

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Volumetric breast density measurement for personalized screening: accuracy, reproducibility, consistency, and agreement with visual assessment

Andreas Fieselmann
Daniel Förnvik
Hannie Förnvik
Kristina Lång
Hanna Sartor
Sophia Zackrisson
Steffen Kappler
Ludwig Ritschl
Thomas Mertelmeier

SPIE.

Andreas Fieselmann, Daniel Förnvik, Hannie Förnvik, Kristina Lång, Hanna Sartor, Sophia Zackrisson, Steffen Kappler, Ludwig Ritschl, Thomas Mertelmeier, "Volumetric breast density measurement for personalized screening: accuracy, reproducibility, consistency, and agreement with visual assessment," *J. Med. Imag.* **6**(3), 031406 (2019), doi: 10.1117/1.JMI.6.3.031406.

Volumetric breast density measurement for personalized screening: accuracy, reproducibility, consistency, and agreement with visual assessment

Andreas Fieselmann,^{a,*} Daniel Förnvik,^b Hannie Förnvik,^b Kristina Lång,^{b,c} Hanna Sartor,^b Sophia Zackrisson,^b Steffen Kappler,^a Ludwig Ritschl,^a and Thomas Mertelmeier^a

^aSiemens Healthcare GmbH, Forchheim, Germany

^bLund University and Skåne University Hospital, Department of Translational Medicine, Malmö, Sweden

^cInstitute for Biomedical Engineering, ETH Zurich, Zurich, Switzerland

Abstract. Assessment of breast density at the point of mammographic examination could lead to optimized breast cancer screening pathways. The onsite breast density information may offer guidance of when to recommend supplemental imaging for women in a screening program. A software application (Insight BD, Siemens Healthcare GmbH) for fast onsite quantification of volumetric breast density is evaluated. The accuracy of the method is assessed using breast tissue equivalent phantom experiments resulting in a mean absolute error of 3.84%. Reproducibility of measurement results is analyzed using 8427 exams in total, comparing for each exam (if available) the densities determined from left and right views, from cranio-caudal and medio-lateral oblique views, from full-field digital mammograms (FFDM) and digital breast tomosynthesis (DBT) data and from two subsequent exams of the same breast. Pearson correlation coefficients of 0.937, 0.926, 0.950, and 0.995 are obtained. Consistency of the results is demonstrated by evaluating the dependency of the breast density on women's age. Furthermore, the agreement between breast density categories computed by the software with those determined visually by 32 radiologists is shown by an overall percentage agreement of 69.5% for FFDM and by 64.6% for DBT data. These results demonstrate that the software delivers accurate, reproducible, and consistent measurements that agree well with the visual assessment of breast density by radiologists. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JMI.6.3.031406]

Keywords: breast density; mammography; breast cancer; personalized screening.

Paper 18215SSR received Sep. 28, 2018; accepted for publication Dec. 27, 2018; published online Feb. 5, 2019.

1 Introduction

1.1 Clinical Background

Breast density is an important topic in breast cancer screening because of two aspects. First, a high amount of dense (fibroglandular) breast tissue is considered to be an independent risk factor for developing breast cancer, and second, sensitivity of mammography is lower in dense breasts due to the masking effect.¹ Supplemental imaging for dense breasts [e.g., breast ultrasound, breast magnetic resonance imaging (MRI)] can thus be useful to increase cancer detection rates in breast cancer screening.² Nowadays, the majority of the US states require that women are informed if they have dense breast tissue and they will receive information about supplemental imaging options.^{3,4}

Radiologists typically estimate breast density during interpretation of the mammograms. However, visual breast density assessment is known to have considerable intra- and inter-reader variability.⁵ Automated breast density assessment by computer software is increasingly used to assist radiologists in reporting breast density more objectively and consistently.

The time when breast density is assessed has a considerable impact on a personalized screening work flow. When breast density is assessed by radiologists, the woman usually has

already left the screening center. Supplemental imaging requires the woman to be called back for an extra assessment.

If, however, automated breast density assessment is provided during the screening appointment, then the work flow might be sped up considerably. If supplemental imaging is recommended, it could be initiated before the woman leaves the screening center (Fig. 1). The women could get the result of the recommended supplemental imaging test on the day of screening, which would reduce psychological distress. This procedure, though, would require an organizational change in scheduling, a problem that should be solvable after having gained experience and after a transition to routine.

1.2 Automated Breast Density Assessment

Several techniques for automated breast density measurement from mammographic x-ray images have been proposed.^{6,7} These techniques either calculate the projected two-dimensional (2-D) area (in cm²) of dense tissue in the x-ray image or quantify the three-dimensional (3-D) volume (in cm³) of the dense tissue in the breast. The calculation of the projected 2-D area of dense tissue requires a segmentation of the dense tissue areas in the image.⁸ For quantification of the 3-D volume of the dense tissue, the physics of the image acquisition process is modeled and it can involve either a precalibration of the system,⁹ a calibration object in the acquired image,¹⁰ or an image-based self-calibration step.¹¹

*Address all correspondence to Andreas Fieselmann, E-mail: andreas.fieselmann@siemens-healthineers.com

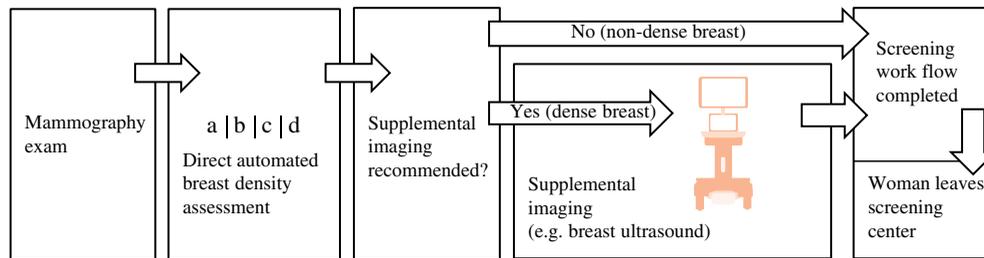


Fig. 1 Screening workflow with breast density assessment at the time of examination allowing to initiate supplemental imaging before the woman leaves the screening center.

A breast density (percentage) value can be computed by dividing the area or volume of dense tissue with the total area or volume of the breast. Because areal and volumetric breast density (VBD) values are computed from different quantities, they have different value ranges and they cannot be compared directly.¹²

Breast density classification has a high clinical relevance. A breast density category can be determined from a measured breast density value by using cut points. Furthermore, machine learning-based algorithms exist that assign a breast density category directly based on extracted image features, such as parenchymal texture or histogram information.⁷ Recently, also, deep machine learning techniques have been applied for breast density classification.¹³

For using an automated breast density assessment software for clinical decision support, it should be validated comprehensively. Ng and Lau⁷ have identified six requirements (denoted in their paper as “sanity checks”) that should be fulfilled by an automated breast density measurement software:

1. “Density should be the same for the identical image of the breast.”
2. “Density should be similar for a breast no matter what the view, in particular cranio-caudal (CC) and medio-lateral oblique (MLO) views.”
3. “Density should be similar for the same breast no matter the imaging equipment, in particular, it should not matter if the equipment is GE, Siemens, Hologic, or if the imaging is done on mammography, tomosynthesis, MRI, or CT.”
4. “Density should be invariant to breast compression.”
5. “Left and right breast densities should be highly correlated but not identical.”
6. “Density should, over a population, generally reduce with age.”

Some studies exist that validate existing software applications for automated breast density assessment.^{14–19} Typical aspects assessed by these studies are accuracy (comparing measured breast density to an objective ground truth), reproducibility (see Ng and Lau’s requirements 1 to 5), consistency (see Ng and Lau’s requirement 6), and agreement with visual assessment (comparing classified breast density to a subjective reference).

Recently, an automated breast density measurement software (Insight BD, Siemens Healthcare GmbH) has been integrated into the acquisition work station of a mammography system (MAMMOMAT Revelation, Siemens Healthcare GmbH; Insight BD and MAMMOMAT Revelation are not commercially

available in all countries. Due to regulatory reasons, the future availability cannot be guaranteed). This allows objective evaluation of breast density directly after the mammographic exam. In this work, we evaluate performance of the software Insight BD to measure VBD. In particular, we evaluate whether the software satisfies the six requirements identified by Ng and Lau. This paper is a substantially expanded version of a previously published conference paper.²⁰

2 Material and Methods

2.1 Volumetric Breast Density Measurement

Insight BD measures VBD based on a physics model of the image acquisition process and an image-based self-calibration.^{11,21} This model appears to be the basis of most commercial software implementations to assess breast density.⁷ The model assumes that the breast consists of two types of tissue, fibroglandular and fatty tissue, with known energy-dependent x-ray attenuation values.

The algorithm receives an unprocessed full-field digital mammogram (FFDM) or an unprocessed central digital breast tomosynthesis (DBT) projection image as input along with the image acquisition parameters such as compressed breast thickness and peak tube voltage. For each detector pixel location, the amount of fibroglandular tissue (measured in mm) located above the pixel is calculated and a 2-D breast density map is created. The total amount of fibroglandular tissue (V_{fg} , measured in cm^3) is determined from the map by numerical integration over the projected breast area. The volume of the breast, V_{breast} , is determined using the known compressed thickness of the breast, its projected surface area, and a 3-D shape model. For determining V_{fg} and V_{breast} , the pectoral muscle region is excluded.

The VBD is calculated by dividing V_{fg} by the total breast volume (V_{breast} , measured in cm^3):

$$\text{VBD} = 100\% \times \frac{V_{fg}}{V_{\text{breast}}}. \quad (1)$$

Breast density categories (a, b, c, and d) correlating with those from ACR BI-RADS fifth ed. atlas²² are assigned by using VBD cut points. To aid classification between categories “b” and “c,” considered to be a nondense and dense breast, respectively, the distribution of dense tissue is also taken into account as described by Fieselmann et al.²¹

The software framework used in this work consists of a core module for the VBD measurement and a wrapper around this module allowing for batch processing of many image files at once. The core module is the same as the one implemented in the Insight BD application of the MAMMOMAT Revelation mammography system.

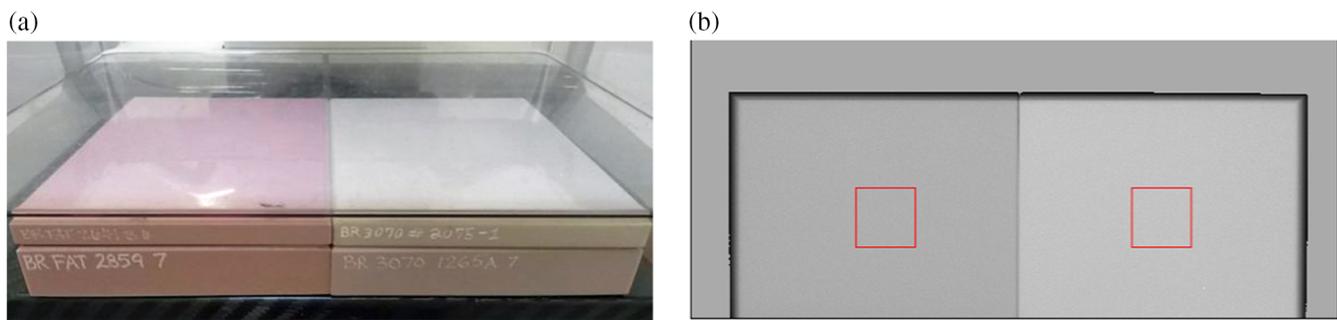


Fig. 2 Setup for phantom-based accuracy evaluation of VBD measurement: (a) photograph, (b) breast density map with measurement regions indicated by red squares.

2.2 Evaluation of Accuracy

Accuracy of breast density measurement is evaluated using phantoms with physical characteristics similar to that of breast tissue (phototimer compensation plates; CIRS Inc., Norfolk, Virginia).

Plates simulating 100% fatty breast tissue and plates simulating fibroglandular breast tissue are placed on the left and right sides, respectively, of the breast support table (Fig. 2). Different glandularities (right side only: 30%, 50%, and 70%) and plate heights (left and right sides: 30 mm, 50 mm, and 70 mm) lead to nine different combinations for evaluation. Images are acquired with a MAMMOMAT Inspiration mammography system (Siemens Healthcare GmbH) using W/Rh anode/filter combination, antiscatter grid in place and automatic exposure control enabled. The tube voltages are chosen automatically depending on the compression paddle height.

The average VBD is measured in two square regions of interest (side length 27.54 mm, Fig. 2) in the 2-D breast density map, one placed in the fatty tissue region and one placed in the fibroglandular tissue region. To measure the accuracy, two quantities are computed. The mean absolute deviation [MAD, measured in percentage points (pp)] is defined as

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^N |x_i - y_i|. \quad (2)$$

The mean absolute percentage error (MAPE, measured in %) is calculated as

$$\text{MAPE} = 100\% \times \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - y_i}{x_i} \right|. \quad (3)$$

In the above equations, x_i and y_i ($i = 1, \dots, N$) denote the known ground truth values and the measured values, respectively. N denotes the number of samples.

2.3 Evaluation of Reproducibility

2.3.1 Same woman, different FFDM views

The reproducibility of the breast density measurement is evaluated using different FFDM views of the same woman. 8150 exams were selected from the Malmö Breast Tomosynthesis Screening Trial (MBTST).²³ The exams were selected based on the availability of raw data (for processing FFDM images and DBT projection images) in the data base. Selection of cases was not influenced by a woman's breast cancer status. Characteristics of this data set are shown in

Table 1 in column "data set 1." Each exam contains anonymized four-view FFDM raw images. The MBTST is an ethics committee-approved prospective trial investigating the accuracy of DBT in a population-based screening program in the city of Malmö in Sweden.²³ In this trial, four FFDM images (CC and MLO views of each breast) and two DBT scans (MLO view of each breast) have been acquired from each participant with a MAMMOMAT Inspiration system.

The average VBD value of both views (CC and MLO) of the left breast is compared with the average VBD of both views of the right breast. This assumes that there is bilateral mammographic density symmetry, and left and right breast densities are correlated but not identical. Furthermore, the average VBD from both CC views of the exam is compared to the average VBD from both MLO views of the exam. Reproducibility is quantified using Pearson correlation coefficient (PCC)²⁴ and the MAD of VBD values.

Similarly, reproducibility of breast volume measurement is evaluated. Reproducibility of fibroglandular tissue volume is not assessed, for the sake of brevity, because it is not statistically independent from VBD and breast volume.

2.3.2 Same woman, FFDM and DBT exams

In the second reproducibility evaluation, FFDM and DBT exams of the same woman acquired during the same breast compression are analyzed. Two data sets are used in this analysis: one data set (denoted as "data set 2" in Table 1) contains 108 exams acquired with a MAMMOMAT Inspiration in Tokyo, Japan; the other data set (denoted as "data set 3" in Table 1) contains 95 exams acquired with a MAMMOMAT Inspiration in Vienna, Austria. For each exam, anonymized four-view FFDM raw images and anonymized four-view DBT projection images are available.

Breast density measures (VBD and breast volume) are calculated by taking the average of the values of all views using FFDM and DBT data, respectively. Reproducibility is quantified using PCC and MAD between the sample values from the two imaging modalities.

2.3.3 Same woman, two FFDM acquisitions

In the third reproducibility evaluation, two FFDM acquisitions of the same woman acquired during the same breast compression with a MAMMOMAT Inspiration are analyzed. The first image was acquired with the antiscatter grid in place; the second one was acquired without antiscatter grid but with software-based scatter correction and reduced x-ray dose. The exams were part of an ethics committee-approved study,²⁵ and 74 anonymized image pairs are available. This data set is denoted as

Table 1 Characteristics of the data sets used for the evaluations.

	Data set 1 (FFDM)	Data set 2 (FFDM + DBT)	Data set 3 (FFDM + DBT)	Data set 4 (FFDM)
Number of exams	8150	108	95	74
Mean patient age at the time of exam	57 ± 9 years (39 to 75 years)	Not available	57 ± 12 years (28 to 85 years)	57 ± 6 years (50 to 72 years)
Mean compressed thickness (FFDM)	51 ± 14 mm (7 to 109 mm)	46 ± 13 mm (19 to 89 mm)	60 ± 13 mm (29 to 91 mm)	57 ± 15 mm (19 to 87 mm)
Mean compression force (FFDM)	112 ± 22 N (24 to 193 N)	89 ± 11 N (48 to 130 N)	93 ± 34 N (28 to 191 N)	63 ± 16 N (39 to 122 N)

“data set 4” in Table 1. This evaluation allows assessment of reproducibility of breast density measurement when different image acquisition conditions (with and without antiscatter grid) are employed.

2.4 Evaluation of Consistency

The calculated breast density in a large population is analyzed with respect to the women’s age. With postmenopausal alteration of fibroglandular breast tissue, it is expected that the density of a woman’s breast will decrease with increasing age.²⁶

The images from data set 1 (Table 1) are used for this analysis. Breast density is calculated from all 8150 four-view exams on a per breast basis (averaging results from CC and MLO views) giving 16,300 separate values for VBD and breast density category, respectively. Mean and standard deviation of VBD as well as frequency of breast density categories are calculated depending on a woman’s age.

2.5 Evaluation of Agreement with Radiologists’ Visual Assessment

600 four-view anonymized FFDM exams had been randomly selected from the MBTST, and 32 experienced radiologists from the US and Canada provided individual breast density classifications for these exams according to the ACR BI-RADS® fifth ed. atlas. Nine radiologists labeled the first set of 200 exams (set “1 to 200”), 10 radiologists labeled the second set of 200 exams (set “201 to 400”), and 13 radiologists labeled the third set of 200 exams (set “401 to 600”). The most frequently chosen category for a certain exam is defined to be the reference density category by the radiologists for this exam (panel majority vote).²¹

The software calculates density categories for each of the 600 FFDM exams. 512 exams have DBT raw projection images (MLO views only) available that were acquired in a different breast compression. For these DBT exams, density categories are calculated by the software as well. These categories, determined by the software using FFDM and DBT exams, respectively, are compared to the reference density categories by the radiologists. Overall percentage agreement and Cohen’s linearly weighted kappa²⁷ values are computed.

3 Results

3.1 Evaluation of Accuracy

The results for the accuracy evaluation are shown in Fig. 3. The measured VBD values are plotted against the ground truth

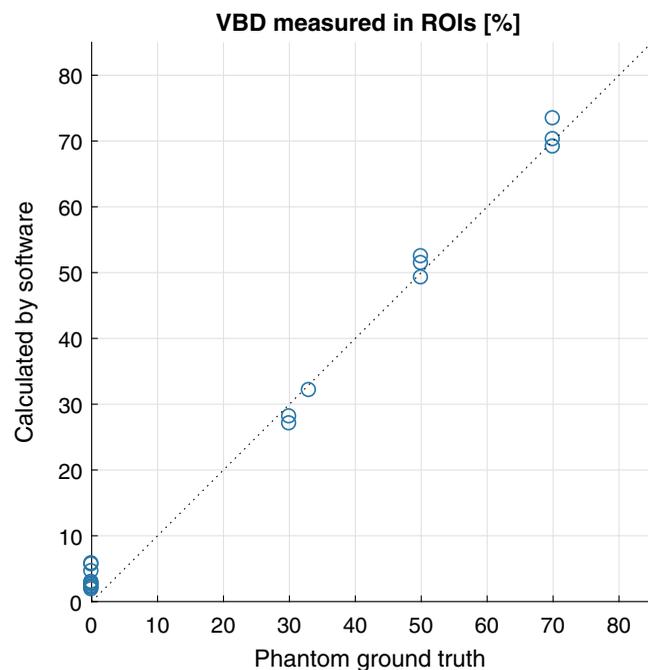


Fig. 3 Results for accuracy evaluation by phantom experiments.

VBD values. One sample point has a ground truth VBD value of 33% instead of 30%, which corresponds to 60-mm plate height with 30% glandularity plus 10-mm plate height with 50% glandularity. The MAD are 3.38 and 1.65 pp for the fatty tissue and dense tissue regions, respectively. MAPE is 3.84% for the dense tissue region. MAPE was not calculated for the fatty tissue region as the denominator would be zero.

3.2 Evaluation of Reproducibility

Scatter plots for the reproducibility evaluation are shown in Figs. 4 and 5 (different FFDM views; color encodes density of sample values), Figs. 6 and 7 (FFDM and DBT exams), and Fig. 8 (two FFDM acquisitions). PCC and MAD values are shown inside the plots.

3.3 Evaluation of Consistency

In Fig. 9, the breast density per breast depending on the age at examination is shown. It is presented as VBD value and as breast density category dichotomized as nondense (a, b) and

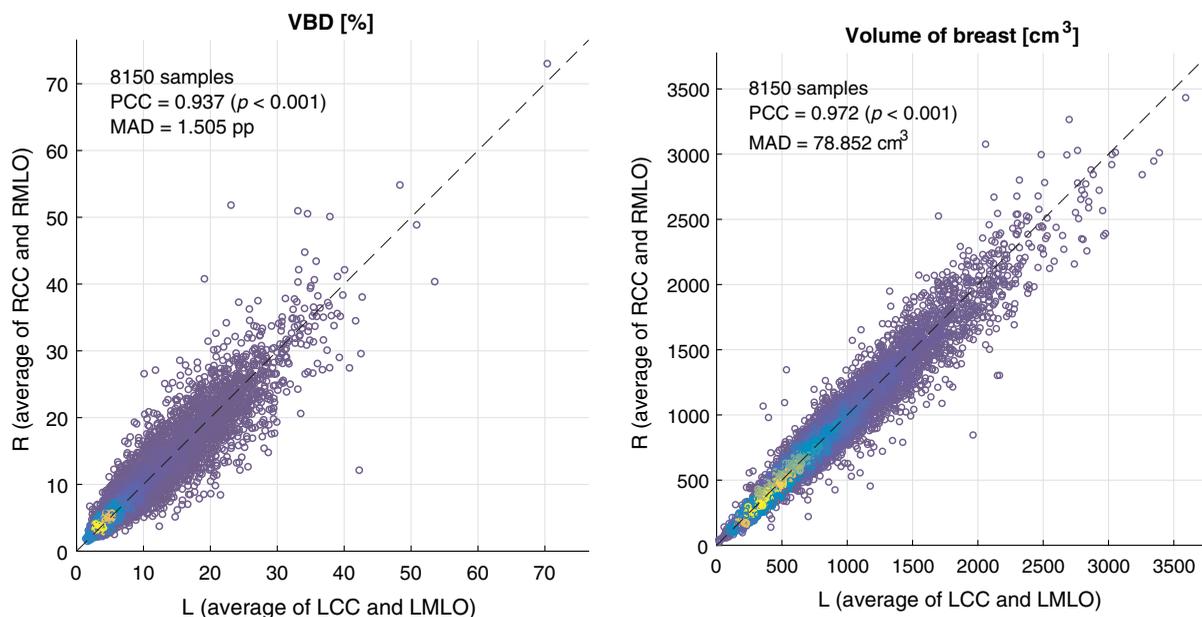


Fig. 4 Results for reproducibility evaluation with data set 1 (left versus right breast).

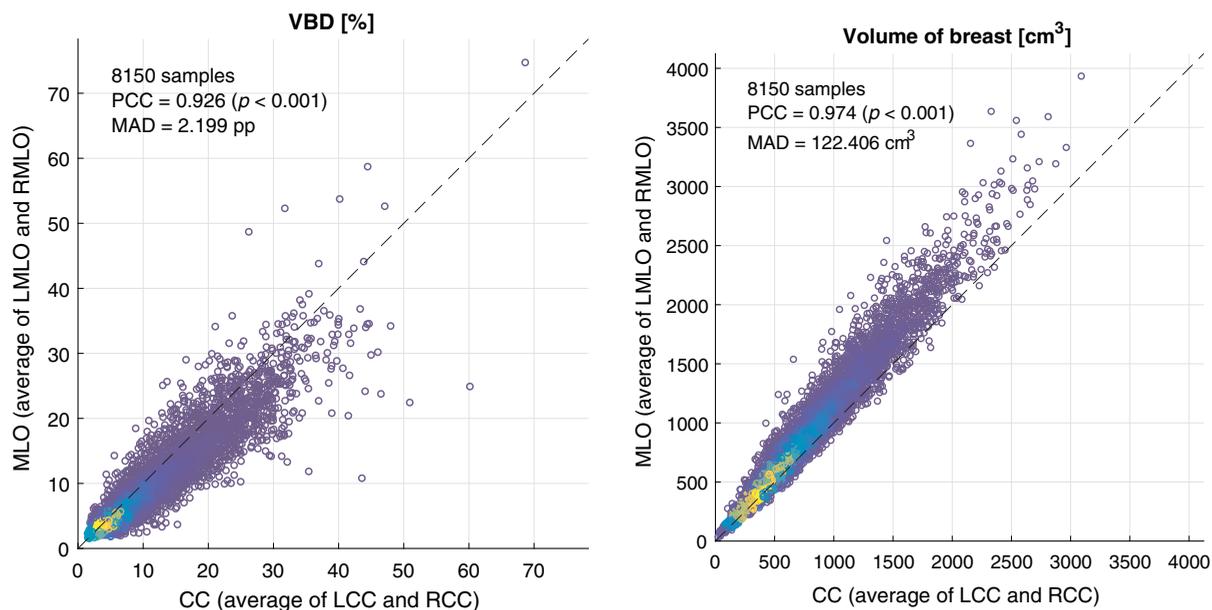


Fig. 5 Results for reproducibility evaluation with data set 1 (CC versus MLO view).

dense (c, d) categories. A histogram of the age at examination for data set 1 is shown in Fig. 10. Proportions of calculated breast density categories for different age groups are shown in Fig. 11. For comparison, proportions for the same age groups as reported by the Breast Cancer Surveillance Consortium (BCSC) are shown as well.

3.4 Comparison with Radiologists' Visual Assessment

Confusion matrices for the evaluation of agreement of the software density categories with radiologists' visual assessment are shown in Table 2. The overall percentage agreement is 69.5% [Cohen's linearly weighted kappa (κ_{lw}) = 0.67] for the categories based on FFDM images and 64.6% (κ_{lw} = 0.59) for

the categories based on DBT projections. When the results are dichotomized into nondense (a, b) and dense (c, d) categories, the agreement is 88.5% (κ_{lw} = 0.76) and 83.0% (κ_{lw} = 0.64), respectively.

4 Discussion

Different studies were carried out to evaluate the performance of breast density measurement with Insight BD. Each evaluation had a focus on one of these four aspects: accuracy, reproducibility, consistency, or agreement with visual assessment. Table 3 displays the results obtained in our evaluations in combination with results from previously published studies using existing software for automated breast density measurement to support interpretation and comparison of our results. A strength of our

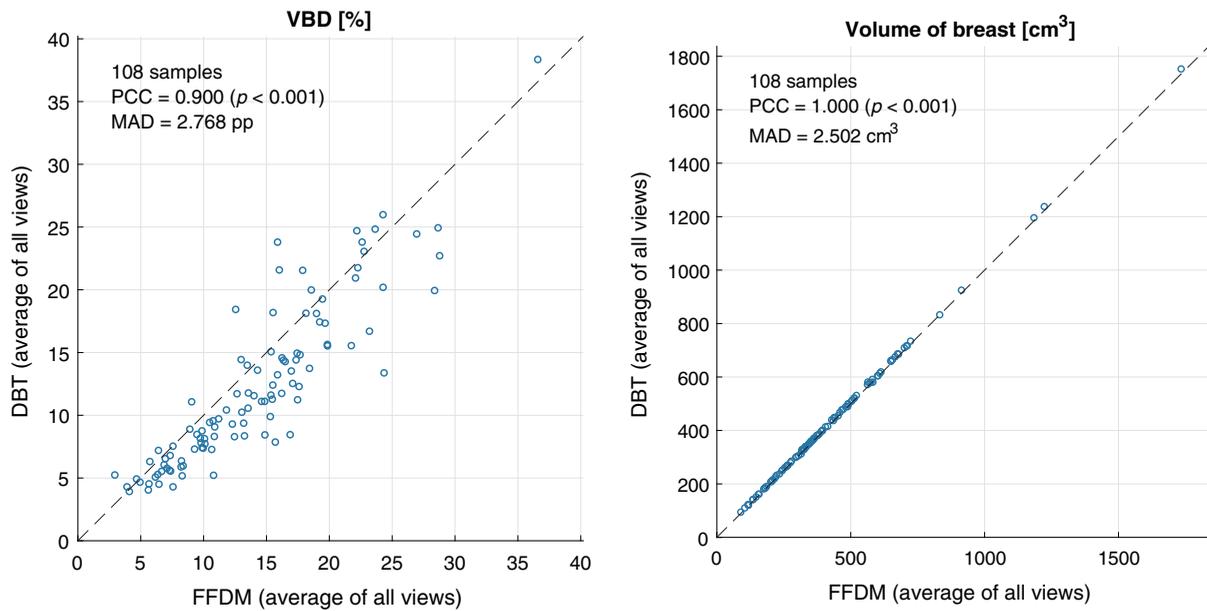


Fig. 6 Results for reproducibility evaluation with data set 2 (FFDM versus DBT).

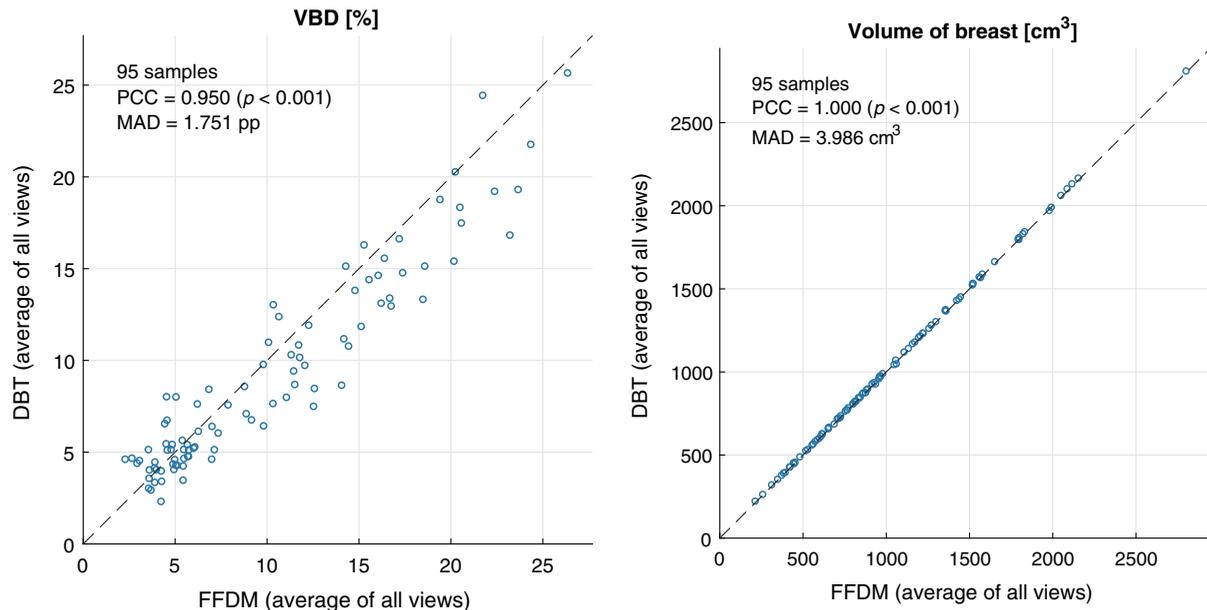


Fig. 7 Results for reproducibility evaluation with data set 3 (FFDM versus DBT).

work is that it addresses all these different relevant aspects of validation in one study.

Accuracy was evaluated based on phantom data, where the breast density is known. An important aspect is the linearity of measured quantities. As can be seen from Fig. 3, the calculated quantities show a high level of linearity. Our results are comparable to those shown in a previous study,¹⁴ where an MAD of 1.1 pp and an MAPE of 6.94% were obtained (Table 3). The current evaluation is based on phantoms that have a very homogeneous distribution of fibroglandular tissue and not a realistic shape in the breast periphery region. It is known that algorithms for breast density assessment may not work well for phantoms that do not have realistic compressed breast edge shapes.⁷ Therefore, we have used regions of interest in the central breast area for the analysis to avoid effects caused by the unrealistic

shape in the breast periphery. Only for this phantom analysis, the evaluation is restricted to the central breast area. In clinical breast images, the full breast is evaluated. Future studies could also evaluate phantoms with a more realistic heterogeneous distribution of fibroglandular tissue.

Reproducibility was evaluated based on clinical data using three different experimental setups. A strong correlation between the results from the left and right breast and also between the two views of the same breast is evident. It should be considered that an existing or developing breast cancer in the exam images may influence the correlation values. However, the cancer prevalence in the data set 1 is expected to be low (breast cancer was detected in 137 of 14,848 women participating in the MBTST²³), and this influence is considered to be negligible.

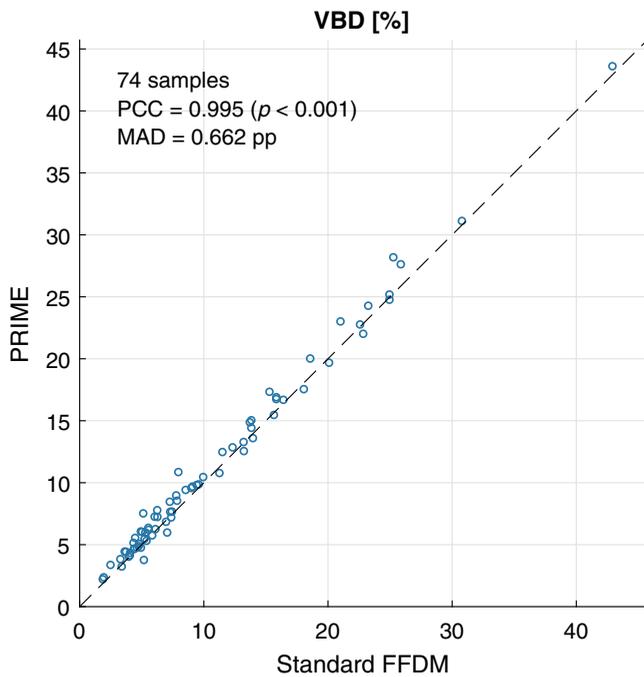


Fig. 8 Results for reproducibility evaluation with data set 4 (two FFDMs, same view).

Breast volume is slightly higher when estimated from MLO views compared to CC views (Fig. 5), which can be explained by the different ways the breast is positioned and visible in the mammograms. In data sets 2 and 3, FFDM and DBT images were acquired in the same breast compression. The estimation of breast volume is thus not influenced by breast positioning, and the breast volume shows a higher correlation compared to the results from data set 1.

For the measurements described in Secs. 2.3.1 and 2.3.2, previous studies exist that show similar correlation values (Table 3). The study by Förnvik et al.²⁹ also investigated the agreement between VBD calculated from FFDM and DBT data based on an initial prototype version of the software assessed in

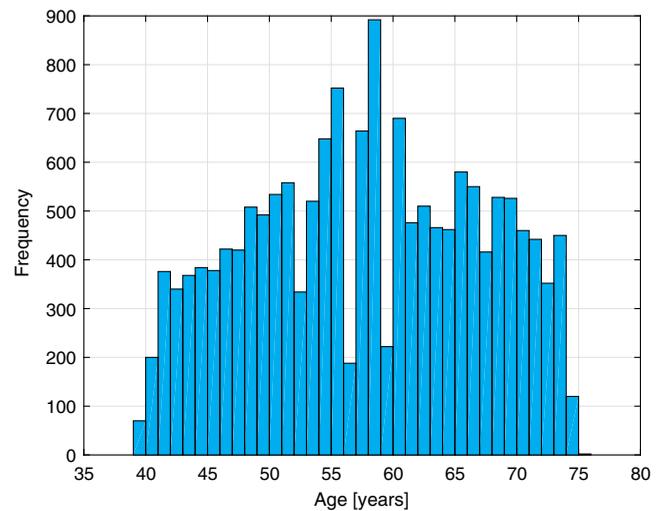


Fig. 10 Histogram of the age at examination for data set 1.

this work. The results in that study (Spearman’s correlation coefficient²⁴ = 0.83) were based on a different data set but indicate high correlation as do the results from our study (PCC = 0.900 to 0.950). For the setup described in Sec. 2.3.3, no previous publication could be identified.

Consistency was evaluated based on a sample of 8150 four-view FFDM images. Results show that VBD decreases with age until it reaches a steady state at about 60 years of age (Fig. 9). Also, the frequency of the classification with breast density category “c” or “d” decreases with age until about 60 years of age (Fig. 9). These results are consistent with the expected behavior that the density of a woman’s breasts will decrease with increasing age.²⁶ Studies evaluating consistency of breast density calculation using other breast density measurement software have also shown a decrease of the woman’s breast density until 60 to 65 years of age.¹⁶ The trend visible in the proportions of calculated breast density categories depending on age group is also similar to the trend visible in the data from the BCSC (Fig. 11). Small differences in the proportions may be explained by the different screening populations

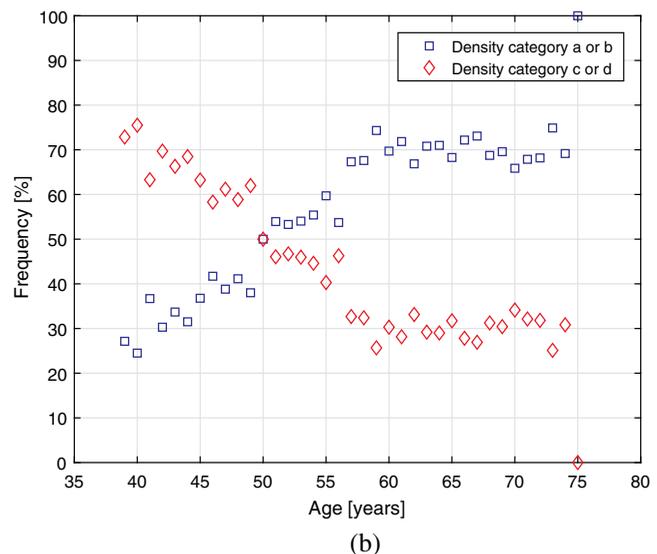
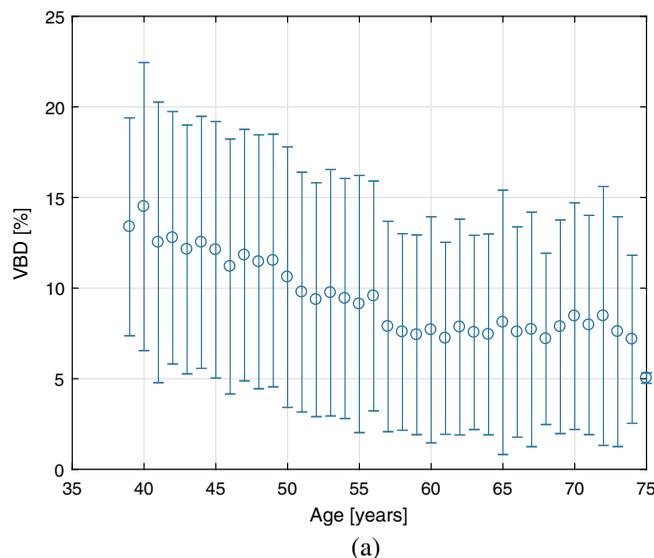


Fig. 9 Breast density per breast depending on age at examination: (a) mean and standard deviation of VBD (b) dichotomous breast density classification.

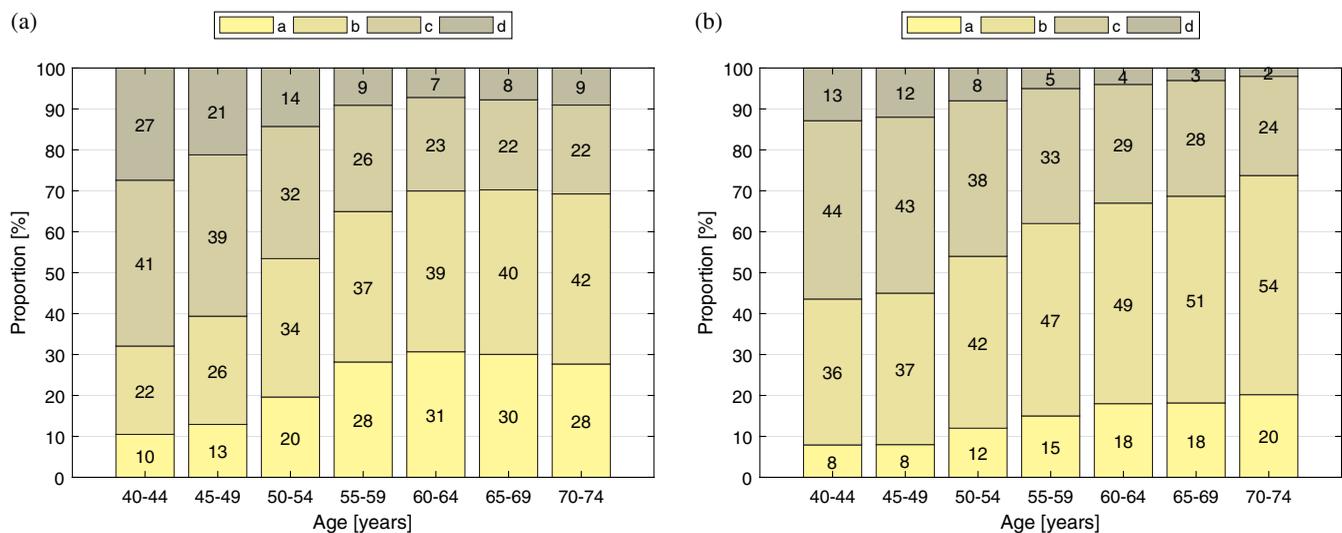


Fig. 11 Proportions of calculated breast density categories depending on age group: (a) our evaluation, (b) for comparison, data from BCSC.²⁸

Table 2 Confusion matrices showing agreement between density categories determined by radiologists and by the software. The software used either FFDM or DBT exam data as input data.

Radiologists' panel majority vote → software density category (from FFDM) ↓	a	b	c	d
a	71	61	0	0
b	25	174	31	1
c	1	36	136	7
d	0	0	21	36

Radiologists' panel majority vote → software density category (from DBT) ↓	a	b	c	d
a	49	33	2	0
b	33	155	38	2
c	1	44	101	7
d	0	0	21	26

(Sweden and USA). The study by Förnvik et al.²⁹ investigated dependency of mean VBD on age using a subset of the MBTST data and a weak correlation has been found (Spearman's correlation coefficient ranging from -0.28 to -0.20). This result is consistent with the results from our study analyzing the age-dependent proportions of breast density categories as well.

The evaluation of agreement with radiologists' visual assessment is based on the radiologists' categories according to the ACR BI-RADS[®] fourth ed. atlas in the comparison studies (Table 3) and the more recent ACR BI-RADS[®] fifth ed. atlas in our study. Our results for the radiologists' agreement are similar to those reported in a previous study³⁰ (four category agreement: 63% to 70%). In that study, an initial prototype version of the software assessed in this work has been evaluated, and the labels were provided by Swedish radiologists.

In Sec. 1.2, the six requirements identified by Ng and Lau for an automated breast density measurement software are quoted. Requirement 1 is satisfied by Insight BD since it is based on a deterministic algorithm. The results from the reproducibility evaluations (Sec. 3.2) show that requirement 2 (density values for CC and MLO views are similar), requirement 3 (density values obtained with mammography and tomosynthesis are similar), and requirement 5 (density values for the left and right breast are highly correlated) are satisfied as well. Requirement 4 has been evaluated implicitly and is also met: in the data sets used for the evaluations, the mean breast compression force was different (Table 1). Finally, requirement 6 is also fulfilled: over a population, breast density values decrease with age as expected (Sec. 3.3).

To conclude, a performance evaluation of Insight BD has been carried out to provide a comprehensive performance assessment of this software. It could be shown that this software satisfies all six requirements identified in the work by Ng and Lau.⁷ It may provide onsite breast density measurement in the exam room for screening pathway guidance. The integration of the software into the acquisition work station of the mammography system makes this information directly available to the radiographer. Other existing software applications for breast density measurement provide this information primarily to the radiologist during image interpretation.

A limitation of this work is that it focuses on a pure technical performance evaluation. The practical impact of onsite breast density evaluation on a screening workflow has not been investigated. Furthermore, the evaluation of accuracy is limited to experiments with simple phantoms. In future studies, accuracy could be evaluated using more realistic breast phantoms and also involve tomographic images (e.g., from breast MRI) providing the ground truth data for comparison.

5 Summary

A software application for VBD measurement (Insight BD) has been evaluated. The results of the performance evaluation show that the software delivers accurate, reproducible, and consistent results that correlate well with the visual assessment done by radiologists. As a feature, this software is directly integrated into the acquisition work station of the mammography system.

Table 3 Summary of results from our evaluations and comparison with published data (SW: software used in previously published study).

Type of evaluation	Results from this study	Comparison with previous studies
Accuracy (phantom tests)	Abs. error (MAD): 1.65 to 3.38 pp rel. error (MAPE): 3.84%	Abs. error (MAD): 1.1 pp rel. error (MAPE): 6.94% (Highnam et al., ¹⁴ SW: Volpara™ 1.2.1)
Reproducibility (left versus right, CC versus MLO)	PCC = 0.937 (left versus right) PCC = 0.926 (CC versus MLO)	PCC = 0.923 (left versus right) PCC = 0.915 (CC versus MLO) (Highnam et al., ¹⁴ SW: Volpara™ 1.2.1)
Reproducibility (FFDM versus DBT)	PCC = 0.900 to 0.950 (four views)	PCC = 0.91 (four views) (Machida et al., ¹⁵ SW: Volpara™ 1.5.1)
Reproducibility (two FFDMs, same view)	PCC = 0.995	Not available
Consistency (calculated breast density versus age)	VBD decreases with age as expected	VBD decreases with age as expected (Highnam et al., ¹⁶ SW: Volpara™ 1.4)
Agreement with radiologists' visual assessment (a–d / nondense versus dense)	64.6% to 69.5% / 83.0% to 88.5%	63.9% to 70.1% / 81.8% to 88.8% (Highnam et al., ¹⁴ SW: Volpara™ 1.2.1) 17.7% / 52.7% (Gubern-Mérida et al., ¹⁷ SW: Volpara™ 1.4.3) 64.9% / 91.5% (Gweon et al., ¹⁸ SW: Volpara™ 1.5.1) 57.1% / 82.2% (Sartor et al., ¹⁹ SW: Volpara™ 1.5.11)

This makes automated breast density measurements in the exam room possible and may allow for a 1-day screening work flow including supplemental imaging for women with dense breasts.

Disclosures

Kristina Lång, Hanna Sartor, and Sophia Zackrisson received grants and speaking fees from Siemens Healthcare GmbH. Andreas Fieselmann, Steffen Kappler, Thomas Mertelmeier, and Ludwig Ritschl are employed by Siemens Healthcare GmbH.

References

- N. F. Boyd et al., "Mammographic density and the risk and detection of breast cancer," *N. Engl. J. Med.* **356**(3), 227–236 (2007).
- B. Wilczek et al., "Adding 3D automated breast ultrasound to mammography screening in women with heterogeneously and extremely dense breasts: report from a hospital-based, high-volume, single-center breast cancer screening program," *Eur. J. Radiol.* **85**(9), 1554–1563 (2016).
- S. Maimone and M. McDonough, "Dense breast notification and supplemental screening: a survey of current strategies and sentiments," *Breast J.* **23**(2), 193–199 (2017).
- M. J. Emaus et al., "MR imaging as an additional screening modality for the detection of breast cancer in women aged 50–75 years with extremely dense breasts: the DENSE trial study design," *Radiology* **277**(2), 527–537 (2015).
- B. L. Sprague et al., "Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study," *Ann. Intern. Med.* **165**(7), 457–464 (2016).
- E. U. Ekpo and M. F. McEntree, "Measurement of breast density with digital breast tomosynthesis—a systematic review," *Br. J. Radiol.* **87**(1043), 20140460 (2014).
- K. Ng and S. Lau, "Vision 20/20: mammographic breast density and its clinical applications," *Med. Phys.* **42**(12), 7059–7077 (2015).
- W. He et al., "A review on automatic mammographic density and parenchymal segmentation," *Int. J. Breast Cancer* **2015**(276217), 31 (2015).
- J. J. Heine, K. Cao, and D. E. Rollison, "Calibrated measures for breast density estimation," *Acad. Radiol.* **18**(5), 547–555 (2011).
- S. Malkov et al., "Single x-ray absorptiometry method for the quantitative mammographic measure of fibroglandular tissue volume," *Med. Phys.* **36**(12), 5525–5536 (2009).
- S. van Engeland et al., "Volumetric breast density estimation from full-field digital mammograms," *IEEE Trans. Med. Imaging* **25**(3), 273–282 (2006).
- R. Highnam et al., "Comparing measurements of breast density," *Phys. Med. Biol.* **52**(19), 5881–5895 (2007).
- C. D. Lehman et al., "Mammographic breast density assessment using deep learning: clinical implementation," *Radiology* **290**(1), 52–58 (2019).
- R. Highnam et al., "Robust breast composition measurement—Volpara™," *Lect. Notes Comput. Sci.* **6136**, 342–349 (2010).
- Y. Machida et al., "Automated volumetric breast density estimation out of digital breast tomosynthesis data: feasibility study of a new software version," *SpringerPlus* **5**(1), 780 (2016).
- R. Highnam et al., "Breast density into clinical practice," *Lect. Notes Comput. Sci.* **7361**, 466–473 (2012).
- A. Gubern-Mérida et al., "Volumetric breast density estimation from full-field digital mammograms: a validation study," *PLoS One* **9**(1), e85952 (2014).
- H. M. Gweon et al., "Radiologist assessment of breast density by BI-RADS categories versus fully automated volumetric assessment," *Am. J. Roentgenol.* **201**(3), 692–697 (2013).
- H. Sartor et al., "Measuring mammographic density: comparing a fully automated volumetric assessment versus European radiologists' qualitative classification," *Eur. Radiol.* **26**(12), 4354–4360 (2016).
- A. Fieselmann et al., "Volumetric breast density measurement for personalized screening: accuracy, reproducibility, and agreement with visual assessment," *Proc. SPIE* **10718**, 107180A (2018).
- A. Fieselmann, A. K. Jerebko, and T. Mertelmeier, "Volumetric breast density combined with masking risk: enhanced characterization of breast density from mammography images," *Lect. Notes Comput. Sci.* **9699**, 486–492 (2016).
- E. A. Sickles et al., *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*, ch. ACR BI-RADS® Mammography, 5th ed., American College of Radiology, Reston, Virginia (2013).
- S. Zackrisson et al., "One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study," *Lancet Oncol.* **19**(11), 1493–1503 (2018).
- K. H. Zou, K. Tuncali, and S. G. Silverman, "Correlation and simple linear regression," *Radiol.* **227**(3), 617–628 (2003).

25. A. Fieselmann et al., "Full-field digital mammography with grid-less acquisition and software-based scatter correction: investigation of dose saving and image quality," *Proc. SPIE* **8668**, 86685Y (2013).
26. C. M. Checka et al., "The relationship of mammographic density and age: implications for breast cancer screening," *Am. J. Roentgenol.* **198**(3), W292–W295 (2012).
27. J. Cohen, "Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit," *Psychol. Bull.* **70**(4), 213–220 (1968).
28. B. L. Sprague et al., "Prevalence of mammographically dense breasts in the United States," *J. Natl. Cancer Inst.* **106**(10), 1–6 (2014).
29. D. Förnvik et al., "Comparison between software volumetric breast density estimates in breast tomosynthesis and digital mammography images in a large public screening cohort," *Eur. Radiol.* **29**, 330–336 (2019).
30. P. Timberg et al., "Breast density assessment using breast tomosynthesis images," *Lect. Notes Comput. Sci.* **9699**, 197–202 (2016).

Biographies of the authors are not available.