



A new automated method to evaluate 2D mammographic breast density according to BI-RADS® Atlas Fifth Edition recommendations

Corinne Balleyguier¹ · Julia Arfi-Rouche¹ · Bruno Boyer¹ · Emilien Gauthier² · Valerie Helin² · Ara Loshkajian¹ · Stephane Ragusa² · Suzette Delalogue¹

Received: 5 March 2018 / Revised: 13 December 2018 / Accepted: 17 January 2019 / Published online: 15 February 2019
© European Society of Radiology 2019

Abstract

Objectives Radiologists' visual assessment of breast mammographic density (BMD) is subject to inter-observer variability. We aimed to develop and validate a new automated software tool mimicking expert radiologists' consensus assessments of 2D BMD, as per BI-RADS V recommendations.

Methods The software algorithm was developed using a concept of Manhattan distance to compare a patient's mammographic image to reference mammograms with an assigned BMD category. Reference databases were built from a total of 2289 pairs (cranio-caudal and medio-lateral oblique views) of 2D full-field digital mammography (FFDM). Each image was independently assessed for BMD by a consensus of radiologists specialized in breast imaging. A validation set of additional 800 image pairs was evaluated for BMD both by the software and seven blinded radiologists specialized in breast imaging. The median score was used for consensus. Software reproducibility was assessed using FFDM image pairs from 214 patients in the validation set to compare BMD assessment between left and right breasts.

Results The software showed a substantial agreement with the radiologists' consensus (unweighted $\kappa = 0.68$, 95% CI 0.64–0.72) when considering the four breast density categories, and an almost perfect agreement (unweighted $\kappa = 0.84$, 95% CI 0.80–0.88) when considering clinically significant non-dense (A-B) and dense (C-D) categories. Correlation between left and right breasts was high ($r_s = 0.87$; 95% CI 0.84–0.90).

Conclusions BMD assessment by the software was strongly correlated to radiologists' consensus assessments of BMD. Its performance should be compared to other methods, and its clinical utility evaluated in a risk assessment model.

Key Points

- A new software tool assesses breast density in a standardized way.
- The tool mimics radiologists' clinical assessment of breast density.
- It may be incorporated in a breast cancer risk assessment model.

Keywords Mammography · Breast density · Breast cancer · Risk assessment · Software

Abbreviations

BI-RADS Breast Imaging Reporting and Data System
BMD Breast mammographic density

CC Cranio-caudal
DBT Digital breast tomosynthesis
Md Manhattan distance
MLO Medio-lateral oblique
MQSA Mammography Quality Standards Act

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-019-06016-y>) contains supplementary material, which is available to authorized users.

✉ Corinne Balleyguier
corinne.balleyguier@gustaveroussy.fr

¹ Imaging Department, Gustave Roussy, 114 rue Edouard Vaillant, 94800 Villejuif, France

² Predlife, Espace Maurice Tubiana, 39 rue Camille Desmoulins, 94800 Villejuif, France

Introduction

Breast mammographic density (BMD) refers to variations in the radiological appearance of the breast tissue due to differences in tissue composition, namely the fibroglandular and fatty tissues. BMD is associated with two risks: the risk of delaying tumor detection due to tumor masking by dense

tissue on mammograms and the increased risk of developing breast cancer. Indeed, BMD has been shown to be a strong risk factor for breast cancer, women with very dense breasts being as much as four to six times more likely to develop breast cancer than those with low breast density [1–5]. In the USA, breast density notification laws have been passed in more than 20 states, requiring physicians to inform patients of their breast density following mammography and its implication with regard to tumor detection and breast cancer risk, and to discuss appropriate screening recommendations, such as supplemental ultrasonography.

BMD is routinely assessed by radiologists and commonly classified using the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) breast density score. The new fifth edition recommendations focus on the masking role of breast density on mammography by defining the categories not by the percentage of dense tissue, but by the dispersion of dense tissue throughout the breast, as follows: (a) almost entirely fatty, (b) scattered fibroglandular density, (c) heterogeneously dense, and (d) extremely dense [6]. However, visual assessments have shown considerable intra- and inter-observer variability that could lead to discordant assessments of dense/non-dense status [7–9]. In order for breast density to be a useful measure in clinical decision-making, whether to identify mammograms that are likely to mask tumors or to predict future risk of developing breast cancer, an objective, reproducible method of BMD assessment method is clearly needed [10]. Several computerized tools have been developed to standardize BMD measurements from mammographic images, including two-dimensional area-based assessments and the more recent volumetric assessments, which are considered to more accurately measure the volume of fibroglandular tissue by taking into account breast thickness [11, 12]. Two automated volumetric assessment software programs are commercially available: Quantra (Hologic) and Volpara (Volpara Solutions). Previous studies on automated volumetric assessment tools have shown variable results in predicting the risk of breast cancer [13–18]. Whether they add value to or can reliably replace clinical BI-RADS assessment has been the subject of much investigation [13, 14].

In a recent study, a breast cancer risk assessment model incorporating both continuous measures of breast density using Volpara and BI-RADS clinical assessment was shown to have improved discriminatory accuracy compared to models with either measure alone, suggesting that different aspects of BMD are measured by quantitative and qualitative methods, all of which are relevant in assessing breast cancer risk [17]. It is currently not known which method most accurately measures BMD and should be used in clinical decision-making [19]. Given the different aspects of BMD, we sought to exploit and optimize what is captured by visual assessment—the quantity as well as distribution of dense

tissue—in an automated, standardized manner, so that its main shortcoming that is the variability would be minimized.

We describe here the development and validation of a new automated method, DenSeeMammo, which mimics radiologists' consensus assessments of BMD based on digital comparison and recognition of mammographic images, combining the advantages of both qualitative and quantitative methods. We evaluated its performance and compared its measurements to a consensus of radiologists' visual assessment, considered as the ground truth.

Materials and methods

Industry support

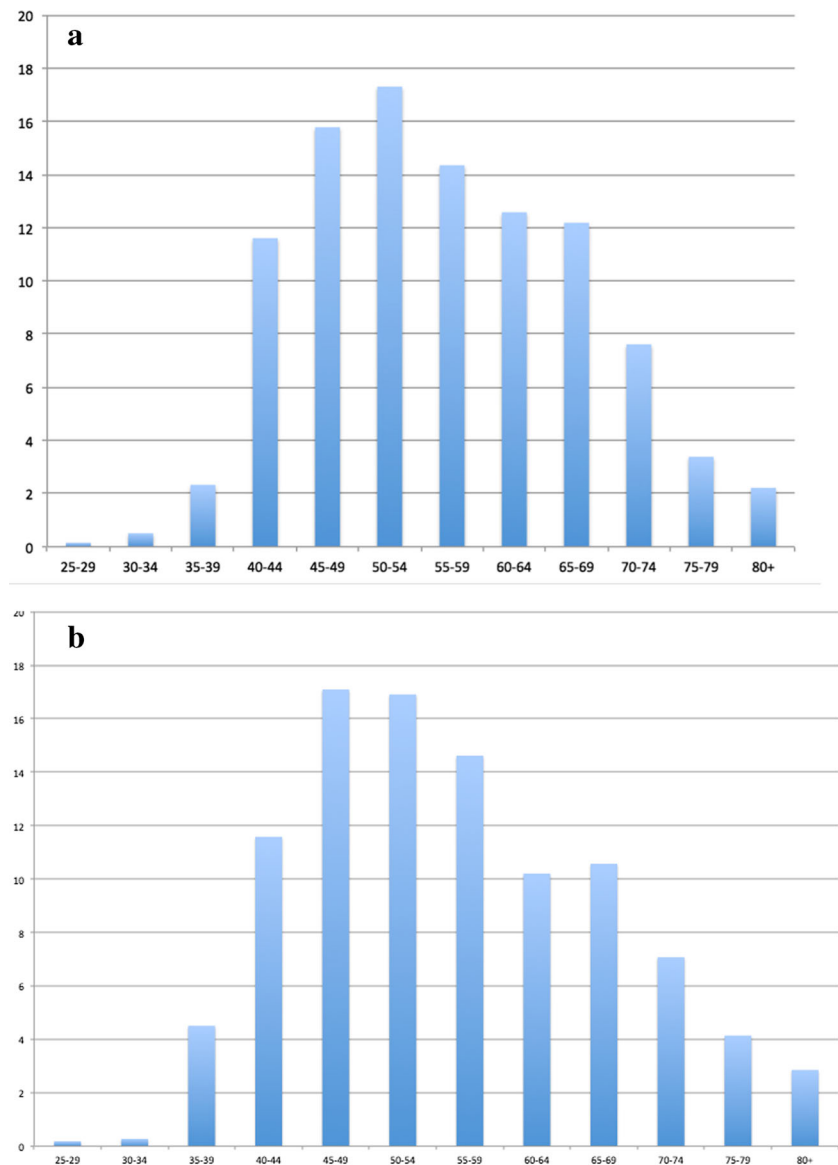
Predlife, which developed the DenSeeMammo algorithm, did not support the study, but provided their software tool for the study. Non-employee authors had complete control of the data and information that might present a conflict of interest to the authors who are employees of Predlife.

Description of the reference database

The reference database was built from consecutive pairs of anonymized and processed “for presentation” 2D full-field digital mammography (FFDM) images from patients undergoing a mammography exam in 2016, either in a radiology practice in Paris, France (1200 image pairs), performed with Senographe Essential (GE Healthcare), or in a radiology department in Monaco (1089 image pairs), performed with Selenia Dimensions (Hologic). Each pair of images consisted of two views of the breast: cranio-caudal (CC) and medio-lateral oblique (MLO). All types of exams were collected for the reference database as long as the breast did not look altered on the image (breast prosthesis, breast surgery, biopsy clips, large microcalcifications were not included). The distribution by age reflected what is expected in the screening population in the USA and Europe (Fig. 1a, b). Patients' characteristics other than age were not known.

Processed images were independently assessed for BMD specifically for this project according to BI-RADS V recommendations [6] by three (GE database) or seven (Hologic database) Mammography Quality Standards Act (MQSA)-qualified radiologists from Brigham and Women's Hospital (Boston, USA): one with 3 years' experience and two with more than 10 years' experience assessing mammograms for GE database, and two with 3 years' experience and five with more than 10 years' experience assessing mammograms for Hologic database. The difference in the number of readers for each database was directly related to the availability of readers at the time each reference database was built. The median score was used for consensus.

Fig. 1 Age distribution of reference images (% per age category). **a** GE reference database ($n = 1200$). **b** Hologic reference database ($n = 1089$)



Validation set

An additional 800 pairs (CC + MLO) of FFDM anonymized and processed images were collected from patients undergoing a mammography exam in 2016, performed either in a radiology practice in Paris, France, with Senographe Essential (400 consecutive FFDM image pairs) or in a radiology department in Monaco with Selenia Dimensions (400 consecutive FFDM image pairs). The images were independently assessed specifically for this project by seven blinded radiologists, different from those who assessed the images in the reference database: four MQSA-qualified radiologists from Brigham and Women's Hospital with more than 10 years' experience, and three radiologists specialized in breast imaging from Gustave Roussy cancer center (Villejuif, France)—two with 3 years' experience

and one with 6 years' experience. The median score was used for consensus.

To evaluate software reproducibility, image pairs (CC + MLO) from 214 patients in the validation set were used to compare the software's BMD assessments between left and right breasts.

Description of the DenSeeMammo algorithm

DenSeeMammo software (version 1.20, Predlife) was used for automated breast density assessment. Image comparison can only be performed for images with the same dimension and resolution. To enable overlapping of mammograms, images were pre-processed and standardized in a completely automatic manner to have the same orientation, dimension, and resolution, as follows: breast margins were determined

(for MLO views, pectoral muscle was removed (see details in [Supplementary material](#))), images were cropped as close as possible around breast margins and vertically divided into two parts from the tip of the breast, each part was resized in a square of 200×200 pixels for GE and 100×100 for Hologic, and the two squares were placed side by side to obtain an image of 200×400 pixels for GE and 100×200 for Hologic (Fig. 2a, b). When a new mammogram was performed, typically four images (two views, two breasts) were analyzed by DenSeeMammo. However, images could also be analyzed when only a CC view and/or views from one breast were available.

Image comparison is based on Manhattan distance (Md) calculation, Md being the simple sum of components. To select the most similar images, first, the most different images were eliminated from the reference database; then, pixel-to-pixel image comparison was performed between the remaining images. For the first step, values were calculated for brightness (average of pixel values) and contrast (standard deviation of pixel values), and the value of the image to be analyzed was compared to that of the reference images. Using Md, the size of the reference database was reduced by 5–15%

by eliminating the most different images with respect to these image features. This step was performed with respect to brightness and contrast for GE images, and with respect to brightness (80th percentile, 40th percentile, then 90th percentile) for Hologic images. Next, pixel-to-pixel image comparison was performed between the remaining images. As shown in Fig. 3, in the DenSeeMammo algorithm, the concept of Md is applied to compare a patient’s mammogram to reference mammograms, based on the difference between image densities of the two mammograms. First, the differences between the gray levels of each pixel having the same coordinates were calculated for each view (CC or MLO). For each breast, Md was then determined by summing the previously calculated distances for each view, between the image to be analyzed and each image in the reference database. The Md obtained therefore corresponds to the sum of absolute values of differences between gray levels of a pair of images (CC + MLO views). Finally, the algorithm retrieved the reference images with the lowest Md (i.e., the most similar images).

The number of nearest neighboring images used by DenSeeMammo was determined empirically by evaluating the agreement between DenSeeMammo and consensus of

Fig. 2 Image adjustment and standardization process using DenSeeMammo software. **a** Overall standardization process for GE and Hologic images. **b** Detailed standardization process for each image

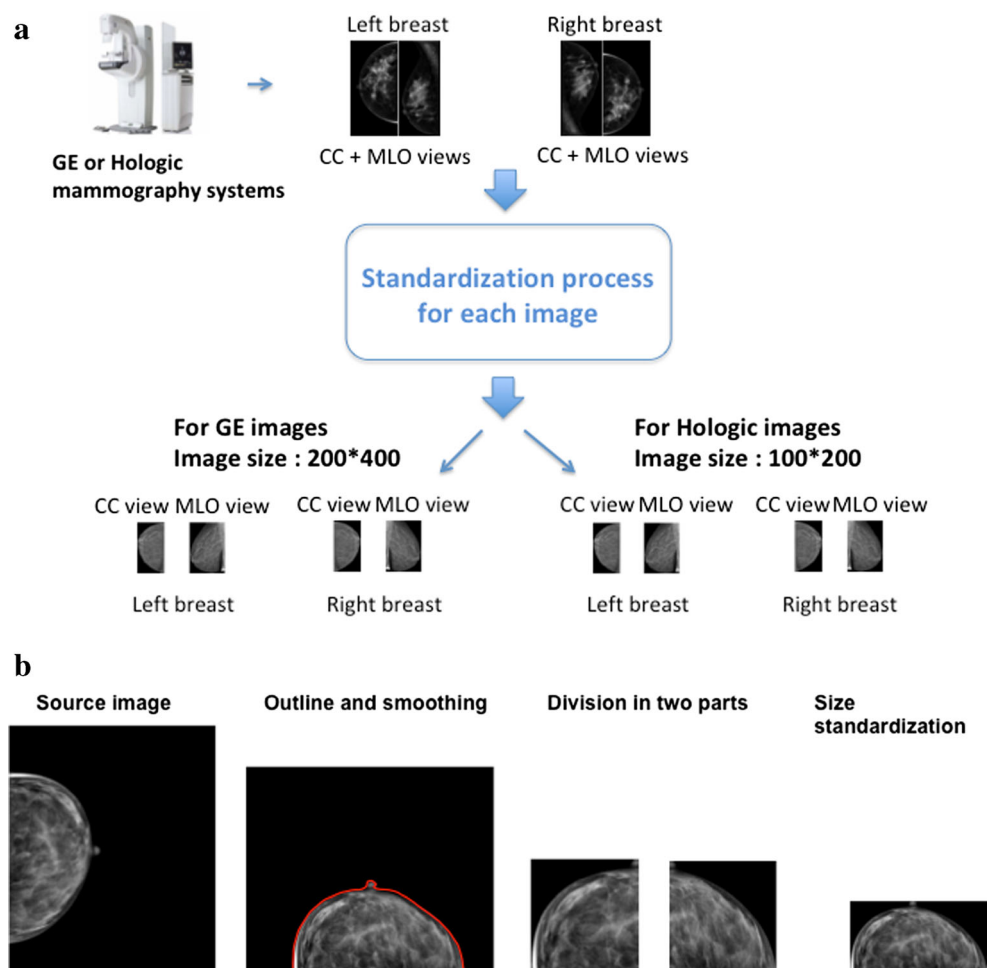
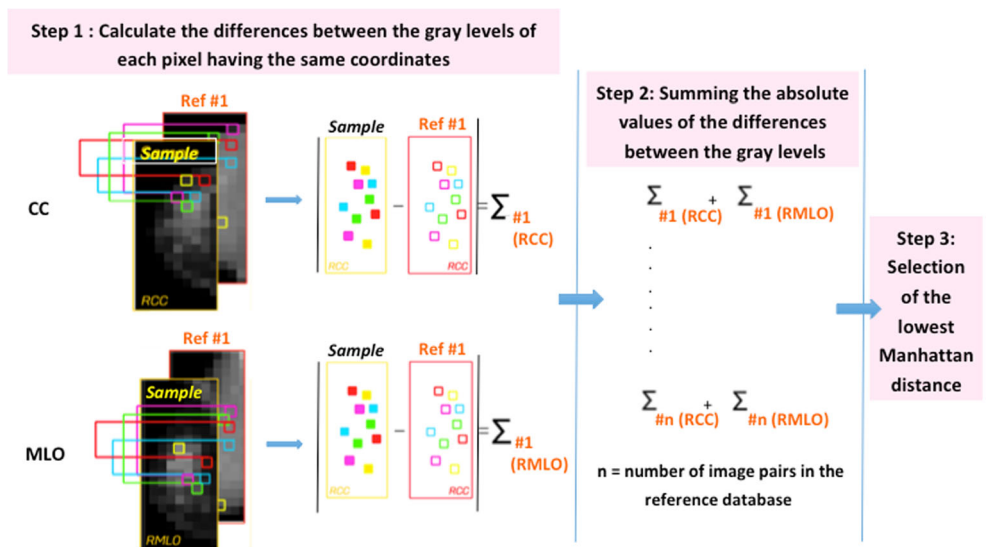


Fig. 3 Image comparison based on Manhattan distance calculation for one breast (R = right): two views (cranio-caudal (CC) and medio-lateral oblique (MLO))



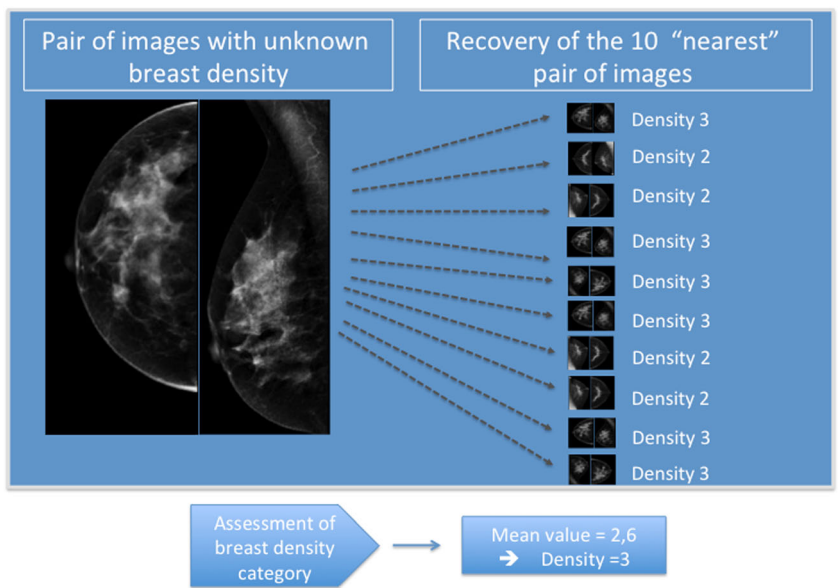
radiologists in the reference database according to the number of neighboring images, and the best kappa value was obtained with the ten nearest neighbors for GE images and 14 for Hologic images (data not shown). The average of the BMD values was therefore calculated from the selected nearest neighboring images, then a breast density category was assigned (Fig. 4) according to the following thresholds: for GE, < 1.5, category A; ≥ 1.5 and < 2.5, category B; ≥ 2.5 and < 3.5, category C; and ≥ 3.5 , category D; for Hologic, < 1.5, category A; ≥ 1.5 and < 2.6, category B; ≥ 2.6 and < 3.3, category C; and ≥ 3.3 , category D. These thresholds gave the best agreement between DenSeeMammo and the consensus of radiologists (unweighted kappa) in the reference database (data not shown). The calculation of the average BMD values was weighted according to distance (see details in the [Supplementary material](#)).

In order to ensure the performance of the algorithm, a security check blocked density assessment in case an image was distorted after pre-processing. However, this error did not occur when testing with 5600 images on a database (data not shown). Furthermore, if breast overlapping was not complete, the distances between images became too large for the non-overlapping image to be considered as one of the nearest images.

Statistical analysis

Unweighted Cohen’s kappa correlation for two raters [20] and weighted Cohen’s kappa correlation for two raters [21] were used to measure agreement between DenSeeMammo and visual assessment. Overall inter-observer kappa was calculated as the mean of Cohen’s kappa obtained side by side for seven

Fig. 4 Software overall process for BMD assessment. After the ten nearest neighboring images are selected from the reference database (for GE images), the average of the BMD values is calculated, and a breast density category is assigned



raters. Spearman rank correlation coefficients were used for reproducibility analysis. All analyses were performed using Statistical Analysis System (SAS, version 9.3, SAS Institute). Kappa values were interpreted as suggested by Landis and Koch [22], as follows: a kappa value equal to or less than 0.20 indicates slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement.

Results

Consensus of radiologists’ visual assessment for the reference database

Consensus of radiologists was used for the reference database (consensus of three or seven radiologists, respectively, for GE and Hologic images). The overall inter-observer agreement (unweighted kappa) when considering the four breast density categories for the GE and Hologic reference databases was, respectively, 0.49 and 0.52. When considering clinically significant non-dense (A-B) and dense (C-D) categories, the inter-observer agreement was almost perfect (unweighted $\kappa = 0.86$) for the GE database and substantial (unweighted $\kappa = 0.75$) for the Hologic database. The distribution across density categories in each reference database according to the radiologists’ consensus assessments was in line with the historical distribution of density categories in the screening population [6], with 9% of images in category 1 (almost entirely fat), 44% in category B (scattered fibroglandular density), 38% in category C (heterogeneously dense), and 9% in category D (extremely dense).

Breast density classification by DenSeeMammo software and visual assessment

For 800 mammographic image pairs (CC + MLO) in the validation set (GE and Hologic), the results of breast density

categories assigned by each of the seven observers and DenSeeMammo are presented in Table 1. The overall unweighted kappa for the seven radiologists’ assessment of BI-RADS density categories showed moderate agreement (unweighted $\kappa = 0.55$; weighted $\kappa = 0.65$). Nevertheless, when BI-RADS categories were grouped into clinically significant non-dense (A-B) and dense (C-D) categories, inter-observer agreement was substantial (unweighted $\kappa = 0.72$).

Table 2 shows the BMD assessment by DenSeeMammo with two views (CC + MLO) and consensus of radiologists in the validation set. When BI-RADS categories were grouped into non-dense (A-B) and dense (C-D) categories, 97% of the non-dense category and 87% of the dense category assigned by the software were assigned in the respective categories by the consensus of radiologists (Table 2). When considering the four breast density categories, the software showed a substantial agreement with the radiologists’ consensus (unweighted $\kappa = 0.68$, 95% CI 0.64–0.72, and weighted $\kappa = 0.74$, 95% CI 0.71–0.77). Furthermore, when considering non-dense (A-B) and dense (C-D) categories, the software showed an almost perfect agreement with the consensus of radiologists’ visual assessment (unweighted $\kappa = 0.84$, 95% CI 0.80–0.88).

Table 3 shows the results obtained when only one view (CC) was used for BMD assessment, from GE and Hologic images. The results were comparable to those obtained with two views, with a substantial agreement between the software and the radiologists’ consensus (unweighted $\kappa = 0.68$, 95% CI 0.64–0.72, and weighted $\kappa = 0.75$, 95% CI 0.72–0.78).

Software compatibility and reproducibility

Table 4 shows that agreement (measured by unweighted kappa) between the software and the consensus of radiologists is equivalent, whatever the mammography equipment, either with GE Senographe Essential or Hologic Selenia Dimensions system. When assessing the breast density category assigned by DenSeeMammo to the left and right breasts of the same patient, a strong positive correlation was found

Table 1 Frequency of breast density categories assigned by each reader and DenSeeMammo in the validation set

Density	No. (%) of mammographic examinations (<i>n</i> = 800)								
	Observers							DSM	
	Observer α	Observer β	Observer γ	Observer δ	Observer X	Observer Y	Observer Z	Consensus	
A	53 (6.63)	133 (16.63)	105 (13.13)	118 (14.75)	37 (4.63)	86 (10.75)	38 (4.75)	79 (9.88)	74 (9.25)
B	375 (46.88)	298 (37.25)	461 (57.63)	311 (38.88)	318 (39.75)	268 (33.50)	450 (56.25)	349 (43.63)	311 (38.88)
C	251 (31.38)	271 (33.88)	224 (28.00)	298 (37.25)	388 (48.50)	352 (44.00)	270 (33.75)	308 (38.50)	357 (44.63)
D	121 (15.13)	98 (12.25)	10 (1.25)	73 (9.13)	57 (7.13)	94 (11.75)	42 (5.25)	64 (8.00)	58 (7.25)

Mammographic exams performed with either Senographe Essential (GE) or Selenia Dimensions (Hologic)

DSM DenSeeMammo

Table 2 Agreement matrix between DenSeeMammo and consensus of radiologists' visual assessment in the validation set (no. (%) of mammographic examinations ($n = 800$))

DSM	Radiologists' consensus				
	A	B	C	D	Total
A	52 (6.50)	22 (2.75)	0	0	74 (9.25)
B	27 (3.38)	273 (34.13)	11 (1.38)	0	311 (38.88)
C	0	54 (6.75)	274 (34.25)	29 (3.63)	357 (44.63)
D	0	0	23 (2.88)	35 (4.38)	58 (7.25)
Total ($n = 602$)	79 (9.88)	349 (43.63)	308 (38.50)	64 (8.00)	800 (100.0)

Note: Mammographic exams performed with either GE Senographe Essential or Selenia Dimensions (Hologic)
DSM DenSeeMammo

($\kappa = 0.87$, 95% CI % 0.84–0.90), which was equivalent to the correlation obtained with the radiologists' consensus ($\kappa = 0.88$, 95% CI % 0.85–0.91).

Discussion

In this study, we have developed a new automated tool based on digital image comparison that mimics radiologists' consensus assessments of BMD according to BI-RADS V recommendations. Our results showed substantial agreement with visual assessment by a consensus of seven radiologists (weighted $\kappa = 0.74$, 95% CI 0.71–0.77) in the validation set of 800 pairs of mammographic examinations. Although the difference in the classification standards (BI-RADS edition) and the use of different sets of images preclude any direct comparison between the methods, there was a higher degree of agreement than that reported in recently published studies on automated volumetric assessment methods [13, 23, 24]. Brandt et al reported a moderate agreement between BI-RADS classification and Volpara (weighted $\kappa = 0.57$; 95% CI 0.55–0.59) and Quantra (weighted $\kappa = 0.46$; 95% CI 0.44–0.47), respectively [13]. In Youk et al, the agreement of density category with visual assessment ranged from moderate to substantial in Quantra (weighted $\kappa = 0.54$ –0.61) and fair to moderate in Volpara (weighted $\kappa = 0.32$ –0.43) [23]. In Sartor et al, the agreement between BI-RADS scores and

volumetric breast density using Volpara was found to be moderate (weighted $\kappa = 0.55$; 95% CI 0.53–0.56) [24].

The present method has a different approach from other automated area-based or volumetric methods in that it is a quantitative tool based on qualitative measures. By taking into account the dispersion of fibroglandular tissue throughout the breast and not the percentage of dense tissue within the breast, this method is more in line with the BI-RADS V breast density classification scheme than volumetric assessments of dense tissue. Recent studies have shown that texture features may be predictors of breast cancer risk independent of percent density, and automated tools for analyzing parenchymal texture features are in development [25–27]. Future studies are needed to compare the performance of DenSeeMammo with that of currently available BMD assessment tools on the same set of mammographic exams.

Accurate measurement of breast density, incorporated in a risk assessment model, could contribute to stratified screening based on individual risk. Indeed, while mammographic screening has been shown to have contributed to reduce breast cancer specific mortality by 20% in the invited population [28], its potential harms, including false-positive recalls, unnecessary biopsies, patient anxiety, overdiagnosis, and over-treatment, plead for an improved balance of benefits and harms at the individual level and implementation of a stratified screening strategy based on individual risk [29]. The present method may be a good candidate to be used for risk estimation

Table 3 Agreement matrix between DenSeeMammo and consensus of radiologists' visual assessment in the validation set using only one view (CC) (no. (%) of mammographic examinations ($n = 800$))

DSM	Radiologists' consensus				
	A	B	C	D	Total
A	49 (6.13)	22 (2.75)	0	0	71 (8.88)
B	24 (3.00)	283 (35.38)	14 (1.75)	0	321 (40.13)
C	0	51 (6.38)	274 (34.25)	32 (4.00)	357 (44.63)
D	0	0	19 (2.38)	32 (4.00)	51 (6.38)
Total ($n = 602$)	73 (9.13)	356 (44.50)	307 (38.38)	64 (8.00)	800 (100.0)

Note: Mammographic exams performed with either Senographe Essential (GE) or Selenia Dimensions (Hologic)
DSM DenSeeMammo

Table 4 Agreement (kappa value) between DenSeeMammo and consensus of radiologists' visual assessment in the validation set according to mammography equipment (GE or Hologic)

Breast density category	Unweighted κ DSM/consensus (GE + Hologic images) (CI 95%)	Unweighted κ DSM/consensus (GE images) (CI 95%)	Unweighted κ DSM/consensus (Hologic images) (CI 95%)
A-B-C-D (4 categories)	0.68 (0.64–0.72)	0.66 (0.60–0.72)	0.78 (0.73–0.83)

DSM DenSeeMammo

in screening programs. In a recent study, Kerlikowske et al reported that the discriminatory accuracy of the Breast Cancer Surveillance Consortium risk model was improved by combining quantitative and qualitative measures of breast density, and urged researchers to assess the combination of automated measures of dense volume and texture features, with clinical assessment [17]. In future work, the present tool could be tested in a risk assessment model as a qualitative measure, in combination with quantitative measures. The tool, which has received FDA clearance in 2016, has been used as part of a risk program management, in a pilot study assessing its integration in a radiology workflow [30].

Due to the limitations of breast density measurement methods that are based on two-dimensional mammographic projection images [31], other technologies are also under investigation, using magnetic resonance imaging (MRI), digital breast tomosynthesis (DBT), computed tomography, optical imaging, and ultrasound [12]. However, given the widespread use of mammography for screening and the relatively high cost of certain technologies such as MRI, it is relevant to develop a robust and accurate breast density measurement method based on mammographic images. Nevertheless, DBT is increasingly used in screening facilities and will likely replace 2D screening mammography [32]. With that perspective, work is currently ongoing to assess breast density using this method in synthesized 2D mammograms from DBT data.

This study has several limitations. The study was designed to compare the software tool's assessment results with the consensus of radiologists' visual assessments. The large inter-observer variability observed when considering the four categories (unweighted $\kappa = 0.49$ for GE and 0.52 for Hologic), comparable to published results [7, 33], can call into question the use of consensus of radiologists' assessments as "gold standard." Nevertheless, when BI-RADS categories were grouped into clinically significant non-dense (A-B) and dense (C-D) categories, the inter-observer agreement was almost perfect (unweighted $\kappa = 0.86$) for GE database and substantial (unweighted $\kappa = 0.75$) for Hologic database. For the validation set, the inter-observer agreement was also substantial (unweighted $\kappa = 0.72$) when considering the two categories. In addition, the off-diagonal classification between categories B and C was limited, with only 8.13% of cases. Furthermore, we

previously found that a consensus of at least three radiologists specialized in breast imaging (with experience of 3 years or more) allows overcoming the intra- and inter-observer variability issue of breast density assessment, and can be considered to represent the radiologist population's assessment, especially considering the clinically meaningful categories of dense/non-dense (data not shown). Although the present tool cannot eliminate the variability that is inherent in visual assessment, and we may find different results with different sets of readers under different study conditions, the reference database in this study, with its variability, represents the radiologists' assessments that have been shown to have a strong relationship with breast cancer risk [17, 18]. We therefore deemed appropriate to use consensus assessments in this study as the "gold standard" for the reference database. Another limitation is the tool's generalizability. The reference databases and the validation set included mammograms generated by GE and Hologic systems. The present results do not allow us to affirm the generalizability of this method when using images generated by other mammography systems. Further research is needed to test the validity of this approach using images generated by other mammography systems.

In summary, BMD assessment by the software was shown to be strongly correlated to radiologists' consensus assessments of BMD. Future research is needed to compare this method with other existing methods and evaluate its generalizability as well as its clinical utility.

Acknowledgments The authors would like to thank Sylvie Phung (Predlife, France) for her technical support and Sandra Canale (Gustave Roussy, France) for her clinical support.

Funding This study has received funding by Fondation ARC pour la Recherche.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Dr. Corinne Balleyguier.

Conflict of interest Emilien Gauthier, Valerie Helin, and Stephane Ragusa, authors of this manuscript, declare relationships with Predlife (Villejuif, France). Predlife, which developed the DenSeeMammo algorithm, did not support the study, but provided their software tool for the

study. Non-employee authors had complete control of the data and information that might present a conflict of interest to the authors who are employees of Predlife.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was not required for this study because it is a retrospective analysis of imaging datasets.

Ethical approval Institutional review board approval was obtained.

Methodology

- retrospective
- observational
- multicenter study

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Harvey JA, Bovbjerg VE (2004) Quantitative assessment of mammographic breast density: relationship with breast cancer risk. *Radiology* 230:29–41
2. Vacek PM, Geller BM (2004) A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiol Biomarkers Prev* 13:715–722
3. Boyd NF, Rommens JM, Vogt K et al (2005) Mammographic breast density as an intermediate phenotype for breast cancer. *Lancet Oncol* 6:798–808
4. McCormack VA, dos Santos Silva I (2006) Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 15:1159–1169
5. Boyd NF, Guo H, Martin LJ et al (2007) Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 356:227–236
6. Sickles EA, D'Orsi CJ, Bassett LW et al (2013) ACR BI-RADS® atlas, breast imaging reporting and data system. American College of Radiology, Reston
7. Ciatto S, Houssami N, Apruzzese A et al (2005) Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. *Breast* 14:269–275
8. Lobbes MB, Cleutjens JP, Lima Passos V et al (2012) Density is in the eye of the beholder: visual versus semi-automated assessment of breast density on standard mammograms. *Insights Imaging* 3:91–99
9. Sprague BL, Conant EF, Onega T et al (2016) Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study. *Ann Intern Med* 165:457–464
10. Ng KH, Yip CH, Taib NA (2012) Standardisation of clinical breast-density measurement. *Lancet Oncol* 13:334–336
11. Yaffe MJ (2008) Mammographic density. Measurement of mammographic density. *Breast Cancer Res* 10:209
12. Chen JH, Gulsen G, Su MY (2015) Imaging breast density: established and emerging modalities. *Transl Oncol* 8:435–445
13. Brandt KR, Scott CG, Ma L et al (2016) Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. *Radiology* 279:710–719
14. Jeffers AM, Sieh W, Lipson JA et al (2017) Breast cancer risk and mammographic density assessed with semiautomated and fully automated methods and BI-RADS. *Radiology* 282:348–355
15. Aitken Z, McCormack VA, Highnam RP et al (2010) Screen-film mammographic density and breast cancer risk: a comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. *Cancer Epidemiol Biomarkers Prev* 19:418–428
16. Wanders JOP, Holland K, Karssemeijer N et al (2017) The effect of volumetric breast density on the risk of screen-detected and interval breast cancers: a cohort study. *Breast Cancer Res* 19:67
17. Kerlikowske K, Ma L, Scott CG et al (2017) Combining quantitative and qualitative breast density measures to assess breast cancer risk. *Breast Cancer Res* 19:97
18. Astley SM, Harkness EF, Sergeant JC et al (2018) A comparison of five methods of measuring mammographic density: a case-control study. *Breast Cancer Res* 20:10
19. Destounis S, Arieno A, Morgan R, Roberts C, Chan A (2017) Qualitative versus quantitative mammographic breast density assessment: applications for the US and abroad. *Diagnostics (Basel)* 7:30
20. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46
21. Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 70:213–220
22. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
23. Youk JH, Gweon HM, Son EJ, Kim JA (2016) Automated volumetric breast density measurements in the era of the BI-RADS fifth edition: a comparison with visual assessment. *AJR Am J Roentgenol* 206:1056–1062
24. Sartor H, Lång K, Rosso A, Borgquist S, Zackrisson S, Timberg P (2016) Measuring mammographic density: comparing a fully automated volumetric assessment versus European radiologists' qualitative classification. *Eur Radiol* 26:4354–4360
25. Gastouniotti A, Conant EF, Kontos D (2016) Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment. *Breast Cancer Res* 18:91
26. Malkov S, Shepherd JA, Scott CG et al (2016) Mammographic texture and risk of breast cancer by tumor type and estrogen receptor status. *Breast Cancer Res* 18:122
27. Wang C, Brentnall AR, Cuzick J, Harkness EF, Evans DG, Astley S (2017) A novel and fully automated mammographic texture analysis for risk prediction: results from two case-control studies. *Breast Cancer Res* 19:114
28. Independent UK Panel on Breast Cancer Screening (2012) The benefits and harms of breast cancer screening: an independent review. *Lancet* 380:1778–1786
29. Kerlikowske K (2009) Evidence-based breast cancer prevention: the importance of individual risk. *Ann Intern Med* 151:750–752
30. Weigert J, Cavanaugh N, Ju T (2018) Evaluating mammographer acceptance of MammoRisk software. *Radiol Technol* 89:344–350
31. Kopans DB (2008) Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. *Radiology* 246:348–353
32. Hooley RJ, Durand MA, Philpotts LE (2017) Advances in digital breast tomosynthesis. *AJR Am J Roentgenol* 208:256–266
33. Gweon HM, Youk JH, Kim JA, Son EJ (2013) Radiologist assessment of breast density by BI-RADS categories versus fully automated volumetric assessment. *AJR Am J Roentgenol* 201:692–697