summary of the report of AAPM Research Committee Task Group 195. *Med Phys* 2015;42(10):5679–91. <https://doi.org/10.1118/1.4928676>.
- [36] Salvat F, PENELOPE-2018: A Code System for Monte Carlo Simulations of Electron and Photon Transport. OECD/NEA: Issy-les-Moulineaux; 2019. <https://doi.org/10.1787/32da5043-en>.
- [37] Hammerstein RG, Miller DW, White DR, Masterson ME, Woodard HQ, Laughlin JS. Absorbed Radiation Dose in Mammography. *Radiology* 1979;130(2):485–91. <https://doi.org/10.1148/130.2.485>.
- [38] Woodard HQ, White DR. The composition of body tissues. *Brit J Radiol* 1986;59(708):1209–18. <https://doi.org/10.1259/0007-1285-59-708-1209>.
- [39] Hernandez AM, Seibert JA, Nosrati A, Boone JM. Generation and analysis of clinically relevant breast imaging x-ray spectra. *Med Phys* 2017;44(6):2148–60. <https://doi.org/10.1002/mp.12222>.
- [40] Siddon RL. Fast calculation of the exact radiological path for a three-dimensional CT array. *Med Phys* 1985;12(2):252–5. <https://doi.org/10.1118/1.595715>.
- [41] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* 2020;17:261–72. <https://doi.org/10.1038/s41592-019-0686-2>.
- [42] Bullock J, Cuesta-Lázaro C, Quera-Bofarull A. In: Gimi B, Krol A, editors. Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging, 10953. International Society for Optics and Photonics; SPIE; 2019. p. 453–63. <https://doi.org/10.1117/12.2512451>.
- [43] Paszke A, Gross S, Massa F, Lerer A, Bradbury J., Chanan G., et al. Pytorch: An imperative style, high-performance deep learning library. In: Wallach H., Larochelle H., Beygelzimer A., d'Alché-Buc F., Fox E., Garnett R., editors. Advances in Neural Information Processing Systems 32. Curran Associates, Inc.; 2019. p. 8024–8035.
- [44] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12:2825–30. <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>.
- [45] Keller BM, Nathan DL, Wang Y, Zheng Y, Gee JC, Conant EF, et al. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. *Med Phys* 2012;39(8):4903–17. <https://doi.org/10.1118/1.4736530>.
- [46] Trevisan Massera R, Tomal A. Estimation of glandular dose in mammography based on artificial neural networks. *Phys Med Biol* 2020;65(9):095009. <https://doi.org/10.1088/1361-6560/ab7a6d>.
- [47] Gennaro G, Bigolaro S, Hill ML, Stramare R, Caumo F. Accuracy of mammography dosimetry in the era of the European Directive 2013/59/Euratom transposition. *Eur J Radiol* 2020;127:108986. <https://doi.org/10.1016/j.ejrad.2020.108986>.
- [48] Suleiman ME, Brennan PC, McEntee MF. Mean glandular dose in digital mammography: a dose calculation method comparison. *J Med Imag* 2017;4(1):13502. <https://doi.org/10.1117/1.JMI.4.1.013502>.
- [49] Tzikopoulos SD, Mavroforakis ME, Georgiou HV, Dimitropoulos N, Theodoridis S. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. *Comput Meth Prog Bio* 2011;102(1):47–63. <https://doi.org/10.1016/J.CMPB.2010.11.016>.
- [50] Feng SSJ, Patel B, Sechopoulos I. Objective models of compressed breast shapes undergoing mammography. *Med Phys* 2013;40(3):031902. <https://doi.org/10.1118/1.4789579>.
- [51] Ma X, Wei J, Zhou C, Helvie MA, Chan HP, Hadjiiski LM, et al. Automated pectoral muscle identification on MLO-view mammograms: Comparison of deep neural network to conventional computer vision. *Med Phys* 2019;46(5):2103–14. <https://doi.org/10.1002/mp.13451>.
- [52] Tyson AH, Mawdsley GE, Yaffe MJ. Measurement of compressed breast thickness by optical stereoscopic photogrammetry. *Med Phys* 2009;36(2):569–76. <https://doi.org/10.1118/1.3065066>.