


Breast cancer detection across dense and non-dense breasts: Markers of diagnostic confidence and efficacy

Acta Radiologica Open
11(1) 1–9
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20584601211072279
journals.sagepub.com/home/arr


Ibrahim Hadadi^{1,2} , William Rae¹, Jillian Clarke¹, Mark McEntee^{1,3}  and Ernest Ekpo^{1,4}

Abstract

Background: The impact of radiologists' characteristics has become a major focus of recent research. However, the markers of diagnostic efficacy and confidence in dense and non-dense breasts are poorly understood.

Purpose: This study aims to assess the relationship between radiologists' characteristics and diagnostic performance across dense and non-dense breasts.

Materials and methods: Radiologists specialising in breast imaging ($n = 128$) who had 0.5–40 (13 ± 10.6) years of experience reading mammograms were recruited. Participants independently interpreted a test set containing 60 digital mammograms (40 normal and 20 abnormal) with similarly distributed breast densities. Diagnostic performance measures were analysed via Jamovi software (version 1.6.22).

Results: In dense breasts, breast-imaging fellowship completion significantly improved specificity ($p = 0.004$), location sensitivity ($p = 0.01$) and the area under the curve (AUC) of the receiver operating characteristic ($p = 0.03$). Only participation in BreastScreen reading significantly improved all performance metrics: specificity ($p = 0.04$), sensitivity ($p = 0.005$), location sensitivity ($p < 0.001$) and AUC ($p < 0.001$). Reading > 100 mammograms weekly significantly improved sensitivity ($p = 0.03$), location sensitivity ($p = 0.001$), and AUC ($p = 0.03$). In non-dense breasts, breast fellowship completion significantly improved sensitivity ($p = 0.02$), location sensitivity ($p = 0.04$) and AUC ($p = 0.002$). Participation in BreastScreen reading and reading > 100 mammograms weekly significantly improved only sensitivity ($p = 0.002$ and $p = 0.003$, respectively) and location sensitivity ($p < 0.001$ and $p < 0.001$, respectively).

Conclusion: Participating in screening programs, breast fellowships and reading > 100 mammograms weekly are important indicators of the diagnostic performance of radiologists across dense and non-dense breasts. In dense breasts, optimal performance resulted from participation in a breast screening program.

Keywords

Mammogram readers, observer performance, mammography, breast density, breast cancer

Received 23 September 2021; Accepted 17 December 2021

¹Medical Image Optimisation and Perception Group, Discipline of Medical Imaging Science, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

²Department of Radiological Sciences, Faculty of Applied Medical Sciences, King Khalid University, Saudi Arabia

³Discipline of Diagnostic Radiography, University College Cork, Cork, Ireland

⁴Orange Radiology, Laboratories and Research Centre, Calabar, Nigeria

Corresponding author:

Ibrahim Hadadi, Medical Image Optimisation and Perception Group, Faculty of Medicine and Health, The University of Sydney, Susan Wakil Health Building, Camperdown Campus, Western Avenue, Sydney, NSW 2006, Australia.

Email: ihad5571@uni.sydney.edu.au



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

Globally, breast cancer is the second most common form of cancer.¹ It is also the second leading cause of cancer-related death among Australian women.² Currently, mammography is the only modality considered effective for breast cancer screening as it has been proven to reduce breast cancer mortality, particularly among women aged 50 to 69.³

Many factors impact the outcome of screening; these include intrinsic limitations of technology, the experience of those interpreting the mammogram, lesion characteristics, and breast density.⁴ Researchers often refer to dense breasts as 'heterogeneously dense' and 'extremely dense' and non-dense breasts as 'fatty' and 'scattered areas of fibroglandular density'. In the United States, it has been estimated that 43% of the screened population has dense breasts.⁵ Chinese and Korean women have a higher prevalence of dense breasts than US women, with 49.2% and 54.4%, respectively.^{6,7} Data on the Australian population is not available, perhaps due to the lack of breast density notification policy. However, Breast Cancer Network Australia is advocating for breast density policy changes.⁸

Increased breast density increases the risk of masking and of cancer.⁹ It has also been shown that dense breast composition is directly linked to risks associated with breast cancer,¹⁰ emphasising the need to optimise early detection in women with dense breasts. There is evidence of wide variation in diagnostic efficacy between radiologists or breast image readers.¹¹ This inter-reader variability requires that intrinsic human factors be considered when designing strategies to improve breast cancer detection. Thus, the impact of radiologists' characteristics has become a major focus of recent research.¹²⁻¹⁴

Several studies have examined the association between reader characteristics such as years of reading mammograms, the number of mammograms read per year, completion of a fellowship in breast imaging and participation in diagnostic workups.^{11,13-17} These studies demonstrated wide variation in the relationships between observers' characteristics and performance in mammography interpretation. However, most published studies assessed the influence of readers' characteristics with little or no consideration for the impact of breast composition. Thus, the markers of diagnostic efficacy and confidence in dense and non-dense breasts are poorly understood, and further in-depth investigation is needed. Therefore, this study aims to assess the relationship between radiologists' characteristics and diagnostic performance across dense and non-dense breasts.

Method and materials

Image test sets

Two digital mammography (DM) test sets were developed from a screening population database. Each DM test set contained 60 cases (40 normal and 20 abnormal). The

normal cases were confirmed to be normal by at least two radiologists and by a follow-up negative mammogram obtained 2–4 years later. The types and characteristics of the lesions were also established by these radiologists. The abnormal cases contained at least one biopsy-proven cancer lesion. Density classification was determined by a consensus of two consultant breast radiologists with more than 20 years of experience in reading screening mammograms. The cases exhibited a range of breast densities classified as non-dense ($\leq 50\%$ glandular tissue) and dense breasts ($> 50\%$ glandular tissue). Australia uses the Royal Australian and New Zealand College of Radiologists (RANZCR) synoptic scale, which is similar to the fourth edition BI-RADS Atlas. To make the test set relevant to countries using the RANZCR synoptic scale and BI-RADS fourth edition, breast density of the cases included was classified according to the fourth edition BI-RADS Atlas.¹⁸ The distribution of breast densities was as follows: DM test set 1 included 40% non-dense and 60% dense cases, while DM test set 2 included 45% non-dense and 55% dense cases.

Participants

A total of 128 radiologists specialising in breast imaging were recruited. The mean age of the participants was 53 ± 11.8 years, and their years of reading mammograms ranged from 0.5 to 40 (mean: 13 ± 10.6 years). All participants had completed a training program overseen by the RANZCR. The training program runs for 5 years and includes system-focused rotations in the last 2 years. By the end of their training, radiologists are competent in diagnostic breast imaging, including mammography and ultrasound, and have exposure to breast MRI and to the investigation and staging of metastatic breast cancer. A total of 730 breast cases are studied during the training program (100 diagnostic mammograms, 500 screening mammograms, 100 ultrasounds, 20 MRI and 10 biopsies). The fellowship-trained breast radiologists had an additional 6 months training to specialise in advanced breast imaging and breast procedures. Of the 128 radiologists, 62 (48.4%) read for the national Australian breast screening program (BreastScreen, Australia) at the time of the study, and 42 (32.8%) had completed a 3- to 6-month breast-imaging fellowship. The characteristics of the participants are summarised in [Table 1](#).

Reading environment

The images were read via the Breast Reader Assessment Strategy (BREAST) platform either at a conference or at different Australian clinical sites using primary displays between 2015 and 2019. Ambient lighting in reading rooms at conferences was set at 15–20 lux to conform with the RANZCR and BreastScreen Australia Accreditation Standards^{19,20} as well as with ambient lighting conditions in

Table 1. Radiologists' demographic information at the time of completing the DM test set.

Radiologists' characteristics	DM test set 1	DM test set 2	Total
Mean age (years)	55 (± 11.6)	48.7 (± 11.3)	53 (± 11.8)
Number of radiologists (M, F)	87 (35, 52)	41 (19, 22)	128 (54, 74)
Breast speciality	100%	100%	100%
Breast screen program readers	55.2%	34.1%	48.4%
Radiologists completed breast fellowship (3 to 6 months)	33.3%	31.7%	32.8%
Mean years reading mammograms	14.5 (± 10.8)	9.6 (± 9.3)	13 (± 10.6)
Number of cases radiologists read weekly			
< 20	21 (24%)	20 (48.8%)	41 (32%)
20–60	17 (19.5%)	5 (12.2%)	22 (17.2%)
61–100	9 (10.4%)	6 (14.6%)	15 (11.7%)
101–150	11 (12.7%)	1 (2.4%)	12 (9.4%)
151–200	16 (18.4%)	6 (14.6%)	22 (17.2%)
> 200	13 (15%)	3 (7.4%)	16 (12.5%)

many clinical settings. Calibrated Barco 5MP medical-grade monochrome liquid crystal display monitors with a resolution of 2049×2560 pixels were used.

Study design

Participants completed an electronic survey of their demographic information and work experience - including position, speciality, completion of a fellowship in breast radiology, number of years reading mammography and the number of cases read per week. Subsequently, each reader independently interpreted the images in the test sets and assigned a confidence rating to each decision. If the reader considered the image to be normal, he/she moved to the next case and the case was automatically rated as 1, meaning no cancer was present. If a lesion was detected, the reader marked the lesion's location and assigned a confidence rating score from 2 to 5, which is compatible with the Tabar/RANZCR classification used in BreastScreen Australia where 2 = benign, 3 = indeterminate/equivocal, 4 = suspicious and 5 = highly suspicious.

A rating of 3, 4 or 5 signified malignancy, with higher ratings denoting higher confidence. If a rating of 3 or above was given, the reader was asked to describe the type of breast lesion detected (discrete mass, architectural distortion, spiculated mass, nonspecific density, stellate and calcification) by checking the appropriate box in a pop-up menu. These marks and ratings were then used to assess reader performance.

Statistical analysis

The radiologists' performances were calculated in terms of specificity, sensitivity, location sensitivity and AUC in dense and non-dense breasts. Location sensitivity was

determined by the distance of the mouse click from the breast lesion centre. If distance was not recorded, this indicated that the radiologist marked outside the correct region or did not give any markings. Diagnostic confidence (radiologists' level of confidence that the detected lesion was malignant) and lesion classification (their ability to correctly classify the lesion into type) in dense and non-dense images were calculated.

The diagnostic performance metrics were compared using an independent-samples *t*-test or a Mann–Whitney *U* test, depending on the distribution of the data. A chi-squared test (χ^2) was conducted to assess the association between radiologists' characteristics and both diagnostic confidence and lesion classification across dense and non-dense breasts. One-way ANOVA and Kruskal–Wallis tests were applied to compare three independent groups depending on data distribution. *p*-value ≤ 0.05 was considered statistically significant. These analyses were conducted via Jamovi software (version 1.6.22).

Results

Radiologists' performances in dense breasts

Table 2 depicts the differences in radiologists' characteristics and performances with DM for dense breasts. All metrics – including specificity ($p = 0.26$), sensitivity ($p = 0.13$), location sensitivity ($p = 0.14$) and AUC ($p = 0.84$) – were similar between radiologists who had read mammography for < 10 years, $10–19$ years and ≥ 20 years. Specificity ($p = 0.004$), location sensitivity ($p = 0.01$) and AUC ($p = 0.03$), but not sensitivity ($p = 0.36$), were significantly higher in radiologists who had completed a breast fellowship than in those who had not. Radiologists who work for the BreastScreen Australia program showed

significantly higher performance in all metrics – specificity ($p = 0.04$), sensitivity ($p = 0.005$), location sensitivity ($p < 0.001$) and AUC ($p < 0.001$) – compared to those who do not. Radiologists who read > 100 mammogram cases weekly showed significantly higher sensitivity ($p = 0.03$), location sensitivity ($p = 0.001$) and AUC ($p = 0.03$), but not specificity ($p = 0.51$).

Radiologists' performances in non-dense breasts

Table 3 shows the differences in radiologists' characteristics and performances with DM for non-dense breasts. All metrics, including specificity ($p = 0.08$), sensitivity ($p = 0.44$), location sensitivity ($p = 0.09$) and AUC ($p = 0.25$), were similar between radiologists who had read

mammography for < 10 years, 10–19 years and ≥ 20 years. Specificity ($p = 0.13$) was similar between radiologists who had completed the breast fellowship and those who had not; however, sensitivity ($p = 0.02$), location sensitivity ($p = 0.4$) and AUC ($p = 0.002$) were significantly higher among those who had completed the breast fellowship. Radiologists who work for the BreastScreen Australia program showed similar specificity ($p = 0.60$) and AUC ($p = 0.09$), but significantly higher sensitivity ($p = 0.002$) and location sensitivity ($p < 0.001$), compared to those who do not. Radiologists who read > 100 mammogram cases weekly showed significantly higher sensitivity ($p = 0.003$) and location sensitivity ($p < 0.001$), but not specificity ($p = 0.18$) or AUC ($p = 0.35$).

Table 2. Comparison of radiologists' performance characteristics in dense breasts.

Characteristic (readers no.)	Specificity (%)	p	Sensitivity (%)	p	Location sensitivity (%)	p	AUC (0–1)	p
No. of years reading mammography								
<10 years (63)	81.8 ± 13.2	0.26	77.15 ± 19.9	0.13	66.3 ± 20.5	0.14	0.794 ± 0.116	0.84
10–19 years (21)	75.9 ± 14.5		83.7 ± 11.3		71.2 ± 10.4		0.797 ± 0.079	
10–19 years (21)	79.6 ± 14.8		77.9 ± 15.5		64.5 ± 17.2		0.784 ± 0.107	
Breast fellowship (3–6 months)								
Yes (42)	87 (78.91)*	0.004	80.5 ± 16.8	0.36	72.4 ± 16.5	0.01	0.820 ± 0.103	0.03
No (86)	81 (66.90)*		77.5 ± 17.6		63.7 ± 18.2		0.777 ± 0.107	
BreastScreen Australia program reader								
Yes (62)	82.7 ± 13.1	0.04	84.6 (77.92)*	0.005	77 (69.83)*	< 0.001	0.828 ± 0.108	< 0.001
No (66)	77.6 ± 14.6		83.3 (62.85)*		67 (54.75)*		0.757 ± 0.093	
No. of cases read per week								
≤ 100 cases (50)	81 ± 15.2	0.51	85 (76.92)*	0.03	76 (67.83)*	0.001	0.817 ± 0.091	0.03
≤ 100 cases (78)	79.4 ± 13.3		83 (67.92)*		68 (54.68)*		0.775 ± 0.113	

(*) signifies median values, including 1st and 3rd quartiles, where significant values resulted from the Mann–Whitney U test. Bold values indicate statistical significance at the p -value ≤ 0.05 level.

Table 3. Comparison of radiologists' performance characteristics for non-dense breasts.

Characteristics (readers no.)	Specificity (%)	p	Sensitivity (%)	p	Location sensitivity (%)	p	AUC (0–1)	p
No. of years reading mammography								
< 10 years (63)	80 ± 15.43	0.08	78.7 ± 17.5	0.44	65.2 ± 22	0.09	0.797 ± 0.104	0.25
10–19 years (21)	78.44 ± 15.63		83.6 ± 15.2		75.25 ± 16.7		0.809 ± 0.105	
≥ 20 years (44)	73 ± 15.9		81.5 ± 16.6		70.9 ± 16.8		0.769 ± 0.102	
Breast fellowship (3–6 months)								
Yes (42)	80.3 ± 15.6	0.13	85 ± 16.3	0.02	74 ± 18.3	0.04	0.829 ± 0.104	0.002
No (86)	75.8 ± 15.8		77.9 ± 16.5		66.5 ± 20.2		0.769 ± 0.098	
BreastScreen Australia program reader								
Yes (62)	76.6 ± 17.4	0.60	85 ± 14.3	0.002	71.4 (71.86)*	< 0.001	0.805 ± 0.105	0.09
No (66)	78 ± 14.1		75.8 ± 17.7		62.5 (43.83)*		0.774 ± 0.101	
No. of cases read per week								
>100 cases (50)	75 ± 18	0.18	85.7 ± 14.5	0.003	73 (71.86)*	< 0.001	0.800 ± 0.107	0.35
≤100 cases (78)	88 ± 14.1		76.8 ± 17.2		71 (43.83)*		0.782 ± 0.102	

(*) signifies median values, including 1st and 3rd quartiles, where significant values resulted from the Mann–Whitney U test. Bold values indicate statistical significance at the p -value ≤ 0.05 level.

Association between radiologists' characteristics and diagnostic confidence when reporting breast cancer across breast densities

Diagnostic confidence among radiologists who had read mammography for < 10 years, 10–19 years and \geq 20 years did not differ significantly in dense ($p = 0.42$) or non-dense ($p = 0.99$) breasts. Radiologists who had completed a breast fellowship showed no significant increase in diagnostic confidence compared to those who had not in either dense ($p = 0.06$) or non-dense ($p = 0.08$) breasts. BreastScreen program readers showed significantly lower confidence levels (i.e. used more scores of 3 (indeterminate/equivocal)) than radiologists who do not read for the program (dense: $p = 0.008$; non-dense $p < 0.001$). Radiologists who read a higher volume (> 100) of mammogram cases weekly also had significantly lower confidence levels than those who read ≤ 100 mammogram cases per week when reporting cancer in dense ($p = 0.004$) and non-dense ($p < 0.001$) breasts, as shown in [Table 4](#).

Association between radiologists' characteristics and their ability to classify lesions into types across breast densities

When classifying breast lesions, the performance of radiologists who had read mammography for < 10 years, 10–19 years and \geq 20 years did not differ significantly in dense ($p = 0.56$) or non-dense ($p = 0.96$) breasts. Completion of the breast fellowship did not impact radiologists' lesion classification performances in dense ($p = 0.16$) or non-dense ($p = 0.44$) cases. However, in dense cases ($p = 0.03$), BreastScreen program readers performed significantly better than those who did not read for the program; this was not true in non-dense cases ($p = 0.12$). Also, radiologists who interpreted > 100 mammogram cases weekly more accurately classified breast lesions in dense breasts ($p = 0.03$) but not in non-dense breasts ($p = 0.17$), as shown in [Table 5](#).

Discussion

This observational study was conducted to assess radiologists' performances and markers of good performance across different breast compositions. The evidence suggests a difference in radiologists' performance characteristics across dense and non-dense breasts, indicating that participation in screening programmes, completing breast fellowship training and reading > 100 mammograms weekly are important diagnostic performance indicators, but not the number of years reading mammography.

The literature shows wide variation in the search, perception and decision-making abilities of radiologists that are concomitant with differences in performance in the

interpretation of mammographic images, suggesting that human limitations significantly impact the efficacy of screening mammography. Differences in reader ability and interaction with radiological images cannot be completely mitigated, but should be exploited to improve diagnostic efficacy.

We thought that in a simulated clinical environment, readers may exhibit a trade-off between sensitivity and specificity in dealing with suspicious cases.²¹ However, BreastScreen readers and those with breast fellowship training were not influenced by this trade-off, as demonstrated by the significantly higher specificity compared to other characteristics, particularly in dense breasts. Whilst the impact of fellowship training and volume read on performance is consistent with most published literature,^{11,17,22,23} it is unclear whether the significantly higher performance observed in BreastScreen readers is due to feedback from the service. It is possible that these three factors are interconnected and work together to improve performance.

The lack of association between years of reading mammograms and performance is reasonable and consistent with data from both data linkage and observer performance studies.^{12,13,17,24} Our findings suggest that years reading mammograms may not necessarily capture experience because factors such as mentorship, participation in diagnostic workup, feedback and interactions with images of different disease presentations may influence how radiologists build image interpretation skills.

Diagnostic confidence is also an important factor of radiologists' performance, as it is associated with greater accuracy in detecting breast cancer and influences decision-making regarding recall for further assessment or biopsy.^{22,25} However, diagnostic confidence is complex, involving visual perceptions and clinical judgements that depend on other factors, such as image quality and the interpreters' capabilities.⁴ Interestingly, BreastScreen Australia and high-volume readers demonstrated significantly lower diagnostic confidence, using a score of 3 when reporting breast cancer across dense and non-dense breasts. This is consistent with the results of a previous observational study that did not consider mammographic density.²² This could be due to a cohort of more risk-averse readers who do not want to over-emphasise the significance of the lesion or who assume that mammograms alone are unreliable and require additional information from supplemental imaging and biopsy for confirmation. In the Tabar/RANZCR scoring scheme, a score of 3 indicates that the lesion requires further investigation, usually through percutaneous needle biopsy.²⁶ Diagnostic confidence based on the assessment categories is very subjective; the decision to recall or not is the key parameter, as any score of 3 or above will result in a recall.

Table 4. Association between radiologists' characteristics and diagnostic confidence when reporting breast cancer.

Reader characteristics (readers no.)	Confidence score (dense cases)			p	Confidence score (non-dense cases)			p
	4 or 5	3			4 or 5	3		
No. of years reading mammography								
< 10 years (63)	36.4% (289/794)	63.6% (505/794)		0.42	34.12% (159/466)	65.88% (307/466)		0.99
10–19 years (21)	35% (92/266)	65% (174/266)			35.7% (55/154)	64.3% (99/154)		
≥ 20 years (44)	40% (225/563)	60% (338/563)			34.9% (111/318)	65.1% (207/318)		
Breast fellowship (3–6 months)								
Yes (42)	44% (187/425)	56% (241/428)		0.06	38.5% (101/262)	61.5% (161/262)		0.08
No (86)	49.5% (418/845)	50.5% (427/845)			45.4% (224/493)	54.6% (269/493)		
BreastScreen Australia program reader								
Yes (62)	44% (292/664)	56% (372/664)		0.008	36% (137/383)	64.2% (246/383)		< 0.001
No (66)	51.6% (314/609)	48.4% (295/609)			50.5% (188/372)	49.5% (184/372)		
No. of cases read per week								
> 100 cases (50)	42.8% (229/535)	57.2% (306/535)		0.004	34.4% (107/311)	65.6% (204/311)		< 0.001
≤ 100 cases (78)	51.1% (377/738)	48.9% (361/738)			49.3% (219/444)	50.7% (225/444)		

Score 3 (indeterminate/equivocal); score 4 (suspicious); score 5 (highly suspicious). Bold values indicate statistical significance at the p-value ≤ 0.05 level.

Table 5. Association between radiologists' characteristics and lesion classification.

Reader characteristics (readers no.)	Lesion classification into types (dense cases)		p	Lesion classification into types (non-dense cases)		p
	True	False		True	False	
No. of years reading mammography						
< 10 years (63)	30% (238/794)	70% (556/794)	0.56	41.2% (192/466)	58.8% (274/466)	0.96
10–19 years (21)	33% (88/266)	67% (178/266)		43.5% (67/154)	56.5% (87/154)	
≥ 20 years (44)	33% (185/563)	67% (378/563)		35% (111/318)	65% (207/318)	
Breast fellowship (3–6 months)						
Yes (42)	43% (184/428)	57% (244/428)	0.16	54.6% (143/262)	45.4% (119/262)	0.44
No (86)	38.7% (327/845)	61.3% (518/845)		51.3% (253/493)	48.7% (240/493)	
BreastScreen Australia program reader						
Yes (62)	43% (286/664)	57% (378/664)	0.03	55.4% (212/383)	44.6% (171/383)	0.12
No (66)	37% (225/609)	63% (384/609)		49.5% (184/372)	50.5% (188/372)	
No. of cases read per week						
> 100 cases (50)	44% (234/535)	56% (301/535)	0.03	55.6% (173/311)	44.4% (138/311)	0.17
≤ 100 cases (78)	37.5% (277/738)	62.5% (461/738)		50.2% (223/444)	49.8% (221/444)	

Lesion types: Stellate, discrete mass, spiculated mass, nonspecific density, calcification and architectural distortion. Bold values indicate statistical significance at the p-value ≤ 0.05 level.

These findings have a few implications for policy and practice. First, radiologists' characteristics associated with performance in dense and non-dense breasts can be used to optimise pairing strategies in countries such as Australia, where independent double reading of mammograms is practiced. This strategy may increase the chance that if a lesion is missed by one radiologist due to breast composition, it will be detected by another. Second, these findings can be used to identify radiologists who may benefit from tailored training in identifying breast cancer in different density breasts and inform educational interventions to improve their performance. Third, radiological lesion classification into types has been associated with specific histological findings,²⁷ and variation in the classification of lesions could affect further assessment and patient management. Therefore, familiarity with lesion features may eliminate some intrinsic errors associated with radiologists' diagnostic performance and decisions,⁴ and test-set data may be useful for training to improve radiologists' ability to detect and classify malignant features on mammograms. These findings suggest the need for observational studies exploring the impact of knowledge of mammographic lesion features on breast cancer detection and confidence levels across dense and non-dense breasts.

This study is not without limitations. First, the number of dense cases was comparable to that of non-dense cases, and such weighting may not reflect real screening populations. However, the number of cases needed to be similarly distributed to avoid selection bias. This is supported by a previous study that found readers show significantly higher sensitivity rates in dense breasts when fewer dense breast cases are included in the dataset.²⁸ Second, these findings may not completely reflect the performance of this cohort of radiologists in actual screening practice because interval cancers were not considered. However, a previous study²⁹ comparing the performance of the same cohort of Australian radiologists in both clinical and test settings showed no difference in performance, suggesting that test set data can reasonably predict performance in a clinical setting. Clinical audits to assess radiologists' screening performance require several years of follow-up to establish true interval cancers and negative mammograms, and the results of these audits are provided to the clinical practice rather than the individual radiologist. Therefore, test set data may provide opportunities to establish reader characteristics associated with performance across breasts of different compositions and feedback for individual radiologists. Third, it is possible that not all testing sites conformed to the RANZCR and BreastScreen Australia Accreditation Standards ambient lighting standards, although such differences should have a negligible impact on the findings.³⁰ To our knowledge, no study has closely examined the characteristics of BreastScreen Australia

readers associated with improved performance in different breast compositions. Therefore, our study provides baseline data to optimise diagnostic efficacy and confidence in dense breasts.

In conclusion, participating in a screening program reading, fellowship in breast imaging and weekly volumes read of greater than 100 mammogram cases are the most important indicators of diagnostic performance across dense and non-dense breasts. In dense breasts, optimal performance was demonstrated by screening program readers. These findings have practical implications for helping breast screening programs achieve better outcomes in dense and non-dense breasts.

Acknowledgement

The authors would like to thank the Breast Reader Assessment Strategy (BREAST) for supporting the collection of the data used in this paper, and the Australian Department of Health, and Cancer Institute New South Wales for funding the BREAST.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Ibrahim Hadadi  <https://orcid.org/0000-0002-2517-7465>
Mark McEntee  <https://orcid.org/0000-0001-7045-2190>

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136: E359–E386.
2. Australian Institute of Health Welfare & Cancer Australia. Breast Cancer in Australia: An Overview. Canberra, 2012. <https://www.aihw.gov.au/reports/cancer/breast-cancer-in-australia-an-overview>.
3. Australian Institute of Health Welfare & Cancer Australia. BreastScreen Australia Monitoring Report 2009–2010. Canberra, 2012. <https://www.aihw.gov.au/reports/cancer-screening/breastscreen-australia-monitoring-2009-2010>.
4. Ekpo EU, Alakhras M, Brennan P. Errors in mammography cannot be solved through technology alone. *Asian Pac J Cancer Prev* 2018;19:291–301. DOI: [10.22034/apjcp.2018.19.2.291](https://doi.org/10.22034/apjcp.2018.19.2.291).
5. Sprague BL, Gangnon RE, Burt V, et al. Prevalence of mammographically dense breasts in the United States. *J Natl Cancer Inst* 2014; 106. DOI: [10.1093/jnci/dju255](https://doi.org/10.1093/jnci/dju255).

6. Dai H, Yan Y, Wang P, et al. Distribution of mammographic density and its influential factors among Chinese women. *Int J Epidemiol* 2014;43:1240–1251. DOI: [10.1093/ije/dyu042](https://doi.org/10.1093/ije/dyu042).
7. Jo H-M, Lee EH, Ko K, et al. Prevalence of women with dense breasts in Korea: results from a nationwide cross-sectional Study. *Cancer Res Treat* 2019;51:1295–1301. DOI: [10.4143/crt.2018.297](https://doi.org/10.4143/crt.2018.297).
8. Breast Cancer Network Australia. Mammographic Density and Screening, 2018. <https://www.bcna.org.au/breast-health-awareness/mammographic-density-and-screening/>.
9. Oiwa M, Endo T, Suda N, et al. Can quantitative evaluation of mammographic breast density, “volumetric measurement”, predict the masking risk with dense breast tissue? Investigation by comparison with subjective visual estimation by Japanese radiologists. *Breast Cancer* 2019;26:349–358.
10. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006;15:1159–1169. DOI: [10.1158/1055-9965.EPI-06-0034](https://doi.org/10.1158/1055-9965.EPI-06-0034).
11. Suleiman WI, Lewis SJ, Georgian-Smith D, et al. Number of mammography cases read per year is a strong predictor of sensitivity. *J Med Imaging* 2014;1:015503. DOI: [10.1117/1.JMI.1.1.015503](https://doi.org/10.1117/1.JMI.1.1.015503).
12. Elmore JG, Wells CK, Howard DH. Does diagnostic accuracy in mammography depend on radiologists’ experience? *J Women’s Health* 1998;7:443–449.
13. Rawashdeh MA, Lee WB, Bourne RM, et al. Markers of good performance in mammography depend on number of annual readings. *Radiology* 2013;269:61–67.
14. Hoff SR, Myklebust T-Å, Lee CI, et al. Influence of mammography volume on radiologists’ performance: results from BreastScreen Norway. *Radiology* 2019;292:289–296.
15. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists’ characteristics associated with accuracy. *Radiology* 2009;253:641–651.
16. Reed WM, Lee WB, Cawson JN, et al. Malignancy detection in digital mammograms. *Acad Radiol* 2010;17:1409–1413.
17. Molins E, Macià F, Ferrer F, et al. Association between radiologists’ experience and accuracy in interpreting screening mammograms. *BMC Health Serv Res* 2008;8:91.
18. Radiology ACo. American College of Radiology Breast Imaging Reporting and Data System (BI-RADS). 4th edition. American College of Radiology, 2003.
19. BreastScreen Australia Accreditation Review Committee. BreastScreen Australia National Accreditation Standards. Breast Imaging: A Guide for Practice BreastScreen. Australia, 2015.
20. Heggie JCP, Barnes P, Cartwright L, et al. Position paper: recommendations for a digital mammography quality assurance program V4.0. *Australas Phys Eng Sci Med* 2017;40:491–543.
21. Gur D, Bandos AI, Cohen CS, et al. The “laboratory” effect: comparing radiologists’ performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology* 2008;249:47–53.
22. Trieu PD, Lewis SJ, Li T, et al. Reader characteristics and mammogram features associated with breast imaging reporting scores. *Br J Radiol* 2020;93:20200363.
23. Miglioretti DL, Gard CC, Carney PA, et al. When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology* 2009;253:632–640.
24. Barlow WE, Chi C, Carney PA, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst* 2004;96:1840–1850. DOI: [10.1093/jnci/djh333](https://doi.org/10.1093/jnci/djh333).
25. Geller BM, Bogart A, Carney PA, et al. Is confidence of mammographic assessment a good predictor of accuracy? *Am J Roentgenol* 2012;199:W134–W141.
26. Breast Imaging: A Guide for Practice. Camperdown, NSW: National Breast Cancer Centre, 2002. <https://www.cancer.gov.au/publications-and-resources/cancer-australia-publications/breast-imaging-guide-practice>.
27. Thurffjell MG, Lindgren A, Thurffjell E. Nonpalpable breast cancer: mammographic appearance as predictor of histologic type. *Radiology* 2002;222:165–170.
28. Tapia KA, Rickard MT, McEntee MF, et al. Impact of breast density on cancer detection: observations from digital mammography test sets. *Int J Radiol Radiat Ther* 2020;7:36–41. DOI: [10.15406/ijrrt.2020.07.00261](https://doi.org/10.15406/ijrrt.2020.07.00261).
29. Soh BP, Lee W, McEntee MF, et al. Screening mammography: test set data can reasonably describe actual clinical reporting. *Radiology* 2013;268: 4–53.
30. Pollard BJ, Samei E, Chawla AS, et al. The influence of increased ambient lighting on mass detection in mammograms. *Acad Radiol* 2009;16:299–304. DOI: [10.1016/j.acra.2008.08.017](https://doi.org/10.1016/j.acra.2008.08.017).