UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

BREAST TISSUE DENSITY ESTIMATION USING AUTOMATED VOLUMETRIC BREAST DENSITY COMPARED TO RADIOLOGISTS' VISUAL ASSESSMENT BY BREAST IMAGING REPORTING AND DATA SYSTEM (BIRADS) AT CHARLOTTE MAXEKE JOHANNESBURG ACADEMIC HOSPITAL (CMJAH), SOUTH AFRICA

Dr Thalitha George

A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Medicine in Diagnostic Radiology.

Johannesburg, 2020

DECLARATION

I, Thalitha George, declare that this thesis is my own work. It is being submitted for the degree of Master of Medicine in Diagnostic Radiology at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

____

Dr Thalitha George

27________day of February________2020__________in Johannesburg______________

DEDICATION

This study is dedicated to my family. Thank you for the support

PUBLICATIONS AND PRESENTATIONS

This work has never been published.

It has never been presented at a congress.

ABSTRACT

There are different methods of assessing breast tissue density, which include automated volumetric assessment and visual assessment using the BI-RADS classification.

The aim of the study was to assess concordance in evaluation of breast density between fully automated volumetric and visual assessment by radiologists. Secondarily, intra-observer concordance was assessed for each reader, re-reading 100 of the images after 8-weeks.

A retrospective study using images from 2017 and 2018 was conducted at Charlotte Maxeke Johannesburg Academic Hospital. The sample included 192 women.

The mammograms were performed on the Full Field Digital Mammogram Hologic Selenia Dimensions. The mammograms were reviewed by four radiologists using BARCO 5 and BARCO 3 megapixel monitors. The volumetric density was obtained using Hologic Quantra 3 software. Agreement assessed with kappa statistics. Experienced readers agreed with the software in 46.9% and 45.8% of the cases (kappa 0,174 and 0.132). The less experienced readers showed 40% and 36.4% (kappa 0.075 and 0.074). Intra-reader concordance was 35% and 60%, (kappa 0.016 and 0.365).

The second inter reader agreement observed between the radiologists' second reading and the machine was assessed. Readers 1 and 2, 2nd reading showed an agreement of 47% and 49% (kappa 0.1519 and 0.1638). Readers 3 and 4, 2nd reading was 41.00% and 37% (kappa 0.1146 and 0.0601).

The study showed that visual assessment of breast density on mammograms is observer depended. Experienced radiologists demonstrated slightly higher agreement with automated assessment than junior radiologists. There was also intra-reader variation when readers were asked to assess the same cases 8-weeks later.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Prof Smilg and Dr Haberfeld for their time, vast knowledge of the topic and continuous support with this project.

I would like to thank my statistician, Mr Lukhanyo Nyati. His assistance has been remarkable.

I would also like to thank my data readers, Professor J Smilg, Dr J Haberfeld, Dr T Buthelezi and Dr K Mahango.

Table of Contents

LIST OF TABLES

NOMENCLATURE

A	Almost fatty breast
B	Scattered fibroglandular breast density
BI-RADS	Breast Imaging-Reporting and Data System
C	Heterogenously dense breast tissue
D	Extremely dense
DY	Dysplasia
N1	Normal
P1	Prominent 1
P2	Prominent
QDY	Quasi-dysplasia

1 INTRODUCTION

1.1 Timeline in density assessment

The first person to describe mammographic breast density as a risk factor for breast cancer was John Wolfe in 1976. He hypothesised that if a relationship between a prominent duct pattern and breast cancer was regularly seen in women with breast cancer, then a prominent duct pattern may precede the development of breast cancer. This described a relationship between a prominent duct pattern and breast cancer[1]. He described four qualitative categories based on parenchymal patterns:

- N1 (normal): described a breast made entirely of fat;
- P1 (prominent 1): breast composed predominantly of fat, but displaying prominent ducts behind the areola or in the upper axillary quadrant occupying not more than 25% of the breast;
- P2 (prominent 2): prominent duct pattern, more than P1 (often in a triangular pattern in the centre of the breast), with a quarter or more of the breast occupied;
- DY (dysplasia): described a general increase in breast density, and a possible minimal involvement of prominent ducts.

Categories that were determined as lower cancer risk included N1 and P1, whereas P2 and DY presented higher risk patterns. Women below the age of 40-45 years were classified under the QDY (quasi-dysplasia) category due to the fact that these women usually show the DY pattern because of their young age; and after menopause, this pattern tends to regress to a lower risk pattern[1].

Visually based assessment has lacked reproducibility. In 1980, Norman Boyd et al. were the first to try to semi-quantify density utilising a six-category classification (SCC). The aim was to decrease the heterogeneity in the risk estimates. This method was the first to be more objective in assessing tissue density patterns by using percentages. The six-category classification is a quantitative area-based measure that has visual estimates of density utilising a thresh-holding method, these being; A: 0% density; B: 0 to <10% density; C: 10% to <25% density; D: 25% to <50% density; E: 50% to <75% density; and F: >75% density. Higher SCC scores were strongly associated with increased breast cancer risk, although this was to a lesser degree compared to the initial Wolfe estimates [1].

In addition, another semi-quantitative approach is the Visual Analogue Scale (VAS), of which readers mark along a continuous scale that represents 0-100% density and these score sheets can then be scanned through software to obtain the percent breast tissue [1].

In 1997, Lazlo Tabar developed another pattern-based qualitative description system of breast density. His model described a mixture of five mammographic blocks that are components of normal breast anatomy. These patterns included nodular densities corresponding to the terminal duct lobular units; linear structures that correspond to either ducts or fibrous or blood vessels; homogenous structure-less densities, which related to fibrous density; and radiolucent areas, which correlated with adipose fatty tissue. Pattern IV and V (correspond to P2 and DY) describe dense breasts carrying a 20% higher incidence of breast cancer [1].

- Pattern 1 was characterised by prominently dense tissue with nodular densities, with regions of fatty tissue;
- Pattern II indicated completely fatty breast;
- Pattern III described a mostly fatty breast with visible ducts behind the areola;
- Pattern IV were predominantly dense breasts with linear and nodular densities;
- Pattern V comprised an increased level of homogenous density[1].

Another visual method of assessing breast density is the Breast Imaging-Reporting and Data System (BI-RADS) that is used for reporting mammographic breast density[2]. The BI-RADS system aims at providing a standardised lexicon of reporting and imaging interpretations.

The first edition of the BI-RADS was published in 1993, and revisions were made in 1995 and 1998, according to Rao(3)] The previous and commonly used (for over 10 years) BI-RADS 4 breast density categories issued in 2003 by the American College of Radiology were descriptive, quantitative definitions of the BI-RADS mammographic density categories which were defined as category 1 (<25% glandular), category 2 (25%-50% glandular), category 3 (51-75% glandular) and category 4 (> 75% glandular).

The 5th BI-RADS edition, published in 2014, predominantly assesses breast density based on the distribution of tissue density and the higher chance of obscuration of underlying lesion it presents, as opposed to percentage quartile ranges for each mammographic breast density category. For example, if the breasts are predominantly fatty, but there is a focal area of dense fibroglandular tissue accumulation, then the breasts are described as heterogeneously dense to indicate that this area may obscure a mass, even though the percentage of the total breasts consists of fibroglandular tissue at less than 50 per cent. BI-RADS(5th edition) classifies breast density as; category 1: almost entirely fatty; category 2: scattered fibroglandular densities; category 3: heterogeneously dense; and category 4: extremely dense [2]. This is illustrated in table 2.

Due to limitations caused by visual density assessment, less subjective and more quantitative measures have been developed since then. Many of these measures include semi-automated density assessments methods such as Cumulus and Madena. These methods are another threshold-based method for density measurement. Training and human input are needed to define the threshold of density with these methods, in an effort to avoid subjectivity and address the challenges in their application in clinical use. Therefore, these methods are for research purposes only[1]. There are other approaches such as volumetric methods of assessing breast density that have also been developed, which estimate the actual volume of fibroglandular tissue in the breast, taking advantage of the information contained in the raw mammographic image. Volumetric methods are anatomically relevant and take the depth of fibroglandular tissue into consideration from tomosynthesis volume, compared to area-based methods. The commercially most frequently used software includes the Quantra volumetric breast density assessment that became available in 2008. This tool was produced by Hologic (Hologic Inc., Bedford, MA, USA). Another volumetric assessment tool is Volpara Density (Volpara), which was developed in 2012 by Volpara Health Technologies, Wellington, New Zealand[1].

Breast tissue density is an important indicator in risk stratification of breast cancer. It has been proven that the risk of developing breast cancer in those with heterogeneous or extremely dense breast tissue density is 2 to 6-fold as high compared to those with fatty or scattered fibroglandular tissue [4].

The aim of this study was to assess concordance in evaluation of breast tissue density between fully automated computed volumetric assessment and visual assessment by radiologists using the BI-RADS categorisation. In addition, intra-reader concordance was examined after an eight-week interval from the initial read. The results are demonstrated in table 3.1 and 3.2.

1.2 Epidemiology of breast cancer in South Africa and worldwide

Breast cancer is considered as both the most common cancer in females and the leading cause of cancer deaths among women worldwide[5]. Based on the 2015 Global Burden of Disease (GBS) study, of the 17.5 million cancer cases detected globally, breast cancer accounted for 2.4

million new cases and 523 000 deaths in 2015. Similarly, breast cancer is the most common form of cancer to affect women in South Africa, and in 2013 was accountable for 20.8 per cent of female cancers and more than 10 per cent of the entire cancer burden [5]. It is predicted that the number of women dying from breast cancer annually will increase if there are no significant advances in screening and treatment efforts in the near future. By 2030, it is estimated that the number of women diagnosed globally with breast cancer will increase to almost 3.2 million per year, nearly double the incidence in 2012 [5].

The incidence of breast cancer appears to be low among black South Africans when compared to other races, and this has been attributed to certain epidemiological factors peculiar to breast cancer such as early age of first birth, multi-parity, universal and prolonged lactation, low use of hormone replacement therapy and a diet low in fat / high in fibre [5].

Sartor indicated that increased mammographic density is associated with increased risk of breast cancer. There has been a developing interest in the assessment of mammographic density in screening programs and for incorporation into risk prediction models. Satisfactory assessment of mammographic density requires a trustworthy measuring tool. The BI-RADS classification for assessing breast density is commonly used. However, this method has inter-reader variability and is subjective[6].

As of May 2017, it has been reported that about 31 states in the United States of America have density notification laws[1]. The laws consider women undergoing screening mammography who have dense breast tissue, meaning the BI-RADS-heterogeneously or very dense categories, have to be given written notification of their breast density. These women are then educated about the implications of dense breast tissue and the associations with increased risk of breast cancer as this may affect the accuracy of mammography[7]. This suggests that supplemental tests such as an MRI and ultrasound may be necessary for further evaluation. There are advantages and disadvantages to this notification process. Overestimation of breast tissue density may cause unnecessary anxiety in the patient and result in additional and costly investigations. There are currently no laws in place in South Africa regarding mandatory notification of patients as to their radiographic breast density.

Furthermore, not only does dense breast tissue increase the risk of breast cancer but it may also mask detection of the disease [8]. It is common knowledge that most breast cancers have similar radiographic density (isodense) to breast tissue.[9]. To detect a breast cancer on a mammogram, a portion of the lesion must be abutted by fatty tissue. The denser the glandular tissue the smaller the chance that there would be fat abutting the cancer. Denser breast tissue would also have more overlapping tissue that could make it more difficult to appreciate the lesion. Therefore, it is pivotal to make sure that breast cancer is not missed and is diagnosed early to prevent unnecessary deaths.

1.3 Study objectives

This study aimed to:

a) Explore agreement of breast density assessment between automated volumetric assessment and visual assessment by radiologists at the CMJAH.
b) Compare intra-reader and inter-reader agreement among radiologists with varying degrees of experience in breast imaging.

2 MATERIALS AND METHODS

A cross sectional, retrospective study was conducted at the Charlotte Maxeke Johannesburg Academic Hospital during the years 2017-2018. The study population were women who presented for mammograms.

2.1 Sample size

A representative sample size was calculated using the formula:

$$\mathrm{n} = \frac{\mathrm{n}^*}{1+\mathrm{n}^*/\mathrm{N}}, \text{ where } n^* = \frac{1}{r^2(p_a - p_e)^2} \qquad [10]$$

This formula is based on calculating the sample size on all kappa-like agreement coefficients, where *n* depends on the relative error *r* and the difference p_a-p_e between the overall agreement probability *pa* and the chance- agreement probability *pe*. N is the number of subjects in the entire population. Data from literature [6] were used for working out the sample size. Given the observed agreement between automated volumetric density and the BI-RADS scores (kappa = 0.55) as well as the relative error, which is an arbitrary 20 per cent, a minimum sample size was obtained comprising 156 reading [6]. This number was rounded up to be as close to 200 as possible.

2.2 Inclusion and exclusion criteria

The final sample included 192 female patients who qualify for breast cancer mammography. Women who were proven to have breast cancer, had mass lesions, had undergone surgery as well as women with prior chemotherapy and radiation therapy were excluded from the study.

2.3 Technical, radiological and clinical aspects

All reviewed digital mammograms were performed on the Full Field Digital Hologic Selenia Dimensions Standard 2D C-view medio-lateral (MLO) and cranio-caudal (CC) views were interpreted.

Mammograms were reviewed on BARCO 5-megapixel monitors. The two junior consultants reviewed the cases on BARCO 3-megapixel monitors. Breast tissue density was visually classified according to the BI-RADS 5th edition. Computer software, Hologic Quantra 3, version 2.1.1 was used to generate automated volumetric breast density.

Four radiologists took part in the study. Two were senior consultants, each with 17 or more years of experience in mammography, and two junior consultants, each with less than five years of experience. The readers were blinded to automated results.

The results of the radiologist BI-RADS classification was compared by the lead researcher to the automated volumetric breast density calculated by the computer.

The radiologists were given 100 of the original 200 cases to review eight weeks later to assess if there was any change in classification. Standardised score sheets were provided to the radiologists.

2.4 Data collection

Data were collected anonymously by allocating random identification (viz., Case 1, Case 2, etc.). The attributed cases were only known by the primary investigator. According to the latest BI-RADS-5 edition, the following categories for mammographic density were used: almost entirely fatty breast tissue; scattered areas of fibroglandular density; heterogeneously dense; and extremely dense. Radiologists used these classifications to categorise the type of breast tissue. The results were compared to automated volumetric breast density assessment.

2.5 Statistical analysis

Inter- and intra-reader agreement (reproducibility and repeatability) were assessed using the Cohen's kappa statistic. All tests were performed at the 95 per cent confidence interval using STATA (StatCorp, USA) version 13.1.

Table 2 Breast Imaging Reporting and Data System Edition 5 (BI-RADS-5) depicts breast density evaluation

BI-RADS density category	**Breast composition**
A	The breasts are almost entirely fatty
B	Scattered areas of fibroglandular density
C	Heterogeneously dense
D	Extremely dense

3 RESULTS

The final number of patients included in this study was 192, and the mean age of the population was 57.99 (standard deviation of 11.9) years. Eight cases were excluded from the original sample of 200 due to breast abnormalities such as breast inflammation, which would have influenced the reading produced by Quantra software (automated volumetric breast density assessment). Comparisons between the reader (radiologist) and the automated volumetric breast density assessment (Quantra software) are presented in Table 3.1

Table 3.1 Contingency table and Cohen's kappa statistics demonstrate inter-reader agreement between readers and the automated volumetric breast density assessment (Quantra software) of 192 cases

	Agreement	**Kappa**	**Standard error**	**Z**	**Prob > Z**
Reader 1	45.83%	0.1738	0.0413	4.20	0.00000
Reader 2	46.87%	0.1324	0.0499	2.65	0.0040
Reader 3	36.46%	0.0745	0.0433	1.72	0.0427
Reader 4	40.10%	0.0746	0.0471	1.58	0.0565

Note: Kappa = 0.20 (no agreement); Kappa = 0.21-0.39 (minimal); Kappa= 0.40-59 (weak); Kappa = 0.60-0.79 (moderate); Kappa = 0.80-90 (strong); above 0.90 (almost perfect) [10]

Z= is a measure of how many standard deviations below or above the population mean a raw score is

Reader 1 demonstrated the overall agreement of 45.83% and kappa 0.173. The lowest agreement was demonstrated by reader 3 with an agreement of 36.46% and kappa score of 0.074. This demonstrated a significant difference on how senior and junior readers differed from the Quantra software assessment.

Table 3.2 Intra-reader agreement between radiologists, comparing first and second readings, on 100 cases (eight-week time interval).

Reader	**Agreement**	**Kappa**	**Standard error**	**Z**	**Prob > Z**
1) 1st Reading vs 2nd reading	35.00%	0.0160	0.0581	0.28	0.3913
2) 1st Reading vs 2nd reading	60.00%	0.3650	0.0678	5.39	0.0000
3) 1st Reading vs 2nd reading	44.00%	0.1644	0.0652	2.52	0.0058
4) 1st Reading vs 2nd reading	50.00%	0.1917	0.0717	2.67	0.0038

There was poor agreement demonstrated by the readers between the first and second readings within the eight-week interval.

Table 3.3 Comparison between second reading (n = 100) cases and the Quantra software

Reader	**Agreement**	**Kappa**	**Standard error**	**Z**	**Prob > Z**
1) 2^{nd} reading vs machine	47.00%	0.1519	0.0706	0.73	0.2312
2) 2^{nd} reading vs machine	49.00%	0.1638	0.0683	2.40	0.0082
3) 2^{nd} reading vs machine	41.00%	0.1146	0.0615	1.87	0.0311
4) 2^{nd} reading vs machine	37.00%	0.0601	0.0608	0.99	0.1614

The senior readers demonstrated slightly higher agreement on their second reading versus the Quantra software compared to junior readers with the highest to lowest agreement ranging from 47.00% to 37.00% respectively.

4. ETHICAL CONSIDERATIONS

Informed consent was not needed for this retrospective study and data was anonymised. Written consent was obtained from the hospital management and similarly clearance certificate was also obtained from the relevant medical school and university ethics committees prior to the conduction of the study. The clearance no.M171128.

5 DISCUSSION

There are many theories, speculating the causes of discordance in assessing breast density. Many of these theories are attributed to a tendency to over- or underestimate scoring of breast density due to work pressure and fatigue as well as viewing conditions, according to Prinsloo [11].

The aim of this study was to assess concordance in evaluation of breast tissue density between fully automated volumetric breast density assessment and visual assessment by radiologists using the BI-RADS classification. In this study, mammographic density assessment was performed on 192 women that presented for mammograms by applying a fully automated volumetric breast density assessment with a Hologic Quantra 3 software. The results from the automated volumetric breast density were compared to the radiologists' classification according to the BI-RADS (5th edition).

The results demonstrated that agreement between radiologists and categorical automated volumetric breast density assessment was poor; however, it was slightly greater among the two senior radiologists compared to the junior radiologist. The results are demonstrated in table 3.1

This study's findings are in line with previous studies, including a study by Prinsloo [11]. The Prinsloo study is similar to the current study in many ways and is also based in a South African context.

Furthermore, it has also been indicated in other studies that there are many reasons that could explain the lower agreement between automated volumetric breast density assessment and the BI-RADS assessment. One of the reasons is that the BI-RADS categories are based on processed images, while the Quantra software analysis is performed on raw data. Also, the automated volumetric breast density assessment is measured on a continuous scale, while the BI-RADS scores are an estimation of four groups[12].

For an example, according to the 5th BI-RADS classification, if the breasts are predominantly fatty, but there is a focal area of dense fibroglandular tissue accumulation, then the breasts are depicted as heterogeneously dense to show that the area may obscure a mass, even though the percentage of the total breasts consists of fibroglandular tissue at less than fifty per cent.

Eight weeks after the first assessments, the radiologists evaluated only 100 of the original 192 cases for the purpose of intra-reader agreement. The observed agreement was poor. The results indicate that assessing mammographic density is subjective among radiologists. A quantitative study conducted by Sprague showed similar variations in mammographic breast density assessments among radiologists in clinical practice. The study found that patient subgroups, which showed the most extreme variation in density assessment by radiologists were found in the categories of patient age and Body Mass Index (BMI) combination groups. In the Sprague study, 17.2% of women who were undergoing breast examinations interpreted by different radiologists had discordant assessment of dense / non-dense status [13].

Correctly identifying breast density, more specifically heterogeneously /extremely dense breasts will prompt the radiologist to use additional imaging tools (such as breast ultrasound / MRI) to make sure that breast cancer is not missed as it is easily masked by increased breast density. It is worth noting that in our setting, a breast ultrasound examination is routinely performed for all patients presenting for screening mammography.

In another study by Carney, accuracy on screening mammography was observed, based on individual and combined effects of age, breast tissue density and hormone replacement therapy. This study demonstrated that accuracy of screening mammography is better evaluated in older women and in women with fatty breasts. Postmenopausal hormone therapy considerably affects mammography precision, but only through its increase in radiographic breast density[14]. Therefore, this highlights the importance of being cognisant of these factors when evaluating breast density.

The current study strongly indicates that it is pivotal for mammographic units to simultaneously evaluate breast density, utilising both the radiologists' assessment (BI-RADS categorisation) and the automated volumetric breast density measurements as there is observed poor agreement between these two measurements. Recognition of such may further assist in exploring other methods that are more accurate and objective in assessing breast tissue density and thus better predict the risk of developing breast cancer. This will facilitate in decision making in terms of patient personalized imaging and in turn maximize the detection rate of breast cancer.

Furthermore, a quantitative multicentre study conducted in the United Kingdom evaluated outcomes of assessing breast density by comparing two software tools used for the measurement of mammographic breast tissue density with observer-based scores. The conclusion was that automated tools for measuring density demonstrate good correlation, but there is poor correlation with observers' assessment. The study also showed that automated volumetric breast tissue density measurements are objective and not dependent on observer variability. However, different software systems do not produce identical results, in this case (Volpara and Quantra software). Agreement between clinical BI-RADS density classifications and Volpara and Quantra software was (Kappa= 0.57 and 0.46) respectively [12].

It was mentioned that differences are likely to occur because as fibroglandular volume increases (especially > 300 cm^3), there is a discrepancy in ratios of fibroglandular, fat tissue and skin allocated to each pixel. It then becomes harder to identify these reference areas, and each manufacturer's solution is likely to result in differences in the final volumetric density results. Another difference was thought to be caused by how skin and fibroglandular tissue are differentiated in the assessment of dense tissue[12].

In another small study, it was found that results from automated volumetric software tools, Quantra and Volpara software correlated accurately with breast MRI results (r= 0.95) by comparing the performance of these software with fuzzy- clustering segmentation method on MRI. [11].

In addition, the degree of radiologists' subjectivity in assessing breast density was further compounded by poor repeatability of this assessment, when there was a time interval in evaluating the same cases. A study by Prinsloo had similar findings[11].

Literature has shown inconsistency in study outcomes regarding the comparison of breast density using qualitative (visual assessment using BI-RADS classification) versus quantitative (automated volumetric breast density) techniques. Most studies have demonstrated discordance between these measurement tools due to many factors, some of which have been alluded to in the above text. The current study is in line with some of the studies that have found significant and poor inter reader and intra-reader agreement.

Currently, there is no gold standard in assessing breast densities as both the quantitative and the qualitative techniques are used in clinical practice. The use of automated volumetric breast density assessment may be a valuable tool to minimise discordance in assessing breast density due to its objectivity. However, this statement does not disregard the pivotal value of human visual assessment using the BI-RADS classification, as automated volumetric assessment tools also have their own flaws / limitations.

There are also other models that have been used to assess mammographic density, which have been investigated and implemented in routine clinical practice, such as the Deep learning model/algorithm (DL). A study by Lehman, using the DL model and experienced radiologists, demonstrated higher agreement with kappa values of 0.67 (95% CL: 0.66, 0.68) and 0.78 (95% CI: 0.73, 0.82) in the test set and reader studies, and showed a good agreement with kappa values of 0.85 (95% CI: 0.84, 0.86) in the non-blinded implementation setting[15]. This could be another suggested option that may be widely used in assessing breast density. However, time and cost implications need to be considered.

5.1 Limitations of the current study

The sample size in this study was sufficient; however, the number of readers was small (n = 4), and this could limit the generalisation of results in terms of comparison between junior and senior readers. Most of the studies in the literature have larger sample sizes and a larger number of readers (6, 13, 16, 17). Another limitation could be the fact that only 100 cases were used for the second reading (to assess intra-reader agreement) instead of the initial 192 cases. However, this number (n = 100) agreed with the statistician.

5.2 Bias

The senior and junior radiologists used different monitors to assess the cases, 5- BARCO megapixel and 3- BARCO megapixel monitors, respectively. The reason was that the junior readers only had access to 3- BARCO megapixel monitors. This could have introduced bias into the study.

5.3 Recommendations

The results of this study have demonstrated subjectivity of radiologists with poor inter and intra-reader agreement amongst the radiologists and in comparison, with the automated volumetric breast density assessment. These findings are in line with previous studies in the literature.

It is recommended that further research is done that focuses on identifying assessment tools that are more objective in assessing breast density. The deep learning algorithms expressed in the latest literature are headed in the promising direction of mastering the art of objectively assessing breast density.

Furthermore, training of radiologists and providing working environment that is less disruptive, may alleviate some of the stressors that radiologists are faced with and in turn offer optimal assessment of breast density.

Table 5: Summary of Literature Review Results in Context -*continues*

STUDY	**STUDY POPULATION**	**DATA SOURCE**	**STUDY FINDINGS**	**CRITIQUE**	**CURRENT STUDY**
Sartor et al., 2016 [6] Sweden	8782	Mammogram/ fully automated software Volpara	The study showed an agreement between the BI-RADS score of different radiologists to be 80.9% (kappa value 0.77). The agreement between volumetric density and the BI-RADS scores was 57.1% (kappa value 0.55). This substantial agreement between BI-RADS scores shows that radiologists evaluate mammographic density similarly. The moderate agreement between the BI-RADS scores and automated volumetric assessment (Volpara software used) illustrate that radiologists assess mammographic breast density differently to the computer automated programmes	The BI-RADS 4th edition's classification was used during the study. Patients who had breast cancer were excluded from the study as it is known that breast tumours can possibly affect the surrounding breast tissue and mammographic density. The outcome of the study may have been different if these patients were included. Information, including patients' use of hormone replacement therapy or patients' reproductive history, was not acquired. These factors are known to have a possible impact in breast density.	The current study was conducted in Johannesburg, South Africa. The study population included 192 screening cases, instead of a large population as in the comparative study. The readers used the BI-RADS-5 edition classification and a similar fully automated software (Quantra). The agreement between the volumetric density and BI-RADS scores between senior and junior readers ranged between 36-47% (kappa values range 0.07 to 0.17). The current study demonstrates that radiologists assess mammographic breast density differently to the automated software, similarly to the Sartor et al. study. We also excluded patients with proven breast cancer and those who had surgery (mastectomy).

Table 5: Summary of Literature Review Results in Context –*continued*

STUDY	STUDY POPULATION	DATA SOURCE	STUDY FINDINGS	CRITIQUE	CURRENT STUDY
S. Yeon Ko et al. 2014 [18] Korea	1129	Mammogram/ Fully automated software - Volpara	The agreement between breast density evaluations by radiologists and volumetric breast density measurement (VBDM) using four grades (D1/D2/D3/D4) was fair (kappa value = 0.26). When the binary classification was used (D1-2/D3-4), the agreement was moderate (K-value = 0.47). 2.4% (27) of women showed VBDM failure. Factors resulting in this failure included small total breast volume, history of partial mastectomy and high breast density.	The study did not look at the radiologists' inter-observer agreement or intra-observer reproducibility in their evaluations of breast density grade.	One of the objectives of our study was to assess intra-observer reproducibility in the evaluation of breast density.

Table 5: Summary of Literature Review Results in Context –*continued*

STUDY	STUDY POPULATION	DATA SOURCE	STUDY FINDINGS	CRITIQUE	CURRENT STUDY
C. Gard et al. 2015 [7] New Mexico	341	Film mammography/CAD (using per cent density)	Intra-radiologist agreement was moderate-substantial (k-value = 0.50-0.81). Intra-radiologist agreement was higher for radiologist with more than 10-year experience. Inter-observer agreement varied across radiologists (k = value 0.02-0.72). A substantial number of clinical BI-RADS examinations (50%) were assigned density *a* instead of *b* and 40% assigned density *d* instead of *c.* This misclassification has implications in clinical practice.	Film mammograms were used. However, nowadays digital mammography is used. The BI-RADS 4th edition was used during the study.	The current study used digital mammography and the latest BI-RADS-5 edition classification.

Table 5: Summary of Literature Review Results in Context *-continued*

STUDY	STUDY POPULATION	DATA SOURCE	STUDY FINDINGS	CRITIQUE	CURRENT STUDY
M. Lobbes et al. 2011 [18] Netherlands	200	Mammogram/ semi-automated software	Experienced readers showed agreement with automated software readings in 58.5% of cases. Overestimation was 35.5%, underestimation in 6.0% of cases. Inexperienced readers had less accurate results. Agreement between both readers and semi-automated evaluation was moderate, k-value = 0.36 (experienced reader) and 0.2 inexperienced reader). This study concluded that visual assessment of breast density on mammograms is not fully accurate and is observer-dependent	A semi-automated software was used, which is still operator-dependent compared to fully automated software.	

Table 5: Summary of Literature Review Results in Context – *continued*

STUDY	STUDY POPULATION	DATA SOURCE	STUDY FINDINGS	CRITIQUE	CURRENT STUDY
E. Prinsloo et al. (2018) South Africa [11]	100	Mammogram/ fully automated software	The study found that there was poor agreement between the breast densities calculated by computed-automated breast density and the observer-based BI-RADS density scores. There was also significant degree of inter-reader and intra-reader variability in assessment of breast density.	The mammograms that were used were not entirely normal: 66% were normal, 26% had masses and 8% had other findings such as architectural distortion, oedema and calcifications, and it could be argued that the lesions could have caused bias towards denser scoring by the reader.	There are a lot of similarities found in this study compared to the current study. The general population assessed by Prinsloo is similar to some extent to the one in this study. Both studies were conducted in-tertiary institutions and are based in the South African context. The inter-rater and intra-reader agreement is poor in both studies.
Oliver W.E. Morrish et al. 2015 UK [12]	8867	Mammograms and two automated volumetric density assessment software (Volpara and Quantra)	Automated software/techniques for assessing breast density demonstrated good correlation. However, there was poor correlation with these and observer scores		

Table 5: Summary of Literature Review Results in Context *-continued*

STUDY	STUDY POPULATION	DATA SOURCE	STUDY FINDINGS	CRITIQUE	CURRENT STUDY
B.L. Sprague et al. 2016 United states of America [13]	269, 741 mammograms	Mammograms	This study demonstrated that there is a wide disparity in assessment of breast density between radiologists and this should be considered by providers and policy-makers when considering supplemental screening modalities. The possibility that a woman being told that she has dense breast varies considerably depending on which radiologist interprets the mammogram	No comparison to automated volumetric assessment was made.	Sprague et al. is a much bigger study compared to the current study. However, the results show notable inter-rater variability among radiologists, and the current study has to some degree a similar outcome.

6 CONCLUSIONS

The current study showed that visual assessment of breast density on mammograms is observer-dependent and varies among radiologists. Comparison between senior (experienced) and junior (less experienced) radiologists' assessment with categorical volumetric density was documented. Experienced radiologists demonstrated slightly higher agreement with the automated assessment than junior radiologists; however, it was still poor. Furthermore, there was significant intra-reader variation when readers were asked to read the same cases within a specific time interval. This confirms subjectivity of visual assessment by radiologists.

7 REFERENCES

1. Destounis S, Arieno A, Morgan R, Roberts C, Chan A. Qualitative Versus Quantitative Mammographic Breast Density Assessment: Applications for the US and Abroad. Diagnostics (Basel) journal. 2017;7(2): 1-33.

2. Youk JH, Kim SJ, Son EJ, Gweon HM, Kim JA. Comparison of Visual Assessment of Breast Density in BI-RADS 4th and 5th Editions With Automated Volumetric Measurement. AJR Am J Roentgenol. 2017;209(3):703-8.

3. Rao AA, Feneis J, Lalonde C, Ojeda-Fournier H. A Pictorial Review of Changes in the BI-RADS Fifth Edition. Radiographics : a review publication of the Radiological Society of North America, Inc. 2016;36(3):623-39.

4. Ekpo EU, McEntee MF. Measurement of breast density with digital breast tomosynthesis--a systematic review. The British journal of radiology. 2014;87(1043):313-21.

5. Lince-Deroche N, Rayne S, Van Rensburg C, Benn C, Masuku S, Holele P. Breast cancer in South Africa : developing an affordable and achievable plan to improve detection and survival. 2017;2017(1):181-8.

6. Sartor H, Lang K, Rosso A, Borgquist S, Zackrisson S, Timberg P. Measuring mammographic density: comparing a fully automated volumetric assessment versus European radiologists' qualitative classification. European radiology. 2016;26(12):4354-60.

7. Gard CC, Aiello Bowles EJ, Miglioretti DL, Taplin SH, Rutter CM. Misclassification of Breast Imaging Reporting and Data System (BI-RADS) Mammographic Density and Implications for Breast Density Reporting Legislation. The breast journal. 2015;21(5):481-9.

8. Holland K, van Gils CH, Mann RM, Karssemeijer N. Quantification of masking risk in screening mammography with volumetric breast density maps. Breast cancer research and treatment. 2017;162(3):541-8.

9. Boyd NF, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast tissue composition and susceptibility to breast cancer. Journal of the National Cancer Institute. 2010;102(16):1224-37.

10. McHugh ML. Interrater reliability: the kappa statistic. Biochemia medica. 2012;22(3):276-82.

11. Prinsloo E, Minné C, Greeff W. Measuring breast density: Comparing computer-automated breast density quantification with an observer-based method in a South African academic context. 2018. 2018;22(2):1358.

12. Morrish OW, Tucker L, Black R, Willsher P, Duffy SW, Gilbert FJ. Mammographic breast density: comparison of methods for quantitative evaluation. Radiology. 2015;275(2):356-65.

13. Sprague BL, Conant EF, Onega T, Garcia MP, Beaber EF, Herschorn SD, et al. Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice: A Multicenter Observational Study. Annals of internal medicine. 2016;165(7):457-64.

14. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Annals of internal medicine. 2003;138(3):168-75.

15. Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. Radiology. 2019;290(1):52-8.

16. Youk JH, Gweon HM, Son EJ, Kim JA. Automated Volumetric Breast Density Measurements in the Era of the BI-RADS Fifth Edition: A Comparison With Visual Assessment. AJR Am J Roentgenol. 2016;206(5):1056-62.

17. Nicholson BT, LoRusso AP, Smolkin M, Bovbjerg VE, Petroni GR, Harvey JA. Accuracy of assigned BI-RADS breast density category definitions. Academic radiology. 2006;13(9):1143-9.

18. Lobbes MB, Cleutjens JP, Lima Passos V, Frotscher C, Lahaye MJ, Keymeulen KB, et al. Density is in the eye of the beholder: visual versus semi-automated assessment of breast density on standard mammograms. Insights into imaging. 2012;3(1):91-9.

APPENDIX A: ETHICS CLEARANCE CERTIFICATE

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Dr Tha itha George

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M171128

NAME: (Principal Investigator)	Dr Thalitha George
DEPARTMENT:	Radiology Charlotte Maxeke Johannesburg Academic Hospital
PROJECT TITLE:	Comparison of Breast Tissue Density Estimation using Automated Volumetric Breast Density Method versus Radiologist's Visual Assessments by BIRADS Categozation in Female Patients at Charlotte Maxeke Johannesburg Academic Hospital, South Africa
DATE CONSIDERED:	24/11/2017
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr J Smiig and Dr J Heberfeld
APPROVED BY:	Prof C Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	16/05/2018

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third floor Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. I agree to submit a yearly progress report. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in November and will therefore be due in the month of November each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

01-12-2017

Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX B: PLAGIARISM REPORT

1318310:MMED_thesis_-_Thalitha_George_(1).docx

ORIGINALITY REPORT

14%	11%	10%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	pubs.rsna.org Internet Source	3%
2	link.springer.com Internet Source	3%
3	www.journalslibrary.nihr.ac.uk Internet Source	1%
4	www.health.gov.za Internet Source	1%
5	www.ajronline.org Internet Source	1%
6	sajr.org.za Internet Source	1%
7	Lecture Notes in Computer Science, 2016. Publication	1%
8	www.hologic.com Internet Source	1%
9	M. B. I. Lobbes, J. P. M. Cleutjens, V. Lima Passos, C. Frotscher et al. "Density is in the eye	<1%

APPENDIX C: RAW DATA

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continues*

Case	Machine	Reader 1	Reader 2	Reader 3	Reader 4
1	Scattered areas of fibroglandular tissue	B	B	C	B
2	Scattered areas of fibroglandular tissue	B	B	B	B
3	Scattered areas of fibroglandular tissue	B	B	A	A
4	Scattered areas of fibroglandular tissue	B	C	B	C
5	Heterogenously dense	B	C	C	C
6	Heterogenously dense	B	A	A	A
7	Heterogenously dense	C	C	A	D
8	Almost entirely fatty	A	B	C	A
9	Scattered areas of fibroglandular tissue	B	B	A	D
10	Scattered areas of fibroglandular tissue	B	B	A	A
11	Heterogenously dense	C	C	B	C
12	Scattered areas of fibroglandular tissue	A	A	A	A
13	Scattered areas of fibroglandular tissue	B	B	A	A
14	Scattered areas of fibroglandular tissue	A	A	A	A
15	Scattered areas of fibroglandular tissue	B	B	B	B
16	Almost entirely fatty	A	A	A	A
17	Scattered areas of fibroglandular tissue	B	C	B	B
18	Scattered areas of fibroglandular tissue	B	B	B	B
19	Scattered areas of fibroglandular tissue	A	A	A	A
20	Almost entirely fatty	D	D	D	D
21	Scattered areas of fibroglandular tissue	B	B	B	B
22	Almost entirely fatty	B	B	C	B
23	Scattered areas of fibroglandular tissue	B	C	C	B
24	Heterogenously dense	B	C	B	B
25	Almost entirely fatty	B	B	B	B
26	Scattered areas of fibroglandular tissue	B	B	C	B

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	Machine	Reader 1	Reader 2	Reader 3	Reader 4
27	Scattered areas of fibroglandular tissue	B	B	B	B
28	Scattered areas of fibroglandular tissue	B	B	B	A
29	Scattered areas of fibroglandular tissue	B	C	C	B
30	Scattered areas of fibroglandular tissue	B	B	B	B
31	Scattered areas of fibroglandular tissue	B	B	A	A
32	Scattered areas of fibroglandular tissue	A	A	A	A
33	Heterogenously dense	C	C	C	B
34	Almost entirely fatty	B	B	B	A
35	Scattered areas of fibroglandular tissue	C	C	C	B
36	Scattered areas of fibroglandular tissue	B	B	B	B
37	Scattered areas of fibroglandular tissue	B	B	A	A
38	Heterogenously dense	D	C	D	C
39	Scattered areas of fibroglandular tissue	C	C	D	B
40	Scattered areas of fibroglandular tissue	B	B	B	B

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
41	Scattered areas of fibroglandular tissue	D	C	D	B
42	Scattered areas of fibroglandular tissue	B	B	B	A
43	Almost entirely fatty	A	A	A	A
44	Almost entirely fatty	A	A	A	A
45	Scattered areas of fibroglandular tissue	B	B	B	B
46	Scattered areas of fibroglandular tissue	B	B	B	B
47	Almost entirely fatty	B	B	A	B
48	Scattered areas of fibroglandular tissue	A	A	A	A
49	Scattered areas of fibroglandular tissue	B	B	A	A
50	Scattered areas of fibroglandular tissue	B	B	B	B
51	Scattered areas of fibroglandular tissue	B	B	A	A
52	Scattered areas of fibroglandular tissue	B	B	B	A
53	Heterogenously dense	D	D	D	C
54	Scattered areas of fibroglandular tissue	B	B	B	B
55	Scattered areas of fibroglandular tissue	A	A	A	A

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
56	Almost entirely fatty	A	A	A	A
57	Almost entirely fatty	A	A	A	A
58	Almost entirely fatty	A	A	A	A
59	Scattered areas of fibroglandular tissue	A	B	A	B
60	Scattered areas of fibroglandular tissue	B	B	B	A
61	Scattered areas of fibroglandular tissue	C	C	C	B
62	Almost entirely fatty	A	A	A	A
63	Scattered areas of fibroglandular tissue	B	B	B	A
64	Almost entirely fatty	B	B	A	A
65	Heterogenously dense	C	C	B	C
66	Scattered areas of fibroglandular tissue	C	C	B	B
67	Scattered areas of fibroglandular tissue	B	B	B	B
68	Scattered areas of fibroglandular tissue	A	A	A	A
69	Scattered areas of fibroglandular tissue	A	A	A	A
70	Almost entirely fatty	A	A	A	A
71	Almost entirely fatty	B	B	A	A

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	Machine	Reader 1	Reader 2	Reader 3	Reader 4
72	Extremely dense	D	D	D	C
73	Scattered areas of fibroglandular tissue	B	B	B	B
74	Almost entirely fatty	B	B	A	A
75	Heterogenously dense	B	C	B	B
76	Heterogenously dense	A	A	A	A
77	Scattered areas of fibroglandular tissue	A	A	A	A
78	Almost entirely fatty	B	B	A	B
79	Scattered areas of fibroglandular tissue	B	B	A	B
80	Scattered areas of fibroglandular tissue	B	B	A	B
81	Scattered areas of fibroglandular tissue	B	B	B	B
82	Scattered areas of fibroglandular tissue	B	B	A	A
83	Heterogenously dense	C	B	B	B
84	Scattered areas of fibroglandular tissue	B	C	D	B
85	Almost entirely fatty	A	B	B	A
86	Scattered areas of fibroglandular tissue	B	A	A	C
87	Scattered areas of fibroglandular tissue	B	C	C	B
88	Almost entirely fatty	A	B	B	A
89	Scattered areas of fibroglandular tissue	B	A	A	A

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
90	Almost entirely fatty	A	B	A	A
91	Heterogenously dense	C	A	A	A
92	Scattered areas of fibroglandular tissue	B	C	C	C
93	Scattered areas of fibroglandular tissue	B	B	A	A
94	Heterogenously dense	B	B	A	A
95	Scattered areas of fibroglandular tissue	B	B	A	A
96	Scattered areas of fibroglandular tissue	B	B	B	B
97	Heterogenously dense	C	B	B	B
98	Almost entirely fatty	A	C	C	B
99	Almost entirely fatty	B	B	A	A
100	Heterogenously dense	C	B	A	A
101	Scattered areas of fibroglandular tissue	B	C	C	C
102	Scattered areas of fibroglandular tissue	A	B	A	A
103	Almost entirely fatty	B	A	A	A
104	Scattered areas of fibroglandular tissue	C	B	A	A
105	Heterogenously dense	C	C	D	C
106	Almost entirely fatty	A	C	C	C
107	Scattered areas of fibroglandular tissue	B	A	A	A
108	Almost entirely fatty	B	B	C	B

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
109	Almost entirely fatty	A	B	A	B
110	Scattered areas of fibroglandular tissue	A	B	A	A
111	Scattered areas of fibroglandular tissue	B	B	A	A
112	Almost entirely fatty	B	B	C	B
113	Scattered areas of fibroglandular tissue	B	B	B	A
114	Scattered areas of fibroglandular tissue	B	B	B	B
115	Heterogenously dense	C	C	B	B
116	Heterogenously dense	C	C	C	B
117	Scattered areas of fibroglandular tissue	A	C	C	C
118	Scattered areas of fibroglandular tissue	B	A	A	A
119	Scattered areas of fibroglandular tissue	A	B	B	A
120	Scattered areas of fibroglandular tissue	B	B	A	A
121	Almost entirely fatty	B	B	A	A
122	Scattered areas of fibroglandular tissue	C	B	B	B
123	Scattered areas of fibroglandular tissue	C	C	C	B
124	Scattered areas of fibroglandular tissue	A	C	B	B

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
125	Almost entirely fatty	B	A	A	A
126	Almost entirely fatty	B	B	B	A
127	Heterogenously dense	C	B	B	A
128	Scattered areas of fibroglandular tissue	B	C	D	B
129	Scattered areas of fibroglandular tissue	B	B	A	A
130	Scattered areas of fibroglandular tissue	B	B	B	B
131	Almost entirely fatty	B	B	A	A
132	Scattered areas of fibroglandular tissue	B	A	A	A
133	Almost entirely fatty	B	B	B	A
134	Scattered areas of fibroglandular tissue	B	B	A	A
135	Scattered areas of fibroglandular tissue	C	B	A	A
136	Scattered areas of fibroglandular tissue	A	C	B	C
137	Scattered areas of fibroglandular tissue	A	A	A	A
138	Scattered areas of fibroglandular tissue	A	A	A	A
139	Almost entirely fatty	A	A	A	A
140	Scattered areas of fibroglandular tissue	B	A	A	A
141	Heterogenously dense	B	B	B	A
142	Scattered areas of fibroglandular tissue	B	B	B	A

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
143	Heterogenously dense	C	A	A	B
144	Almost entirely fatty	B	C	D	A
145	Scattered areas of fibroglandular tissue	C	B	A	B
146	Scattered areas of fibroglandular tissue	B	C	C	A
147	Heterogenously dense	C	B	C	B
148	Heterogenously dense	C	C	C	A
149	Scattered areas of fibroglandular tissue	A	C	B	C
150	Scattered areas of fibroglandular tissue	B	A	A	A
151	Heterogenously dense	C	B	B	A
152	Almost entirely fatty	A	C	D	B
153	Heterogenously dense	D	A	A	B
154	Scattered areas of fibroglandular tissue	B	D	D	A
155	Scattered areas of fibroglandular tissue	B	B	A	C
156	Scattered areas of fibroglandular tissue	B	B	A	A
157	Almost entirely fatty	C	B	B	A
158	Scattered areas of fibroglandular tissue	B	B	B	C
159	Almost entirely fatty	B	C	C	B
160	Scattered areas of fibroglandular tissue	B	B	A	B
161	Scattered areas of fibroglandular tissue	B	C	B	B
162	Scattered areas of fibroglandular tissue	C	B	A	A
163	Almost entirely fatty	A	C	C	B
164	Heterogenously dense	B	B	A	B
165	Heterogenously dense	C	B	B	B
166	Scattered areas of fibroglandular tissue	B	D	D	A
167	Extremely dense	C	A	A	B
168	Scattered areas of fibroglandular tissue	C	C	D	A
169	Heterogenously dense	C	C	C	C
170	Scattered areas of fibroglandular tissue	B	C	B	C
171	Extremely dense	C	C	B	C
172	Scattered areas of fibroglandular tissue	A	C	D	B
173	Scattered areas of fibroglandular tissue	B	B	D	C

Table 6: Raw data demonstrating inter-reader agreement. Readers' assessment of breast density is compared to machine readings (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
174	Heterogenously dense	B	B	B	B
175	Heterogenously dense	C	C	B	B
176	Scattered areas of fibroglandular tissue	B	C	B	B
177	Heterogenously dense	B	B	A	A
178	Heterogenously dense	B	B	B	B
179	Heterogenously dense	B	B	B	B
180	Almost entirely fatty	B	B	B	B
181	Heterogenously dense	B	A	A	A
182	Heterogenously dense	C	C	C	B
183	Scattered areas of fibroglandular tissue	B	C	B	B
184	Heterogenously dense	C	B	B	B
185	Heterogenously dense	C	C	C	B
186	Scattered areas of fibroglandular tissue	B	C	C	C
187	Scattered areas of fibroglandular tissue	A	C	D	B
188	Almost entirely fatty	B	A	A	B
189	Scattered areas of fibroglandular tissue	B	B	A	A
190	Scattered areas of fibroglandular tissue	B	B	A	A
191	Scattered areas of fibroglandular tissue	B	A	A	A
192	Scattered areas of fibroglandular tissue	B	B	B	A

Table 7: Demonstrates intra-reader agreement between radiologists. The radiologists' first readings are compared to their second readings within an interval of eight weeks, using the same cases, but only the first 100 cases were compared (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continues*

	READER 1		**READER 2**		**READER 3**		**READER 4**	
Case	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**
1	B	B	B	B	C	D	B	B
2	B	B	B	C	B	C	B	B
3	B	B	B	B	A	A	A	A
4	B	C	C	C	B	D	C	C
5	B	C	C	C	C	D	C	C
6	B	B	A	B	A	A	A	A
7	C	C	C	C	A	C	D	D
8	A	A	B	B	C	A	A	A
9	B	B	B	B	A	B	D	A
10	B	B	B	B	A	A	A	B
11	C	B	C	B	B	A	C	A
12	A	C	A	C	A	C	A	A
13	B	A	B	A	A	A	A	D
14	A	B	A	B	A	A	A	A

Table 7: Demonstrates intra-reader agreement between radiologists. The radiologists' first readings are compared to their second readings within an interval of eight weeks, using the same cases, but only the first 100 cases were compared (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

	READER 1		**READER 2**		**READER 3**		**READER 4**	
Case	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**
15	B	A	B	A	B	A	B	A
16	A	B	A	B	A	B	A	A
17	B	A	C	A	B	A	B	B
18	B	B	B	B	B	B	B	A
19	A	B	A	B	A	B	A	B
20	D	B	D	B	D	A	D	A
21	B	D	B	D	B	D	B	A
22	B	B	B	C	C	C	B	D
23	B	B	C	B	C	C	B	B
24	B	B	C	C	B	C	B	B
25	B	C	B	B	B	C	B	B
26	B	B	B	B	C	C	B	C
27	B	B	B	B	B	C	B	B
28	B	B	B	B	B	B	A	B

Table 7: Demonstrates intra-reader agreement between radiologists. The radiologists' first readings are compared to their second readings within an interval of eight weeks, using the same cases, but only the first 100 cases were compared (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

	READER 1		**READER 2**		**READER 3**		**READER 4**	
Case	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**
29	B	B	C	C	C	B	B	B
30	B	B	B	A	B	C	B	B
31	B	B	B	B	A	B	A	B
32	A	B	A	A	A	A	A	A
33	C	A	C	C	C	A	B	A
34	B	C	B	B	B	C	A	B
35	C	B	C	C	C	A	B	A
36	B	C	B	B	B	D	B	C
37	B	B	B	B	A	B	A	A
38	D	B	C	C	D	A	C	C
39	C	C	C	C	D	D	B	B
40	B	C	B	B	B	B	B	A
41	D	B	C	C	D	B	B	D

Table 7: Demonstrates intra-reader agreement between radiologists. The radiologists' first readings are compared to their second readings within an interval of eight weeks, using the same cases, but only the first 100 cases were compared (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

	READER 1		**READER 2**		**READER 3**		**READER 4**	
Case	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**
42	B	C	B	B	B	D	A	C
43	A	B	A	A	A	A	A	B
44	A	A	A	A	A	A	A	D
45	B	A	B	C	B	A	B	A
46	B	C	B	B	B	B	B	A
47	B	B	B	B	A	B	B	A
48	A	B	A	A	A	B	A	B
49	B	A	B	B	A	A	A	B
50	B	B	B	B	B	B	B	A
51	B	B	B	B	A	B	A	C
52	B	B	B	B	B	A	A	A
53	D	B	D	D	D	B	C	A
54	B	C	B	B	B	D	B	A
55	A	B	A	A	A	A	A	A
56	A	A	A	A	A	A	A	B
57	A	A	A	A	A	A	A	B
58	A	B	A	A	A	A	A	A

Table 7: Demonstrates intra-reader agreement between radiologists. The radiologists' first readings are compared to their second readings within an interval of eight weeks, using the same cases, but only the first 100 cases were compared (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

	READER 1		**READER 2**		**READER 3**		**READER 4**	
Case	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**
59	A	A	B	A	A	A	B	D
60	B	B	B	B	B	A	A	A
61	C	B	C	C	C	B	B	A
62	A	C	A	A	A	B	A	A
63	B	A	B	B	B	A	A	A
64	B	B	B	B	A	B	A	A
65	C	B	C	B	B	A	C	A
66	C	C	C	C	B	B	B	A
67	B	B	B	C	B	B	B	B
68	A	B	A	C	A	B	A	A
69	A	A	A	A	A	A	A	A
70	A	B	A	A	A	A	A	A
71	B	A	B	A	A	A	A	C
72	D	B	D	B	D	D	C	C
73	B	D	B	D	B	A	B	B
74	B	B	B	C	A	B	A	A
75	B	B	C	B	B	B	B	A
76	A	C	A	C	A	C	A	A

Table 7: Demonstrates intra-reader agreement between radiologists. The radiologists' first readings are compared to their second readings within an interval of eight weeks, using the same cases, but only the first 100 cases were compared (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

	READER 1		**READER 2**		**READER 3**		**READER 4**	
Case	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**	**First reading**	**Second reading**
77	A	B	A	A	A	A	A	A
78	B	B	B	A	A	A	B	D
79	B	B	B	B	A	B	B	B
80	B	B	B	B	A	B	B	B
81	B	B	B	B	B	B	B	A
82	B	B	B	B	A	B	A	B
83	C	B	B	B	B	B	B	B
84	B	B	C	B	D	D	B	C
85	A	C	B	C	B	B	A	B
86	B	B	A	B	A	B	C	B
87	B	A	C	A	C	A	B	A
88	A	C	B	C	B	B	A	A
89	B	B	A	B	A	B	A	B
90	A	B	B	A	A	A	A	A
91	C	B	A	B	A	B	A	A
92	B	A	C	A	C	A	C	B
93	B	C	B	D	A	D	A	A
94	B	B	B	A	A	A	A	A
95	B	B	B	B	A	B	A	B
96	B	B	B	B	B	B	B	A
97	C	C	B	B	B	B	B	B
98	A	B	C	B	C	C	B	B
99	B	C	B	C	A	A	A	A
100	C	B	B	B	A	A	A	A

Table 8: Comparison between radiologists' second reading and the machine (automated volumetric breast density assessment), 100 cases (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continues*

		Reading 2			
Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
1	Scattered areas of fibroglandular tissue	B	B	D	B
2	Scattered areas of fibroglandular tissue	B	C	C	B
3	Scattered areas of fibroglandular tissue	B	B	A	A
4	Scattered areas of fibroglandular tissue	C	C	D	C
5	Heterogenously dense	C	C	D	C
6	Heterogenously dense	B	B	A	A
7	Heterogenously dense	C	C	C	D
8	Almost entirely fatty	A	B	A	A
9	Scattered areas of fibroglandular tissue	B	B	B	A
10	Scattered areas of fibroglandular tissue	B	B	A	B
11	Heterogenously dense	B	B	A	A
12	Scattered areas of fibroglandular tissue	C	C	C	A
13	Scattered areas of fibroglandular tissue	A	A	A	D
14	Scattered areas of fibroglandular tissue	B	B	A	A
15	Scattered areas of fibroglandular tissue	A	A	A	A
16	Almost entirely fatty	B	B	B	A
17	Scattered areas of fibroglandular tissue	A	A	A	B
18	Scattered areas of fibroglandular tissue	B	B	B	A
19	Scattered areas of fibroglandular tissue	B	B	B	B
20	Almost entirely fatty	B	B	A	A
21	Scattered areas of fibroglandular tissue	D	D	D	A
22	Almost entirely fatty	B	C	C	D
23	Scattered areas of fibroglandular tissue	B	B	C	B
24	Heterogenously dense	B	C	C	B
25	Almost entirely fatty	C	B	C	B
26	Scattered areas of fibroglandular tissue	B	B	C	C
27	Scattered areas of fibroglandular tissue	B	B	C	B
28	Scattered areas of fibroglandular tissue	B	B	B	B

Table 8: Comparison between radiologists' second reading and the machine (automated volumetric breast density assessment), 100 cases (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

		Reading 2			
Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
29	Scattered areas of fibroglandular tissue	B	C	B	B
30	Scattered areas of fibroglandular tissue	B	A	C	B
31	Scattered areas of fibroglandular tissue	B	B	B	B
32	Scattered areas of fibroglandular tissue	B	A	A	A
33	Heterogenously dense	A	C	A	A
34	Almost entirely fatty	C	B	C	B
35	Scattered areas of fibroglandular tissue	B	C	A	A
36	Scattered areas of fibroglandular tissue	C	B	D	C
37	Scattered areas of fibroglandular tissue	B	B	B	A
38	Heterogenously dense	B	C	A	C
39	Scattered areas of fibroglandular tissue	C	C	D	B
40	Scattered areas of fibroglandular tissue	C	B	B	A
41	Scattered areas of fibroglandular tissue	B	C	B	D
42	Scattered areas of fibroglandular tissue	C	B	D	C
43	Almost entirely fatty	B	A	A	B
44	Almost entirely fatty	A	A	A	D
45	Scattered areas of fibroglandular tissue	A	C	A	A
46	Scattered areas of fibroglandular tissue	C	B	B	A
47	Almost entirely fatty	B	B	B	A
48	Scattered areas of fibroglandular tissue	B	A	B	B
49	Scattered areas of fibroglandular tissue	A	B	A	B
50	Scattered areas of fibroglandular tissue	B	B	B	A
51	Scattered areas of fibroglandular tissue	B	B	B	C
52	Scattered areas of fibroglandular tissue	B	B	A	A
53	Heterogenously dense	B	D	B	A
54	Scattered areas of fibroglandular tissue	C	B	D	A
55	Scattered areas of fibroglandular tissue	B	A	A	A
56	Almost entirely fatty	A	A	A	B
57	Almost entirely fatty	A	A	A	B
58	Almost entirely fatty	B	A	A	A

Table 8: Comparison between radiologists' second reading and the machine (automated volumetric breast density assessment), 100 cases (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

		Reading 2			
Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
59	Scattered areas of fibroglandular tissue	A	A	A	D
60	Scattered areas of fibroglandular tissue	B	B	A	A
61	Scattered areas of fibroglandular tissue	B	C	B	A
62	Almost entirely fatty	C	A	B	A
63	Scattered areas of fibroglandular tissue	A	B	A	A
64	Almost entirely fatty	B	B	B	A
65	Heterogenously dense	B	B	A	A
66	Scattered areas of fibroglandular tissue	C	C	B	A
67	Scattered areas of fibroglandular tissue	B	C	B	B
68	Scattered areas of fibroglandular tissue	B	C	B	A
69	Scattered areas of fibroglandular tissue	A	A	A	A
70	Almost entirely fatty	B	A	A	A
71	Almost entirely fatty	A	A	A	C
72	Extremely dense	B	B	D	C
73	Scattered areas of fibroglandular tissue	D	D	A	B
74	Almost entirely fatty	B	C	B	A
75	Heterogenously dense	B	B	B	A
76	Heterogenously dense	C	C	C	A
77	Scattered areas of fibroglandular tissue	B	A	A	A
78	Almost entirely fatty	B	A	A	D
79	Scattered areas of fibroglandular tissue	B	B	B	B
80	Scattered areas of fibroglandular tissue	B	B	B	B
81	Scattered areas of fibroglandular tissue	B	B	B	A
82	Scattered areas of fibroglandular tissue	B	B	B	B
83	Heterogenously dense	B	B	B	B
84	Scattered areas of fibroglandular tissue	B	B	D	C

Table 8: Comparison between radiologists' second reading and the machine (automated volumetric breast density assessment), 100 cases (A= almost entirely fatty, B= scattered areas of fibroglandular tissue, C= heterogenously dense, D= extremely dense) - *continued*

		Reading 2			
Case	**Machine**	**Reader 1**	**Reader 2**	**Reader 3**	**Reader 4**
85	Almost entirely fatty	C	C	B	B
86	Scattered areas of fibroglandular tissue	B	B	B	B
87	Scattered areas of fibroglandular tissue	A	A	A	A
88	Almost entirely fatty	C	C	B	A
89	Scattered areas of fibroglandular tissue	B	B	B	B
90	Almost entirely fatty	B	A	A	A
91	Heterogenously dense	B	B	B	A
92	Scattered areas of fibroglandular tissue	A	A	A	B
93	Scattered areas of fibroglandular tissue	C	D	D	A
94	Heterogenously dense	B	A	A	A
95	Scattered areas of fibroglandular tissue	B	B	B	B
96	Scattered areas of fibroglandular tissue	B	B	B	A
97	Heterogenously dense	C	B	B	B
98	Almost entirely fatty	B	B	C	B
99	Almost entirely fatty	C	C	A	A
100	Heterogenously dense	B	B	A	A