Clinical Radiology 79 (2024) e1549–e1556

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Access to prior screening mammograms affects the specificity but not sensitivity of radiologists' performance

J.D. Akwo [a,*], P. D. (Yun) Trieu [a], M.L. Barron [a], T. Reynolds [a], S.J. Lewis [a,b]

[a] *Medical Image Optimisation and Perception Group, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia*
[b] *School of Health Sciences, Western Sydney University, Campbelltown, Australia*

ARTICLE INFORMATION

Article history:
Received 25 July 2024
Received in revised form
5 September 2024
Accepted 12 September 2024

AIMS: To establish the impact that access to prior mammograms has on radiologists' performance and the influence of radiologists' characteristics and breast density on their subsequent performance.

METHODS: Eight participants independently interpreted 72 digital screening mammograms in two reading sessions using the Royal Australian and New Zealand College of Radiologist's classification. In the first reading session, participants were given access to current and prior mammograms. In the second reading session six months later, participants only had access to the current mammograms. Radiologists' specificity, sensitivity, lesion sensitivity, Receiver Operating Characteristic (ROC) curve, and Jacknife Alternative Free-response ROC (JAFROC) were calculated. A Paired T-test was used to compare readings with and without prior mammograms, and to assess if breast density influenced participants performance. Independent Sample T-test was used to compare performance across radiologists' characteristics. A relative risk analysis was conducted to assess the probability of false positives and false negatives when prior mammograms were available.

RESULTS: Access to prior mammograms improved specificity in dense and non-dense breasts ($p \leq 0.01$) and reduced false positives ($p = 0.01$) but had no effect on sensitivity ($p = 0.37$), lesion sensitivity ($p = 0.67$), ROC ($p = 0.16$), and JAFROC ($p = 0.24$). Prior mammogram also reduced the probability of false positives (RR = 0.38; 95%CI:0.26–0.57, $p < 0.0001$) without affecting the false negative rate (RR = 1.14; 95%CI:0.88–1.49, $p = 0.30$). The impact of prior mammograms on performance was not influenced by breast density or radiologists' characteristics.

CONCLUSIONS: Access to prior mammograms improves radiologists' specificity and reduces false positives without affecting sensitivity and the false negative rate regardless of radiologists' characteristics and breast density.

* Guarantor and correspondent: J. D. Akwo, Medical Image Optimisation and Perception Group, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia.
E-mail address: jakw2899@uni.sydney.edu.au (J.D. Akwo).

https://doi.org/10.1016/j.crad.2024.09.007
0009-9260/

Background

Female breast cancer is the second most common cancer and leading cause of deaths due to cancer in women worldwide.[1,2] Early detection is key to reducing deaths from the disease.[3] To facilitate early detection, Australia established a population-based screening program for women aged 50–74 years in 1991, called BreastScreen Australia (BSA).[4] Mammography is the standard imaging tool for breast cancer screening in Australia, and the images produced are interpreted by trained readers, normally radiologists. However, the interpretation of mammographic images is challenging.[5,6] About 20–30% cancers are missed in mammograms and radiologists demonstrate wide variability in their ability to detect cancer.[5]

The mammographic diagnosis of breast cancer is made based on the identification of features of abnormality on a woman's mammogram. Sometimes, these features are subtle and may be missed or not judged to be cancer by radiologists interpreting the mammograms.[5] Breast cancer contains rapidly proliferating cells, which causes the cancer lesions to grow over time and sometimes change the architecture of the breast tissues.[7,8] Therefore, previous mammograms may serve as a baseline to identify features or changes in current mammograms suggestive of breast cancer, which could help improve diagnostic decision-making[9,10] With the digitisation of the mammography screening process, it has become easier to retrieve and compare previous and current images of women undergoing screening for abnormal beast changes. However, screening programs are also challenged by storage capacity due to the big data generated from current and previous screening rounds in the digital era. In cases where previous mammograms are available, radiologists usually spend additional time reviewing and comparing these previous mammograms to the current screening exam, which could increase their workload, both operationally and visually.

Radiologists demonstrate variability in the interpretation of screening mammograms, and experience and case load are both documented to be predictors of observer performance.[11] It is possible that some radiologists more than others may benefit from looking at both prior and current mammograms together when interpreting mammograms, but this is poorly understood. Another factor affecting the interpretation of screening mammograms is the composition of a woman's breast, with the sensitivity of mammography documented to be lower in women with extremely dense breasts compared to those with fatty breast.[12,13] Women with dense breasts are more frequently recalled at screening and often receive false positive outcomes due to summation artefacts in their mammograms.[14,15]

Screening programs aim to keep recall rates low and is generally one of the major criteria for judging the performance of the program. It is possible that comparing current and previous mammograms, particularly of women with dense breasts may also help screening programs to reduce recall rates. Therefore, identifying the characteristics of radiologists and women who benefit from prior mammograms being reviewed with current mammograms could lead to informed strategies to improve screening outcomes. This study aims to establish the impact that access to prior mammograms has on accurate interpretation of screening mammograms and to establish the characteristics of radiologists and breast compositions that benefit from prior mammograms.

Methodology

This study received institutional review board approval to conduct an observer performance study involving breast radiologists and other designated screen readers for BSA interpreting different mammographic cases at two different time points: first with access to prior mammograms and second, without access to prior mammograms (Human Research Ethics Committee Number 2023/101).

Participant recruitment

A simple convenience sampling strategy was used to recruit readers at the Breast Imaging Group (BIG) conference and the Royal Australian and New Zealand College of Radiologists (RANZCR) annual scientific meeting in 2023. Eleven radiologists with varying levels of experience and workload characteristics were initially recruited for the study and 8 participants completed both reading sessions.

Mammogram test set

A total of 72 full-field digital mammograms were sampled from the Breastscreen REeader Assessment Strategy STrategy (BREAST) databank of cases, ensuring even distribution of women of breast densities (A-B-C-D), breast cancer types and characteristics (calcifications, discrete mass, architectural distortion, and non-specific density), and ages. Briefly, BREAST was established to provide breast screening readers the opportunity to assess their performance in mammography interpretation and receive feedback on false negative and false positive errors like PERFORMS (Personal Performance in Mammographic Screening) in the United Kingdom National Health Service Breast Screening Programme (NHSBSP).[16,17] BREAST is used to support continuous professional development and serves as a national quality training tool in BSA as PERFORMS is to the NHSBSP.[16,18] Therefore, BREAST mammography test-set cases are derived from the imaging libraries of BSA like the selection of mammograms for PERFORMS are derived from the NHSBSP.[16] The presence of cancer was confirmed by at least two expert radiologists and a positive biopsy result. The normal mammograms with no cancer were confirmed to be cancer-free using negative mammograms acquired 2–4 years later. All cases had a prior mammogram available through the BREAST databank, with the interval between the current and prior mammograms ranging from 2 to 4 years to ensure that slow growing cancers developed enough to be detected. The 72 cases contained prior and current mammograms from different vendors (General Electric, Sectra, Hologic, Fujifilm and Siemens HealthCare).

Experimental design

The cases in the test sets were uploaded to the BREAST platform for viewing and interaction. The eight participants independently interpreted the digital screening mammograms in two reading sessions using the Royal Australian and New Zealand College of Radiologists' (RANZCR) classification: 1 (normal); 2 (benign); 3 (indeterminate); 4 (suspicious); 5 (highly suspicious). If the reader detected a lesion, they then used a mouse to click on the lesion and assign a rating. If a rating of 3 or higher was assigned, the reader was asked to state the mammographic feature of the lesion detected (Fig 1). A rating of 2 denoted that reader considered the detected lesion as benign. The first reading session involved mammograms acquired in the current screening round and their priors (mammograms acquired 2–4 years prior to the current screening round). The second reading session involved the interpretation of the same 72 cases, but the prior mammograms were hidden so that the readers only had access to the current mammograms of these women. The two reading sessions were conducted at least six months apart to reduce memory effect. All cases were read on 5MP monitors and ambient lighting was kept between 15–20 lux to conform with the RANZCR standard. Prior to the assessment, each participant completed a short survey, which enabled us to collect demographic data and practice-related characteristics such as age, gender, years of experience reading mammograms, number of mammograms read per week, whether they are working in screening services. No feedback was given to the readers about their performance after the first reading session.

Statistical analysis

At the end of each the reading session, the BREAST platform automatically calculates the performance of each participant using metrics such as sensitivity, specificity, false positive rates, receiver operating characteristic (ROC) Area Under the curve (AUC), Jack-knife alternative free-response receiver operating characteristic (JAFROC), and lesion sensitivity. We collected and compared the diagnostic performance metrics for the participants in the two reading scenarios (with versus without access to prior mammograms).

Shapiro-wilk test demonstrated that data were normally distributed ($P>0.05$ for all performance metrics). Therefore, a Paired T-Test, which compares the means of two assessments from the same participants was used to compare performance with and without prior mammograms. Independent sample T-test, which compares the means of two independent groups was used to compare the performance of groups of readers with different characteristics. The thresholds used for the grouping were based on the characteristics of the participants who completed the two readings and ensured that participants were evenly distributed between groups. To assess the impact of prior mammograms across breast densities, we grouped data into dense (BI-RADS C – D) and non-dense breasts (BI-RADS A – B) and compared participants' performance with and without prior mammograms in dense and non-dense breasts. A relative risk analysis (RR) analysis was performed to assess the probability of false positive and false negative outcomes when prior mammograms were available. For the RR analysis, sample size power calculation showed that 16 true positive and 16 true negative cases, and a cohort of five participants each reading with and without prior mammograms were sufficient to yield 80% power at

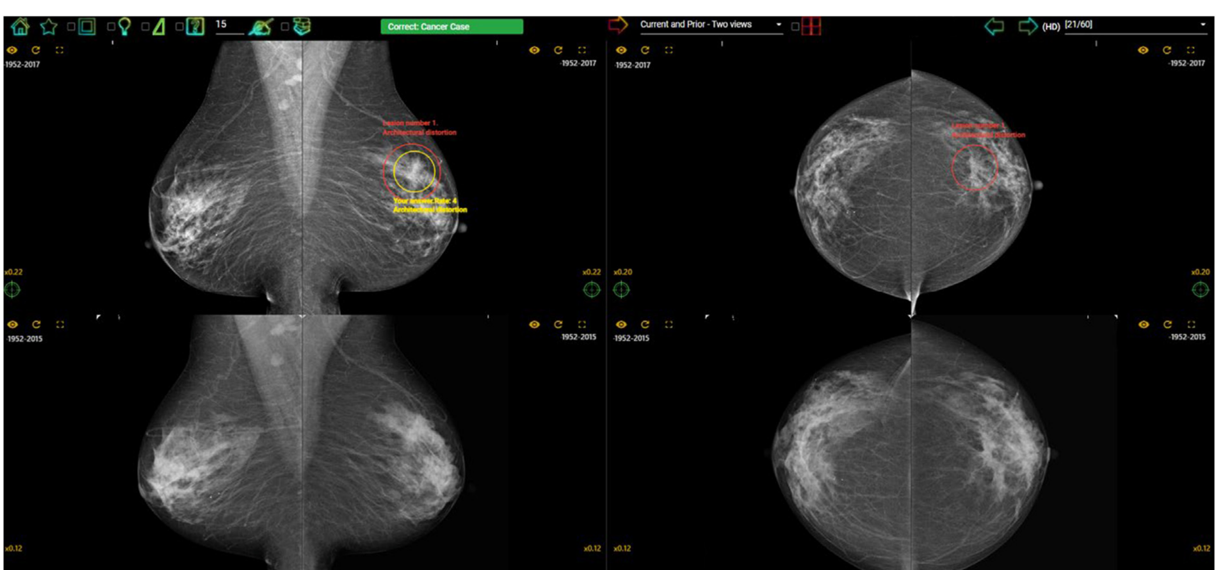

Figure 1 BREAST platform showing readers lesion marks. Yellow circle represents the reader's mark denoting the location of cancer and red circle represents the true location of the cancer.

Table 1
Participants' performance with compared to without prior mammograms.

Performance metric	With priors	Without priors	*p*-value
Specificity	91.1 (80–100)	76.3 (68–85)	0.009*
Sensitivity	67.5 (44–91)	71.3 (53–88)	0.37
Lesion sensitivity	64.1 (38–91)	66.9 (41–88)	0.67
ROC	80.7 (69.5–91.8)	77.6 (69.4–89.6)	0.16
JAFROC	78.1 (66.4–91.8)	74.7 (63.3–89.6)	0.24

ROC: Receiver Operating Characteristic curve; JAFROC: Jacknife Alternative Free-Response Receiver Operating Characteristic curve. Values in bracket represent the minimum and maximum scores.

95% confidence interval and <5% margin of error.[19] P-values ≤ 0.05 were considered statistically significant.

Results

A total of eight participants (six radiologists, one senior radiology trainee, and one breast physician) completed the two reading sessions. The years in their current roles varied from four to 35 years (mean = 20.3), with years of reading mammograms ranging from three to 35 years (Mean = 17.3). The number of hours spent reading mammograms per week varied from four to 30 hours, and the numbers of mammograms cases read per week ranged from 20 to 200.

Table 1 summarises the performance outcomes of the eight participants 'with prior mammograms' compared to 'without prior mammograms'. Only specificity was significantly higher when radiologists had access to prior mammograms compared to when prior mammograms were not available. True negatives were significantly higher with prior availability compared to without prior mammograms (mean = 36.4 vs. 30.5; p=0.01) and false positives were significantly lower with priors compared to without priors (mean = 3.6 vs 9.5; p = 0.01).

The probability of a false positive result was significantly lower when participants had access to prior mammograms (RR= 0.38; 95%CI: 0.26–0.57, p<0.0001). The probability of a false negative screening result was not significantly different between readings with and without prior mammograms (RR = 1.14; 95%CI: 0.88–1.49, p = 0.30).

When the analyses were stratified by breast density, participants' performance was not significantly different between dense and non-dense breast cases (p>0.05 for specificity, sensitivity, and lesion sensitivity). However, specificity was significantly higher when radiologists had access to prior mammograms in dense (p = 0.01) and non-dense (p = 0.01) compared to when prior mammograms were not available (Table 2). No differences were detected in sensitivity and lesion sensitivity when readings with priors were compared to those without prior mammograms.

Independent sample T-test showed that participants with >20 years of specialisation as breast radiologists or as breast physicians demonstrated significantly higher specificity (p=0.04) and true negatives (p=0.04), and lower false positives (p=0.04) when prior mammograms were available compared to those with ≤ 20 years of specialisation. No differences were observed between participants with different characteristics in other performance metrics (Table 3). No significant differences were observed between these groups of participants when prior mammograms were not available. There were also no differences between participants who spent ≤4 hours per week and > 4 hours per week reading mammograms in all performance metrics regardless of whether prior mammograms were available.

When prior mammograms were available, no significant differences were found between participants who have read mammograms for ≤ 10 years and those who had more than 10 years of mammography reading experience. Conversely, when prior mammograms were not available, participants with more than 10 years of mammography reading experience demonstrated better sensitivity, lesion sensitivity, ROC, JAFROC, and true positives (p≤0.05 for all). These group of participants also demonstrated fewer false negatives (p = 0.05). We did not find any significant differences between participants who read fewer cases (≤150) per week and those who reader more than 150 cases per week.

When the readings with and without prior mammograms were compared for participants employment characteristics, specificity emerged as the only performance metric that discriminated readings with versus without prior mammograms (Table 4). Specificity was significantly higher for participants with greater than 20 years of experience as radiologist or breast physician (p = 0.008), those who spent >4 hours/week (p = 0.05) or ≤4 hours/week (p = 0.02), have >10 years (p = 0.05) or ≤10 years (p = 0.01) of mammography reading experience, and those who read more than 150 mammograms per week (p = 0.04) when prior mammograms were available compared with when prior mammograms were not available.

Discussion

The findings demonstrate that access to mammograms from the previous screening round improves radiologists' ability to correctly identify women who have no signs of cancer on their images without affecting the detection and classification breast cancer lesions in women who have

Table 2
Comparison of participants' performance in dense and non-dense breasts with compared to without prior mammograms.

Performance	Dense breasts (BI-RADS C – D)			Non-dense breasts (BI-RADS A – B)		
	With priors	Without priors	*p*-value	With priors	Without priors	*p*-value
Specificity	92.5 (80–100)	76.9 (60–95)	0.01*	92.5 (80–100)	75.6 (60–90)	0.01*
Sensitivity	66.4 (50–93.8)	68.8 (50–87.5)	0.88	68.7 (37.5–93.8)	73.4 (56.3–87.5)	0.37
L. Sensitivity	62.5 (43.8–87.5)	64.8 (37.5–87.5)	0.66	65.6 (31.3–93.8)	68.8 (43.8–87.5)	0.62

L.Sensitivity: lesion sensitivity. Values in bracket represent the minimum and maximum scores.

Table 3
Comparison of the performance of participants with different characteristics across readings with and without prior mammograms.

Performance metric	With priors			Without priors		
	Years qualified as radiologist or breast physician/specialty					
	≤20 years (n=4)	>20 years (n=4)	*p*-value	≤20 years (n=4)	>20 years (n=4)	*p*-value
Specificity	86.5 (80–90)	95.8 (88–100)	0.04*	75.6 (68–85)	76.8 (68–85)	0.87
Sensitivity	71.8 (62–78)	63.3 (44–91)	0.45	68.6 (53–78)	73.6 (66–88)	0.54
L. Sensitivity	67.3 (56–72)	61 (38–91)	0.61	61.8 (41–75)	72 (62–88)	0.35
ROC	81.4 (75.5–79)	80.2 (69.5–91.8)	0.82	76.8 (69.4–82.8)	78.5 (70.3–89.6)	0.76
JAROC	77 (72.5–79)	79 (66.4–91.8)	0.75	71.2 (63.3–82.8)	77.6 (68.3–89.6)	0.39
	Hours spent reading mammograms per week					
	≤4 hours/week (n=4)	>4 hours/week (n=4)	p	≤4 hours/week (n=4)	>4 hours/week (n=4)	*p*-value
Specificity	91.5 (88–100)	90.6 (80–100)	0.89	75.8 (68–85)	76.8 (68–85)	0.87
Sensitivity	68.5 (62–78)	66.5 (44–91)	0.86	68.8 (53–78)	73.8 (66–88)	0.53
L. Sensitivity	64 (59–72)	64.3 (38–91)	0.98	61 (41–75)	72.8 (62–88)	0.27
ROC	81.5 (75.5–85.6	80 (65.5–91.8)	0.78	76.3 (69.4–80.9)	78.9 (70.3–89.6)	0.63
JAROC	77.3 (72.5–81.8)	78.9 (66.4–91.8	0.78	70.9 (63.3–79.3)	78.5 (68.3–89.6)	0.26
	Number of years reading mammograms					
	≤10 years (n=4)	>10 years (n=4)	p	≤10 years (n=4)	>10 years (n=4)	*p*-value
Specificity	90.3 (88–95)	91.6 (80–100)	0.82	74 (68–82)	77.6 (68–85)	0.57
Sensitivity	61.3 (44–91)	71.2 (59–72)	0.39	62.7 (53–88)	76.4 (66–78)	0.05*
L. Sensitivity	55.3 (38–91)	69.4 (56–72)	0.25	54.3 (41–88)	74.4 (62–75)	0.04*
ROC	76.7 (69.5–91.8)	83.1 (79.1–85.6)	0.22	72.2 (69.4–89.6)	80.9 (70.3–82.8)	0.04*
JAROC	73.6 (66.4–91.8)	80.7 (75–79.7)	0.20	67.6 (63.3–89.6)	78.9 (68.3–82.8)	0.05*
	Number of mammograms read per week					
	≤150 reads/week (n=3)	>150 reads/week; n=5	p	≤150 reads/week; n=3	>150 reads/week; n=5	p-value
Specificity	90.6 (88–100)	92 (95–100)	0.80	75 (68–85)	78.3 (68–85)	0.60
Sensitivity	68.2 (59–78)	66.3 (44–91)	0.87	74.6 (53–75)	65.7 (66–88)	0.26
L. Sensitivity	65.2 (56–72)	62.3 (38–91)	0.82	73.2 (41–72)	56.3 (62–88)	0.10
ROC	81.1 (75.5–85.2)	80.1 (69.5–91.8)	0.85	79.3 (69.4–80.9)	74.8 (70.3–89.6)	0.42
JAROC	78.1 (72.5–81.8)	78 (66.4–91.8)	0.99	77.7 (63.3–79.3)	69.6 (68.3–89.6)	0.24

L. Sensitivity: lesion sensitivity; ROC: receiver operating characteristics curve; JAFROC: Jacknife Alternative Free-response receiver operating characteristics curve; *significantly different. Values in bracket represent the minimum and maximum scores.

cancer. Data generated also show that the availability of prior mammograms for radiologists to view reduces the chances of false positive results for women undergoing screening. After adjustments were made for breast density and radiologists' characteristics such as years of specialisation, number of years reading mammograms, volume of mammograms read per week, and hours spent reading mammograms, the availability of prior mammograms consistently led to improvement in specificity and a reduction in false positive rate without affecting the false negative rate. It is important to note that while sensitivity and lesion sensitivity were lower with priors, they were not significantly different from the values obtained when readings occurred without priors. Also, despite the slightly lower sensitivity and lesion sensitivity with priors, ROC and JAFROC, which measures observers' abilities to detect the lesion and simultaneously rate the level of malignancy were higher when prior mammograms were available even though these were not significantly greater than chance.

These findings can be explained because breast cancer elicit perturbations in mammograms, which mimic normal breast parenchyma or benign changes.[7,20] However, unlike perturbations from normal breast parenchyma or benign lesions, the perturbations due to breast cancer gradually change overtime as the cancer grows or invades surrounding tissues.[7,8] Therefore, comparison with prior mammograms should help radiologists detect changes in mammographic features that are most discriminative of breast cancer. These findings are relevant because they suggest that viewing prior mammograms can improve the efficiency of screening programs by ensuring that women who have no cancer are correctly informed thereby reducing unnecessary recall and testing. However, due to population mobility across state lines, access to prior images is not always available to readers. The findings support screening guidelines and policies that recommend that mammograms of women undergoing screening should be retained for future reference.[21] Whilst archiving, cost, and retention issues are factors to consider when retaining prior mammograms,[22] the utility of prior mammograms in reducing false positives,[23,24] which mitigates psychosocial harms and cost of assessments for women wrongly recalled emphasise the importance of prior mammograms to screening programs.[25,26]

Our findings align with published evidence which shows that availability of prior mammograms to radiologists improves specificity and reduce false positives and recall rates without affecting cancer detection and sensitivity.[24,27–32] For example, experimental observer performance and retrospective studies have reported a 4–14.5%

Table 4
Comparison between readings with prior compared to without prior mammograms for participants of similar characteristics.

	Years qualified as radiologist or breast physician/specialty					
	With priors (>20 years; n=4)	Without priors (>20 years; n=4)	*p*-value	With priors (≤20 years; n=4)	Without priors (≤20 years; (n=4)	*p*-value
Specificity	95. (95–100)	76.8 (72–85)	0.008*	86.5 (80–90)	75.8 (68–85)	0.07
Sensitivity	63.3 (44–91)	73.8 (66–88)	0.38	71.8 (62–78)	68.8 (53–78)	0.66
L. Sensitivity	61 (38–91)	72 (62–88)	0.41	67.3 (56–72)	61.8 (41–75)	0.56
ROC	80 (70–92)	78.5 (73–90)	0.79	81.4 (76–86)	76.8 (69–83))	0.28
JAROC	79 (66–92)	77.6 (68–90)	0.84	77.1 (73–82)	71.7 (63–83)	0.11
	Hours spent reading mammograms per week					
	With priors (>4 hours/week; n=4)	Without priors (>4 hours/week; n=4)	*p*-value	With priors (≤4 hours/week; n=4)	Without priors (≤4 hours/week; n=4)	*p*-value
Specificity	90.8 (80–100)	76.8 (68–85)	0.05*	91.5 (88–100)	75.6 (68–85)	0.02*
Sensitivity	66.5 (44–91)	73. (66–88)8	0.54	68.5 (59–78)	68.8 (53–78)	0.97
L. Sensitivity	64.3 (38–91)	72.8 (62–88)	0.52	64 (56–72)	61 (41–75)	0.74
ROC	80 (70–92)	79 (70–90)	0.87	81.5 (76–86)	76.3 (69–81)	0.20
JAROC	78.9 (66–92)	78.5 (68–90)	0.95	77.3 (73–80)	70.9 (63–79)	0.18
	Number of years reading mammograms					
	With priors (>10 years; n=4)	Without priors (>10 years; n=4)	*p*-value	With priors (≤10 years; n=4)	Without priors (≤10 years; n=4)	*p*-value
Specificity	92.5 (80–100)	76.5 (68–85)	0.05*	89.8 (88–95)	76 (68–82)	0.01*
Sensitivity	66.3 (59–75)	73.5 (66–78)	0.19	68.8 (44–91)	69 (53–88)	0.98
L. Sensitivity	64 (56–72)	71 (62–75)	0.20	64.3 (38–91)	62.8 (56–88)	0.92
ROC	81 (79–86)	78.7 (70–83)	0.49	80.5 (70–92)	76.8 (69–90)	0.57
JAROC	78 (75–80)	76.2 (68–83)	0.61	78.1 (66–92)	73.1 (63–90)	0.55
	Number of mammograms read per week					
	With priors (>150 reads/week; n=5	Without priors (>150 reads/week; n=5	*p*-value	With priors ≤150 (reads/week; n=3	Without priors (≤150 reads/week; n=3	*p*-value
Specificity	91 (80–100)	73.7 (68–85)	0.04*	86.3 (88–100)	77.3 (82–85)	0.19
Sensitivity	68 (44–75)	73 (66–88)	0.40	68.2 (59–78)	70 (53–75)	0.89
L. Sensitivity	66.2 (38–91)	69.7 (62–88)	0.56	64.6 (56–72)	61.7 (41–72)	0.85
ROC	81.4 (70–92)	77.3 (70–90)	0.22	79 (76–85)	77.7 (69–81)	0.85
JAROC	78.9 (66–92)	74 (68–90)	0.13	77.1 (73–82)	73.1 (63–79)	0.64

L. Sensitivity: lesion sensitivity; ROC: receiver operating characteristics curve; JAFROC: Jacknife Alternative Free-response receiver operating characteristics curve; *significantly different. Values in bracket represent the minimum and maximum scores.

improvement in specificity, no significant change in sensitivity,[24,27–31] and 8–13.5% reduction in false positive rates[23,24] when prior mammograms were available. This current study found a 14.8% improvement in specificity and a 6% reduction in the false positives rate with prior mammograms. Relative risk analysis demonstrated a 62% reduction in the absolute probability of false positives when prior mammograms were available. It is worth noting that assessment of sensitivity, specificity, and false positive rates involve classification tasks, which requires observers to state whether mammograms contain cancer without indicating the location of cancer in the mammogram. The JAFROC and lesion sensitivity methodologies used in our study capture observers correct lesion detection and the level of malignancy. The only previous study that has used these methodologies compared the performance of radiologists who read different test sets, some with and others without prior mammograms,[30] which may not completely capture the impact of prior mammograms on performance. This study is the first to examine the impact of prior mammograms where the same cohort of participants read the same mammograms with and without prior images, analysed performance with observer performance methodologies at the level of the lesion, and adjusted for factors that could influence their performance. Therefore, our findings account for both the detection and malignancy rating of cancer lesions and provide stronger evidence for the impact of prior examinations in identifying the normal cases better than cancer cases via screening mammograms.

The literature shows that factors such as the characteristics of women screened and radiologists who interpret these mammograms influence observer performance in interpreting screening mammograms.[11,13] Studies based BREAST and PERFORMS have demonstrated the impact of these platforms in identifying the influence that technology, patient, and radiologists' characteristics have on performance.[17,18,33] For example, analysis of data of over 400 PERFORMS participants showed that individual and practice-related characteristics were the most discriminative of high and low performing screen readers.[33] To account for these confounding factors, we adjusted for breast density and radiologists' characteristics such as years of experience, years reading mammograms, number of mammograms read per week, and hour spent reading mammograms using the demographic and practice-related information self-reported by participants in the BREAST platform. Regardless of the adjustments that were made, access to prior mammograms consistently improved specificity without affecting sensitivity, lesion sensitivity, ROC, and JAFROC. When readings without prior mammograms were considered separately for each reader characteristic, only radiologists with greater than 10 years of experience were able to outperform their colleagues with fewer years of experience in terms of sensitivity, lesion sensitivity, ROC, and JAFROC (Table 3). These findings confirm that the improvement observed in specificity were not due to reader characteristics or the composition of the breast of women whose mammograms were used for the study as demonstrated by lack of difference in performance between dense and non-dense breasts. Rather, this study indicates that improved performance was due to the availability of prior mammograms as shown in Tables 2 and 4. These findings highlight the importance of prior mammograms to all radiologists regardless of their years of experience and workload-related characteristics, and women of all breast densities, and we consider this the first time such an effect has been shown.

A few limitations of the study must be acknowledged. The current study was based on an enriched mammography test-set with even distribution of breast densities. While this may not be representative of the clinical environment, the test-set design ensures that all cancer lesion types and breast densities were adequately represented, and accounts for the smaller number of images used in experimental test-set observer performance studies. It should also be noted that studies based on BREAST and PERFORMS have shown that test-set performance can predict screen readers' performance in a real-life clinical setting.[27,34,35] Therefore, our findings can be used to support the use of priors in screening programs and the establishment of a national reference mammography database for women participating in screen mammography. Secondly, only the assessments of eight participants who completed both test sets (readings with and without prior mammograms) were used in the analysis. However, this number is above the 75th percentile of the participant sample sizes used in similar previous studies.[24,27–32] Three participants did not respond to communication to complete the second test set and were therefore excluded from the study. The number of cases (n=72) could limit the generalisability of findings; however, it was important to develop an experimental protocol that could be completed in a reasonable time (2–3 hours per test set) so as not to fatigue the readers. Nonetheless, the sample of cases and readers in this study fall within the range recommended for phase II observer performance studies of medical imaging.[19,36]

Conclusion

Availability of prior mammograms to radiologists improves specificity and reduces false positive rates without reducing the detection and characterisation of breast cancer in mammograms. Therefore, the simultaneous display of prior and current mammograms may be a useful strategy to improve the efficiency of screening programs by reducing the number of women who are incorrectly called back to assessment clinics.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contribution

1. Guarantor of integrity of the entire study: Sarah J. Lewis.
2. Study concepts and design: Judith D. Akwo, Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynolds, Sarah J. Lewis.
3. Literature research: Judith D. Akwo.
4. Clinical studies: N/A.
5. Experimental studies/data analysis: Judith D. Akwo, Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynolds, Sarah J. Lewis.
6. Statistical analysis: Judith D. Akwo, Phuong Dung (Yun) Trieu.
7. Manuscript preparation: Judith D. Akwo.
8. Manuscript editing: Judith D. Akwo, Phuong Dung (Yun) Trieu, Melissa L. Barron, Tess Reynolds, Sarah J. Lewis.

Conflict of interest

The authors declare no conflict of interest.

References

1. Azamjah N, Soltan-Zadeh Y, Zayeri F. Global trend of breast cancer mortality rate: a 25-year study. *Asian Pac J Cancer Prev* 2019;**20**:2015–20.
2. Huang J, Chan PS, Lok V, *et al*. Global incidence and mortality of breast cancer: a trend analysis. *Aging (Albany NY)* 2021;**13**:5748–803.
3. Morrell S, Taylor R, Roder D, *et al*. Mammography screening and breast cancer mortality in Australia: an aggregate cohort study. *J Med Screen* 2012;**19**:26–34.
4. Olver IN, Roder D. History, development and future of cancer screening in Australia. *Public Health Res Pract* 2017;**27**.
5. Ekpo EU, Alakhras M, Brennan P. Errors in mammography cannot be solved through technology alone. *Asian Pac J Cancer Prev* 2018;**19**:291–301.
6. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights into imaging 2017;**8**(1):171–82.
7. Stachs A, Stubert J, Reimer T, *et al*. Benign breast disease in women. *Dtsch Arztebl Int* 2019;**116**:565–74.
8. Cho N. Imaging features of breast cancer molecular subtypes: state of the art. *J Pathol Transl Med* 2021;**55**:16–25.
9. Hakim CM, Catullo VJ, Chough DM, *et al*. Effect of the availability of prior full-field digital mammography and digital breast tomosynthesis images on the interpretation of mammograms. *Radiology* 2015;**276**:65–72.
10. Hayward JH, Ray KM, Wisner DJ, *et al*. Improving screening mammography outcomes through comparison with multiple prior mammograms. *AJR Am J Roentgenol* 2016;**207**:918–24.
11. Wong DJ, Gandomkar Z, Lewis S, *et al*. Do reader characteristics affect diagnostic efficacy in screening mammography? A systematic review. *Clin Breast Cancer* 2023;**23**:e56–67.
12. Hadadi I, Rae W, Clarke J, *et al*. Diagnostic performance of adjunctive imaging modalities compared to mammography alone in women with

non-dense and dense breasts: a systematic review and meta-analysis. *Clin Breast Cancer* 2021;**21**:278–91.
13. Hadadi I, Rae W, Clarke J, *et al*. Breast cancer detection across dense and non-dense breasts: markers of diagnostic confidence and efficacy. *Acta Radiol Open* 2022;**11**:20584601211072279.
14. Hadadi I, Clarke J, Rae W, *et al*. Diagnostic efficacy across dense and non-dense breasts during digital breast tomosynthesis and ultrasound assessment for recalled women. *Diagnostics (Basel)* 2022;**12**.
15. Payne NR, Hickman SE, Black R, *et al*. Breast density effect on the sensitivity of digital screening mammography in a UK cohort. *Eur Radiol* 2024, https://doi.org/10.1007/s00330-024-10951-w.
16. Gale A, Chen Y. A review of the PERFORMS scheme in breast screening. *Br J Radiol* 2020;**93**:20190908.
17. Suleiman ME, Rickard M, Brennan PC. Perfecting detection through education. *Radiography (Lond)* 2020;**26**:S49–53.
18. Suleiman WI, Rawashdeh MA, Lewis SJ, *et al*. Impact of Breast Reader Assessment Strategy on mammographic radiologists' test reading performance. *J Med Imaging Radiat Oncol* 2016;**60**:352–8.
19. Wang X, Ji X. Sample size estimation in clinical research: from randomized controlled trials to observational studies. *Chest* 2020;**158**:S12–20.
20. Fraker JL, Clune CG, Sahni SK, *et al*. Prevalence, impact, and diagnostic challenges of benign breast disease: a narrative review. *Int J Womens Health* 2023;**15**:765–78.
21. Albus K. Record Keeping: Mammography. 2019 (Revised 12/12/2019) [cited 2024/09/07]; https://accreditationsupport.acr.org/support/solutions/articles/11000051613-record-keeping-mammography-revised-12-12-19-.
22. Mezrich JL, Siegel E. Who owns the image? Archiving and retention issues in the digital age. *J Am Coll Radiol* 2014;**11**:384–6.
23. Taylor-Phillips S, Wallis MG, Duncan A, *et al*. Use of prior mammograms in the transition to digital mammography: a performance and cost analysis. *Eur J Radiol* 2012;**81**:60–5.
24. Thurfjell MG, Vitak B, Azavedo E, *et al*. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. *Acta Radiol* 2000;**41**:52–6.
25. Roelofs AAJ, Karssemeijer N, Wedekind N, *et al*. Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 2007;**242**:70–7.
26. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med* 2013;**11**:106–15.
27. Soh BP, Lee WB, McEntee MF, *et al*. Mammography test sets: reading location and prior images do not affect group performance. *Clin Radiol* 2014;**69**:397–402.
28. Roelofs AA, Karssemeijer N, Wedekind N, *et al*. Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 2007;**242**:70–7.
29. Kim WH, Chang JM, Koo HR, *et al*. Impact of prior mammograms on combined reading of digital mammography and digital breast tomosynthesis. *Acta Radiologica* 2017;**58**:148–55.
30. Trieu PDY, Borecky N, Li T, *et al*. The impact of prior mammograms on the diagnostic performance of radiologists in early breast cancer detection: a focus on breast density, lesion features and vendors using wholly digital screening cases. *Cancers (Basel)* 2023;**15**.
31. Varela C, Karssemeijer N, Hendriks JH, *et al*. Use of prior mammograms in the classification of benign and malignant masses. *Eur J Radiol* 2005;**56**:248–55.
32. Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance?-a scoping review. *BJR Open* 2023;**5**:20230038.
33. Scott HJ, Gale AG. What makes a good mammographer? Looking at the homogeneity of those groups who perform best and least well on the PERFORMS self-assessment scheme. *Breast Cancer Res* 2008;**10**:P74.
34. Qenam BA, Li T, Alshabibi A, *et al*. Test-set results can predict participants' development in breast-screen cancer detection: an observational cohort study. *Health Sci Rep* 2024;**7**:e2161.
35. Chen Y, James JJ, Cornford EJ, *et al*. The relationship between mammography readers' real-life performance and performance in a test set-based assessment scheme in a national breast screening program. *Radiol Imaging Cancer* 2020;**2**:e200016.
36. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? *Am J Roentgenology* 2004;**182**:867–9.