Radiology

ORIGINAL RESEARCH • BREAST IMAGING

Screening Outcomes of Mammography with AI in Dense Breasts: A Comparative Study with Supplemental Screening US

Su Min Ha, MD, PhD • Myoung-jin Jang, PhD • Inyoung Youn, MD, PhD • Heera Yoen, MD • Hye Ji, MD • Su Hyun Lee, MD, PhD • Ann Yi, MD, PhD • Jung Min Chang, MD, PhD

From the Department of Radiology (S.M.H., H.Y., H.J., S.H.L., J.M.C.) and Medical Research Collaborating Center (M.J.J.), Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea; Department of Radiology, Seoul National University College of Medicine, Seoul, Republic of Korea (S.M.H., S.H.L., J.M.C.); Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Republic of Korea (S.M.H.); Department of Radiology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea (I.Y.); and Department of Radiology, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, Republic of Korea (A.Y.). Received December 14, 2023; revision requested February 8, 2024; revision received April 11; accepted April 19. **Address correspondence to** J.M.C. (email: *imchangjm@gmail.com*).

Supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grant HI20C2092), and Seoul National University Hospital Research Fund (grants 06-2022-2110 and 06-2022-3040).

Conflicts of interest are listed at the end of this article.

See also the editorial by Whitman and Destounis in this issue.

Radiology 2024; 312(1):e233391 • https://doi.org/10.1148/radiol.233391 • Content codes: BR AI US

Background: Comparative performance between artificial intelligence (AI) and breast US for women with dense breasts undergoing screening mammography remains unclear.

Purpose: To compare the performance of mammography alone, mammography with AI, and mammography plus supplemental US for screening women with dense breasts, and to investigate the characteristics of the detected cancers.

Materials and Methods: A retrospective database search identified consecutive asymptomatic women (≥40 years of age) with dense breasts who underwent mammography plus supplemental whole-breast handheld US from January 2017 to December 2018 at a primary health care center. Sequential reading for mammography alone and mammography with the aid of an AI system was conducted by five breast radiologists, and their recall decisions were recorded. Results of the combined mammography and US examinations were collected from the database. A dedicated breast radiologist reviewed marks for mammography alone or with AI to confirm lesion identification. The reference standard was histologic examination and 1-year follow-up data. The cancer detection rate (CDR) per 1000 screening examinations, sensitivity, specificity, and abnormal interpretation rate (AIR) of mammography alone, mammography with AI, and mammography plus US were compared.

Results: Among 5707 asymptomatic women (mean age, 52.4 years ± 7.9 [SD]), 33 (0.6%) had cancer (median lesion size, 0.7 cm). Mammography with AI had a higher specificity (95.3% [95% CI: 94.7, 95.8], $P = .003$) and lower AIR (5.0% [95% CI: 4.5, 5.6], $P = .004$) than mammography alone (94.3% [95% CI: 93.6, 94.8] and 6.0% [95% CI: 5.4, 6.7], respectively). Mammography plus US had a higher CDR (5.6 vs 3.5 per 1000 examinations, $P = .002$) and sensitivity (97.0% vs 60.6%, $P = .002$) but lower specificity (77.6% vs 95.3%, $P < .001$) and higher AIR (22.9% vs 5.0%, $P < .001$) than mammography with AI. Supplemental US alone helped detect 12 cancers, mostly stage 0 and I (92%, 11 of 12).

Conclusion: Although AI improved the specificity of mammography interpretation, mammography plus supplemental US helped detect more node-negative early breast cancers that were undetected using mammography with AI.

Supplemental material is available for this article.

Screening mammography allows earlier breast cancer detection and has contributed to improved prognosis and mortality reduction (1). However, mammography is interpreted by the visual perceptions of radiologists, which leads to missed cancers (20%–33%) owing to reasons that include dense parenchyma, perceptual oversight, and interpretation errors (2–4). Because the sensitivity of mammography is as low as 30%–48% in women with dense breasts (5,6), the need for supplemental screening examinations in women with dense breasts has been discussed. However, only a few women will truly benefit from additional supplemental screening tests to detect early breast cancers, while the majority with negative or benign findings will not gain any benefit (7–10). Indeed, detection of these small node-negative invasive cancers comes at the cost of additional health care expenses, low specificity, and an increased biopsy rate without survival benefit (7,10).

Traditional computer-assisted detection software was developed for mammography to improve radiologist performance but has not improved accuracy due to a high number of false-positive recalls and low specificity (11,12). Recent deep learning advances have renewed interest in evaluating a more sophisticated artificial intelligence (AI) system, particularly in its application to mammography (13,14). The current AI system enables radiologist-level interpretation, improving the detection of hidden cancers

Abbreviations

AI = artificial intelligence, AIR = abnormal interpretation rate, BI-RADS = Breast Imaging Reporting and Data System, CDR = cancer detection rate

Summary

Supplemental screening US demonstrated additional node-negative early breast cancers that were overlooked by mammography and artificial intelligence in women with dense breasts, while increasing false-positive findings.

Key Results

- In 5707 asymptomatic women (≥40 years of age) with dense breasts who underwent mammography with supplemental US (33 with cancer), mammography with artificial intelligence (AI) had a higher specificity than mammography alone (95.3% vs 94.3%, $P = .003$) but lower cancer detection rate (3.5 vs 5.6 per 1000 examinations, $P = .002$) and sensitivity (60.6% vs 97.0%, $P = .002$) than mammography plus US.
- Supplemental US had an incremental cancer detection rate of 2.2 per 1000 examinations after negative results according to mammography with AI.

while reducing false-positive results and potentially enhancing reading efficiency (13–16). A recent meta-analysis evaluating standalone AI for screening digital mammography showed that it performed as well as or better than individual breast radiologists or average reader outcomes (17). Improvement in the diagnostic performance of mammographic readings with AI has been noted in women with both dense and nondense breasts (16,18). However, certain cancers remain undetected by AI (19). The comparative performance evaluation between AI and breast US as a supplementary modality for women with dense breasts undergoing screening mammography is of great interest. Moreover, the potential of AI to replace or diminish the need for supplemental screening tests in women with dense breasts remains uncertain, prompting the question of whether additional supplemental screening tests are necessary when AI-aided mammographic readings yield negative results.

Thus, the aim of this study was to compare the performance of radiologists reading mammograms, mammography aided by a commercially available AI system, and mammography with supplemental US for the screening of asymptomatic women with dense breasts, as well as to investigate the characteristics of the detected cancers.

Materials and Methods

The institutional review board of Seoul National University Hospital approved this retrospective study and waived the requirement for informed consent. A total of 1798 women in the current study were previously reported in a study that evaluated outcomes of mammography combined with US compared with digital breast tomosynthesis combined with US (20).

Study Cohort

A retrospective search was performed at Seoul National University Hospital Healthcare System Gangnam Center for screening breast examinations performed from January 2017 to December 2018 in consecutive asymptomatic women (≥40 years of age) with mammographic densities classified as Breast Imaging Reporting and Data System (BI-RADS) categories c (heterogeneously dense) or d (extremely dense) (21) who underwent routine screening mammography and supplemental US. Only the earliest screening examination was used for women who were screened twice during the study period. Women were excluded for having a follow-up period of less than 1 year, augmentation mammoplasty, and personal history of breast cancer.

Image Acquisition

All imaging data were obtained prospectively as part of our routine clinical practice and stored in a picture archiving and communication system. All two-dimensional digital mammographic imaging data were acquired using a full-field unit (Selenia Dimensions [Hologic], Senographe 2000D [GE HealthCare]). The standard protocol at the study institution includes acquisition of bilateral two-view (mediolateral oblique, craniocaudal) mammograms. Whole-breast handheld US examinations were performed by one of three dedicated breast radiologists, one of whom was an author (A.Y., with 16 years of experience), who used a 14–16-MHz linear transducer with an EUB-8500 (Hitachi Medical) system. The US imaging protocol is detailed in Appendix S1. At the time of image acquisition, an AI system was not installed at the institution.

Image Interpretation

When women were scheduled for mammography and US on the same day, the radiologist who performed US reported the mammographic findings separately using the fifth edition of BI-RADS, with scores of 1–5 (21), before the US examination; they subsequently reported the combined final assessment after the US examination. In the combined mammography plus US reports, cases assigned a BI-RADS category of 3, 4, or 5 were considered positive screening results, and those assigned a BI-RADS category of 1 or 2 were considered negative (Appendix S1). Recommendations for routine annual follow-up examinations were made for negative cases, whereas a short-interval follow-up of 6 months, additional mammographic views, or biopsy were recommended for positive cases (Fig S1).

AI System

A commercially available AI system (Lunit INSIGHT MMG, version 1.1.7.1; Lunit) that was previously validated for breast cancer detection on mammograms in a multinational study was used for AI analysis (16). Using this algorithm, the AI drew marks on mammograms at the location of abnormalities and provided corresponding abnormality scores. These scores were presented as percentages (0%–100%) for each craniocaudal and mediolateral oblique view. The maximum value per breast was considered the final score for each breast. Lesions with abnormality scores greater than or equal to 10% were considered positive findings (16).

Mammographic Review with or without AI Aid

The AI system was first implemented at the study institution in 2020. The reader study was designed and conducted from November 2022 to April 2023. Because the AI system was

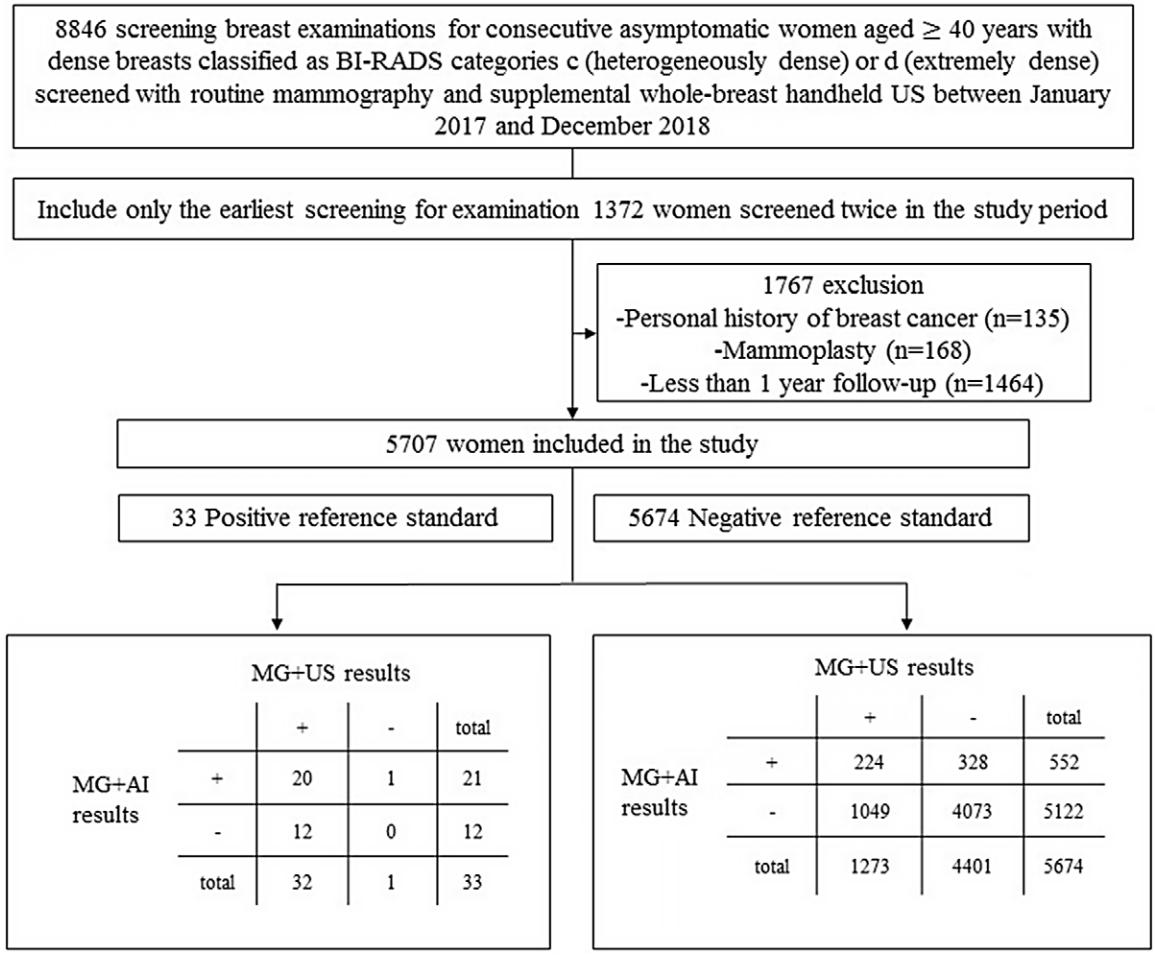

Figure 1: Flowchart shows study inclusion and exclusion. AI = artificial intelligence, BI-RADS = Breast Imaging Reporting and Data System, MG = mammography.

not installed at the time of mammographic acquisition, it was retrospectively applied to all included mammograms acquired from January 2017 to December 2018 in the study cohort. Screening mammograms and AI output results were provided to the readers and were reviewed using the picture archiving and communication system. To acquire the data on final decisions for recall using combined mammography and AI outputs by the radiologists, 1141 or 1142 cases were assigned per radiologist to perform a retrospective reader study for all cases. Sequential reading for the AI-unaided and AI-aided mammography interpretations was performed independently by five breast radiologists (I.Y., H.Y., H.J., S.H.L., and J.M.C. with 12, 3, 2, 12, and 16 years of experience, respectively). Radiologists were blinded to any information about the women, including previous radiology and histopathology reports. Radiologist decisions on whether or not to recall based on mammographic results, without referring to the AI-provided results, were recorded. Subsequently, the AI output was displayed to the radiologist in the same reading session (without a washout period) to decide on recall, with the understanding that maximum pixel-level abnormality scores of greater than or equal to 10% in the AI reports were actionable findings. For both reading sessions, the radiologists recorded the locations of the findings in the recalled cases. Radiologists were not allowed to change their assessments for the AI-unaided readings (Fig S2).

Data Collection and Reference Standards

Details of the data collection are provided in Appendix S1. The risk for invasive breast cancer and interval invasive breast cancer were estimated using the Breast Cancer Surveillance Consortium (BCSC) 5-year risk model, version 2 (22). Data were collected from the aforementioned reader study to make the final decision on recall for mammography alone and combined mammography with AI results. All markers at sites without cancer were considered false-positive findings. For the final decision on combined mammography plus supplemental US, the report initially made by the performing radiologists without a retrospective reader study of the static US images was used. Preoperative breast MRI, mammograms obtained after wire localization, or both were used to determine the reference location of the cancer. A dedicated breast radiologist (S.M.H., with 10 years of experience) reviewed the marks reported by the five readers for mammography alone and mammography with AI to confirm whether the readers correctly identified the lesions. Histologic examinations and 1-year follow-up data were used as reference standards. Lesions diagnosed as ductal carcinoma

Table 1: Characteristics of the Screening Cohort

Characteristic	Entire Screening Cohort (*n* = 5707)	Women with Breast Cancer (*n* = 33)	Women Without Breast Cancer (*n* = 5674)	*P* Value
Age (y)	52.4 ± 7.9 (40–77)	51.6 ± 6.5 (40–69)	52.4 ± 7.9 (40–77)	.53
First-degree family history of breast cancer				.03
No	5386 (94)	28 (85)	5358 (94)	
Yes	321 (6)	5 (15)	316 (6)	
Menopausal status				.26
Premenopausal	2393 (42)	17 (52)	2376 (42)	
Postmenopausal	3314 (58)	16 (48)	3298 (58)	
Hormone replacement therapy				.11
No	4979 (87)	32 (97)	4947 (87)	
Yes	728 (13)	1 (3)	727 (13)	
BI-RADS breast density				.75
Heterogeneously dense (category c)	3608 (63)	20 (61)	3588 (63)	
Extremely dense (category d)	2099 (37)	13 (39)	2086 (37)	
BCSC 5-year risk				.11
Low	2296 (40)	14 (42)	2282 (40)	
Average	2841 (50)	12 (36)	2829 (50)	
Intermediate	460 (8)	7 (21)	453 (8)	
High or very high	62 (1)	0 (0)	62 (1)	
Not calculated	48 (1)	0 (0)	48 (1)	

Note.—Except where indicated, data are numbers of women, with percentages in parentheses, or means ± SDs, with ranges in parentheses. For continuous and categorical variables, the *t* test or Wilcoxon rank sum test and χ^2 test or Fisher exact test were used, respectively. Five-year risk of invasive breast cancer was evaluated using the BCSC risk calculator, version 2 (22), and categorized as low (<1.0%), average (1.00%–1.66%), intermediate (1.67%–2.49%), high (2.50%–3.99%), and very high (>3.99%). Risk was not calculated for women older than 74 years. BCSC = Breast Cancer Surveillance Consortium, BI-RADS = Breast Imaging Reporting and Data System.

Table 2: Comparison of Diagnostic Performance between Mammography Alone, Mammography with AI, and Mammography Plus US

Variable	Mammography Alone	Mammography with AI	Mammography Plus US	*P* Value*	*P* Value†	*P* Value‡
CDR§	3.3 (2.1, 5.2) [19/5707]	3.5 (2.3, 5.4) [20/5707]	5.6 (3.9, 7.9) [32/5707]	.95	.001	.002
Sensitivity	57.6 (40.8, 72.8) [19/33]	60.6 (43.7, 75.3) [20/33]	97.0 (84.7, 99.5) [32/33]	.95	.001	.002
Specificity	94.3 (93.6, 94.8) [5349/5674]	95.3 (94.7, 95.8) [5408/5674]	77.6 (76.5, 78.6) [4401/5674]	.003	<.001	<.001
AIR	6.0 (5.4, 6.7) [344/5707]	5.0 (4.5, 5.6) [286/5707]	22.9 (21.8, 24.0) [1305/5707]	.004	<.001	<.001
PPV	5.5 (3.6, 8.5) [19/344]	7.0 (4.6, 10.6) [20/286]	2.5 (1.7, 3.4) [32/1305]	.005	<.001	<.001
NPV	99.7 (99.6, 99.8) [5349/5363]	99.8 (99.6, 99.9) [5408/5421]	99.9 (99.9, 100.0) [4401/4402]	.75	.03	.04

Note.—Except where indicated, data are percentages, with 95% CIs in parentheses and numbers of examinations in brackets. The screening cohort of 5707 women included 33 with breast cancer and 5674 without breast cancer, and one screening examination per woman was assessed. Mammography alone and mammography with AI, which entailed AI-assisted mammography interpretation, refer to the retrospective reader study. Mammography plus US refers to the prospectively assessed combined reading of mammography and US. *P* values were obtained using the McNemar test, or using generalized estimating equations for PPV and NPV, with Bonferroni correction. AI = artificial intelligence, AIR = abnormal interpretation rate, CDR = cancer detection rate, NPV = negative predictive value, PPV = positive predictive value of recall.

* Mammography alone versus mammography with AI.

† Mammography alone versus mammography plus US.

‡ Mammography with AI versus mammography plus US.

§ Data are reported as the rate per 1000 examinations.

Table 3: Summary of Cancer Characteristics

Characteristic	Total Cancers (n = 33)	Cancers Detected Using Mammography with AI or Standalone AI (n = 21)*	Cancers Detected Using Supplemental US After Negative Result from Mammography with AI (n = 12)†	P Value
Histologic type				.46
Noninvasive	13 (39)	7 (33)	6 (50)	
Invasive	20 (61)	14 (67)	6 (50)	
Invasive ductal	15 (75)	11 (79)	4 (67)	
Invasive lobular	2 (10)	1 (7)	1 (17)	
Mixed ductal and lobular	1 (5)	1 (7)	0 (0)	
Mucinous	2 (10)	1 (7)	1 (17)	
Lesion size (cm)‡	0.7 (0–1.5)	1.0 (0–1.5)	0.3 (0–1.5)	.28
Lymph node status				.52
No metastasis	31 (94)	19 (90)	12 (100)	
Metastasis	2 (6)	2 (10)	0 (0)	
Stage				.50
DCIS, stage 0	13 (39)	7 (33)	6 (50)	
Invasive cancer, stage I	14 (42)	9 (43)	5 (42)	
Invasive cancer, stage II	6 (18)	5 (24)	1 (8)	
Molecular subtype§				.91
ER or PR positive	23 (70)	14 (67)	9 (75)	
HER2 positive	6 (18)	4 (19)	2 (17)	
Triple negative	2 (6)	2 (10)	0 (0)	
Unknown	2 (6)	1 (4)	1 (8)	

Note.—Except where indicated, data are numbers of cancers, with percentages in parentheses. A total of 33 cancers were identified in 33 women (Table S2). Mammography with AI, which entailed AI-assisted mammography interpretation, refers to the retrospective reader study. For continuous and categorical variables, the t test or Wilcoxon rank sum test and χ^2 test or Fisher exact test were used, respectively. AI = artificial intelligence, DCIS = ductal carcinoma in situ, ER = estrogen receptor, HER2 = human epidermal growth factor receptor 2, PR = progesterone receptor.

* Includes one cancer that was not detected with mammography during the reader study but was detected by standalone AI.

† No cancers were detected with mammography alone or mammography with AI during the reader study, or standalone AI.

‡ Data are medians, with IQRs in parentheses.

§ Hormone receptor positivity was defined as ER and/or PR positivity of greater than or equal to 1% with nuclear staining using standard immunohistochemistry methods. HER2 positivity was defined as a HER2 score of 3+ or gene amplification according to fluorescence in situ hybridization in tumors with a HER2 score of 2+.

in situ or invasive carcinoma at biopsy or surgery were considered malignant. High-risk lesions were considered benign. Lesions diagnosed as benign at biopsy or surgery or those that did not change over follow-up for 1 year or more were considered benign (Appendix S1).

Statistical Analysis

In the mammography alone or mammography with AI reader study, reader assessments of no recall were considered test-negative, while reader assessments of recall were considered test-positive. In combined mammography plus US reports, cases assigned a BI-RADS category of 1 or 2 were considered test-negative, while those assigned a BI-RADS category of 3, 4, or 5 were considered test-positive. The cancer detection rate (CDR), sensitivity, specificity, and abnormal interpretation rate (AIR) of mammography, mammography with AI, and mammography plus US were calculated for the entire study sample (n = 5707). Definitions of CDR, sensitivity, specificity, and AIR are provided in Appendix S1. The characteristics of the study cohort and identified cancers were compared. For continuous and categorical variables, the t test or Wilcoxon rank sum test and the χ^2 test or Fisher exact test were used, respectively, and 95% CIs were estimated using the Wilson score method. P values for the comparison of outcomes were obtained using the McNemar test or through generalized estimating equations, with Bonferroni correction. Two-sided $P < .05$ was considered indicative of a statistically significant difference. Statistical analysis was performed using SAS (version 9.4; SAS Institute). The post hoc power analysis was performed using PASS 2022 (NCSS) (Appendix S1).

Results

Cohort Characteristics

A total of 8846 examinations were conducted in consecutive asymptomatic women with dense breasts who were screened with digital mammography and whole-breast handheld US during the study period. For 1372 women screened twice in the study period, only the earliest examination was included, and we

excluded 1767 women for having less than 1 year of follow-up (n = 1464), augmentation mammoplasty (n = 168), and a personal history of breast cancer (n = 135). Demographic characteristics of the 1464 women excluded due to inadequate follow-up are provided in Table S1. Finally, 5707 women were included in the study (Fig 1). Demographic characteristics of the entire screening cohort and according to women with and without breast cancer are described in Table 1. The women ranged in age from 40 to 77 years (mean age, 52.4 years ± 7.9 [SD]). Of the 5707 women, 33 (0.6%) had cancer identified at screening and 5674 (99.4%) had benign or negative results, and breast density was classified as heterogeneously dense in 3608 (63%) and extremely dense in 2099 (37%). Among the 5707 women, 2393 (42%) were premenopausal, 728 (13%) had undergone hormone replacement therapy, and 321 (6%) had a family history of breast cancer. Additionally, 5137 of 5707 women (90%) had low or average (<1.67%) Breast Cancer Surveillance Consortium 5-year risk scores, while 62 of 5707 (1%) had high (2.50%–3.99%) or very high (>3.99%) risk scores.

Outcome Measures of CDR, Sensitivity, Specificity, and AIR

The screening cohort of 5707 women included 33 with breast cancer and 5674 without breast cancer, and one screening examination per woman was assessed. The CDR, sensitivity, specificity, and AIR data are presented in Table 2. The CDR was 3.3 (19 of 5707; 95% CI: 2.1, 5.2) per 1000 examinations for mammography alone and 3.5 (20 of 5707; 95% CI: 2.3, 5.4) per 1000 examinations for mammography with AI (P = .95). The sensitivity was 57.6% (19

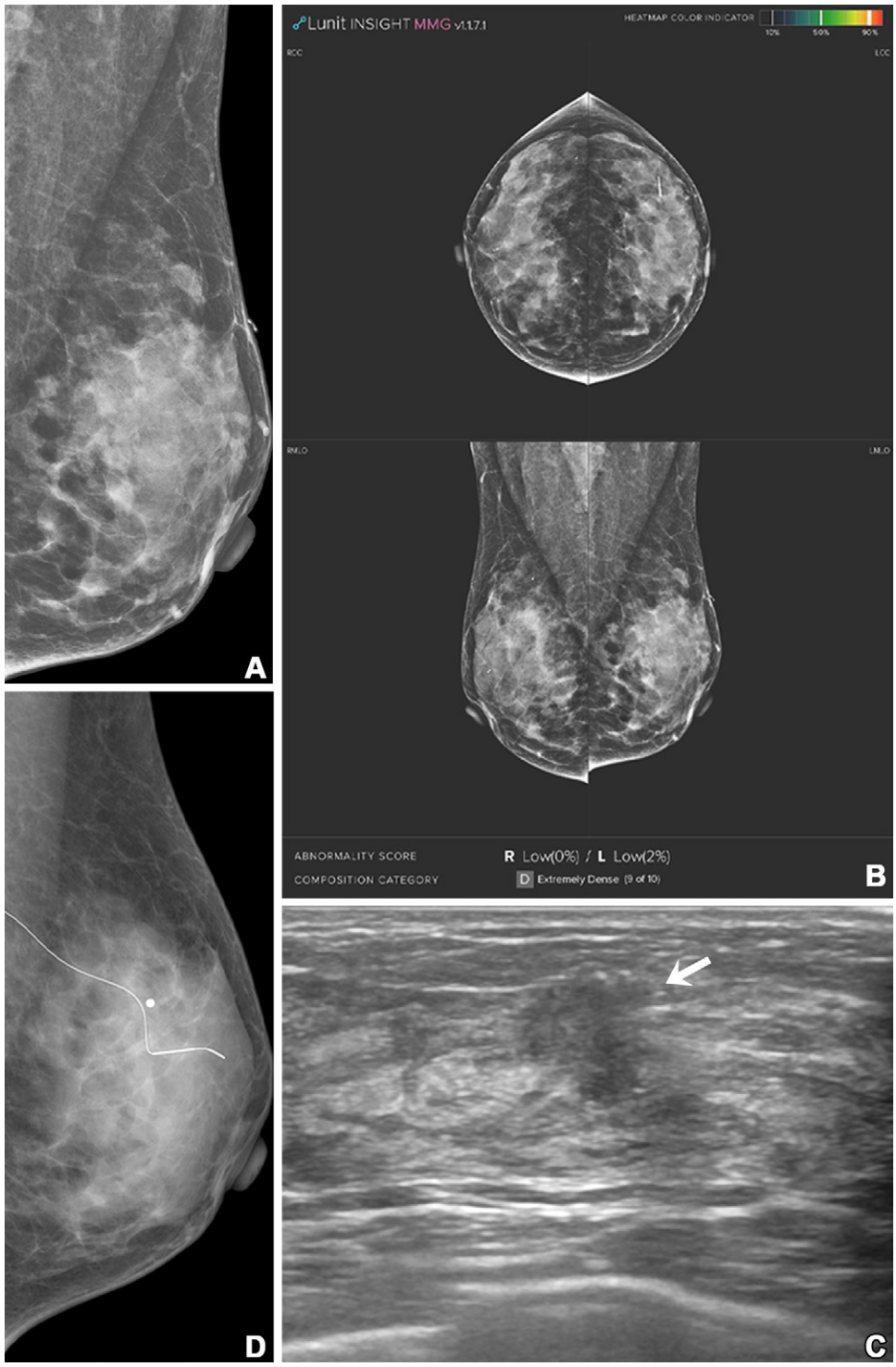

Figure 2: Imaging in a 51-year-old woman with extremely dense breasts (Breast Imaging Reporting and Data System category d). **(A, B)** Left mediolateral oblique digital mammogram **(A)** and right and left craniocaudal (RCC, LCC) and right and left mediolateral oblique (RMLO, LMLO) mammograms with artificial intelligence (Lunit INSIGHT MMG, version 1.1.7.1; Lunit) **(B)** show no abnormality. The radiopaque bar marker, which is attached on the skin to mark a previous benign fibroadenoma excision site, is visible. **(C)** Supplemental transverse US image shows a 1.2-cm irregular mass (arrow) in the left breast. **(D)** Left mediolateral oblique digital mammogram after US-guided wire localization shows no suspicious findings, even in retrospect. The radiopaque round marker, which is attached on the skin to mark the wire insertion site, is visible. The patient was treated with breast-conserving surgery, and the mass was proven to be invasive ductal carcinoma (pT1N0, 1.5 cm in size, estrogen receptor and progesterone receptor positive, human epidermal growth factor receptor 2 negative, histologic grade 1).

of 33 examinations; 95% CI: 40.8, 72.8) for mammography and 60.6% (20 of 33 examinations; 95% CI: 43.7, 75.3) for mammography with AI (P = .95). The specificity was 94.3% (5349 of 5674 examinations; 95% CI: 93.6, 94.8) for mammography and 95.3% (5408 of 5674 examinations; 95% CI: 94.7, 95.8) for mammography with AI (P = .003). Mammography with AI showed a lower AIR of 5.0% (286 of 5707 examinations; 95% CI: 4.5, 5.6) than the 6.0% (344 of 5707 examinations; 95% CI: 5.4, 6.7; P = .004) for mammography alone. Compared with mammography alone, mammography plus supplemental US showed a higher CDR of 5.6 (32 of 5707; 95% CI: 3.9, 7.9; P = .001) per 1000 examinations and sensitivity of 97.0% (32 of 33 examinations; 95% CI: 84.7, 99.5; P = .001), with a lower specificity of 77.6% (4401 of 5674 examinations; 95% CI: 76.5, 78.6; P < .001) and a higher AIR of 22.9% (1305 of 5707 examinations; 95% CI: 21.8, 24.0; P < .001). When comparing mammography with AI and mammography plus US, the latter showed a higher CDR (5.6 vs 3.5 per 1000 examinations, P = .002) and sensitivity (97.0% vs 60.6%, P = .002), with a lower specificity (77.6% vs 95.3%, P < .001) and higher AIR (22.9% vs 5.0%, P < .001). Among 5421 women with negative results based on mammography with AI, supplemental US helped detect 12 cancers, with an incremental CDR of 2.2 (12 of 5421; 95% CI: 1.1, 3.8) per 1000 examinations.

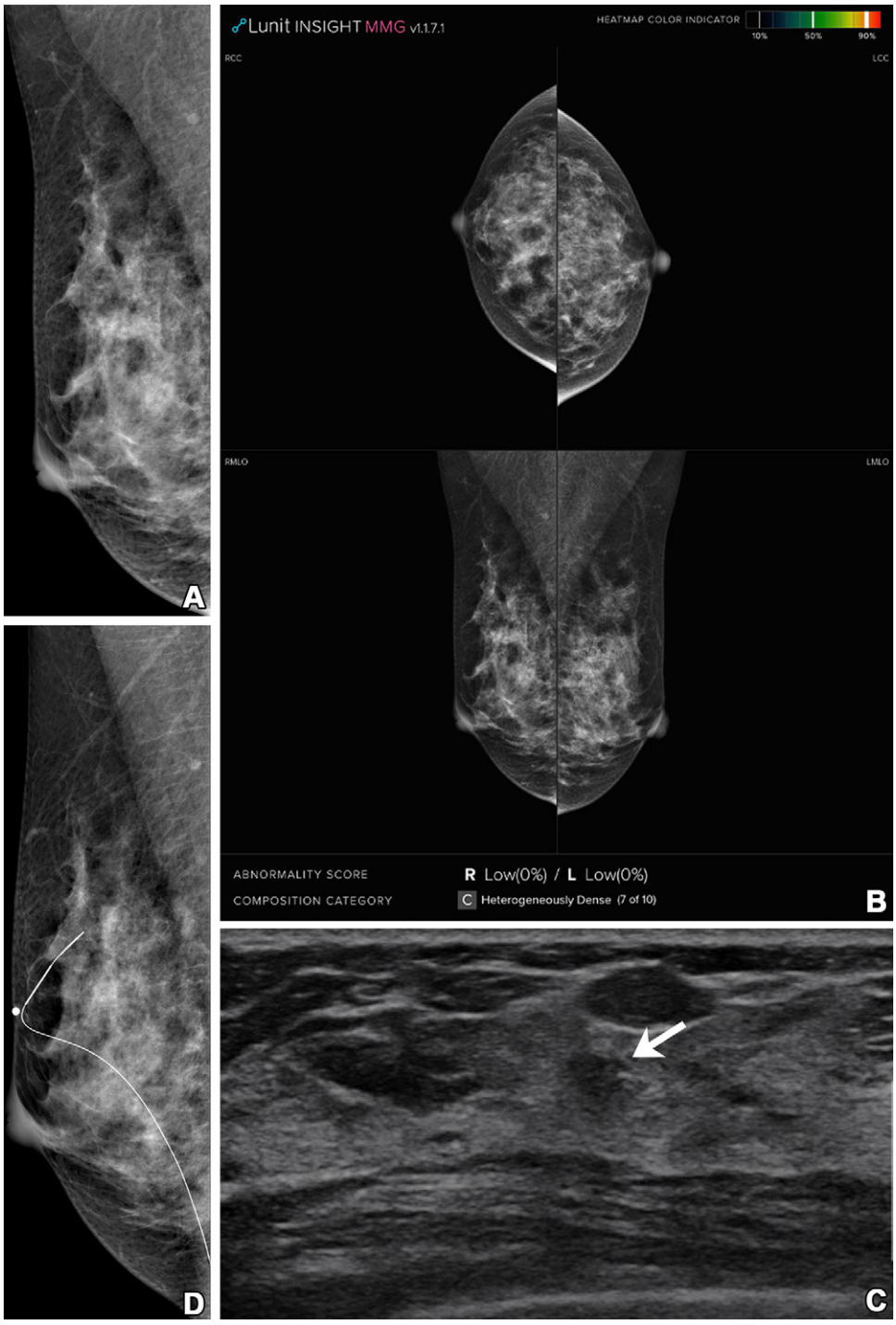

Figure 3: Imaging in a 41-year-old woman with heterogeneously dense breasts (Breast Imaging Reporting and Data System category c). **(A, B)** Right mediolateral oblique digital mammogram **(A)** and right and left craniocaudal (RCC, LCC) and right and left mediolateral oblique (RMLO, LMLO) mammograms with artificial intelligence (Lunit INSIGHT MMG, version 1.1.7.1; Lunit) **(B)** show no abnormality. **(C)** Supplemental transverse US image shows a 1.0-cm irregular mass (arrow) in the right breast. **(D)** Right mediolateral oblique digital mammogram after US-guided wire localization shows no suspicious findings, even in retrospect. The radiopaque round marker, which is attached on the skin to mark the wire insertion site, is visible. The patient was treated with breast-conserving surgery, and the mass was proven to be invasive ductal carcinoma (pT1N0, 0.7 cm in size, estrogen receptor and progesterone receptor positive, human epidermal growth factor receptor 2 negative, histologic grade 1).

Characteristics of Detected and Missed Breast Cancers at Mammography with AI

The 33 detected cancers (median lesion size, 0.7 cm; IQR, 0–1.5 cm) consisted of 20 (61%) invasive cancers and 13 (39%)

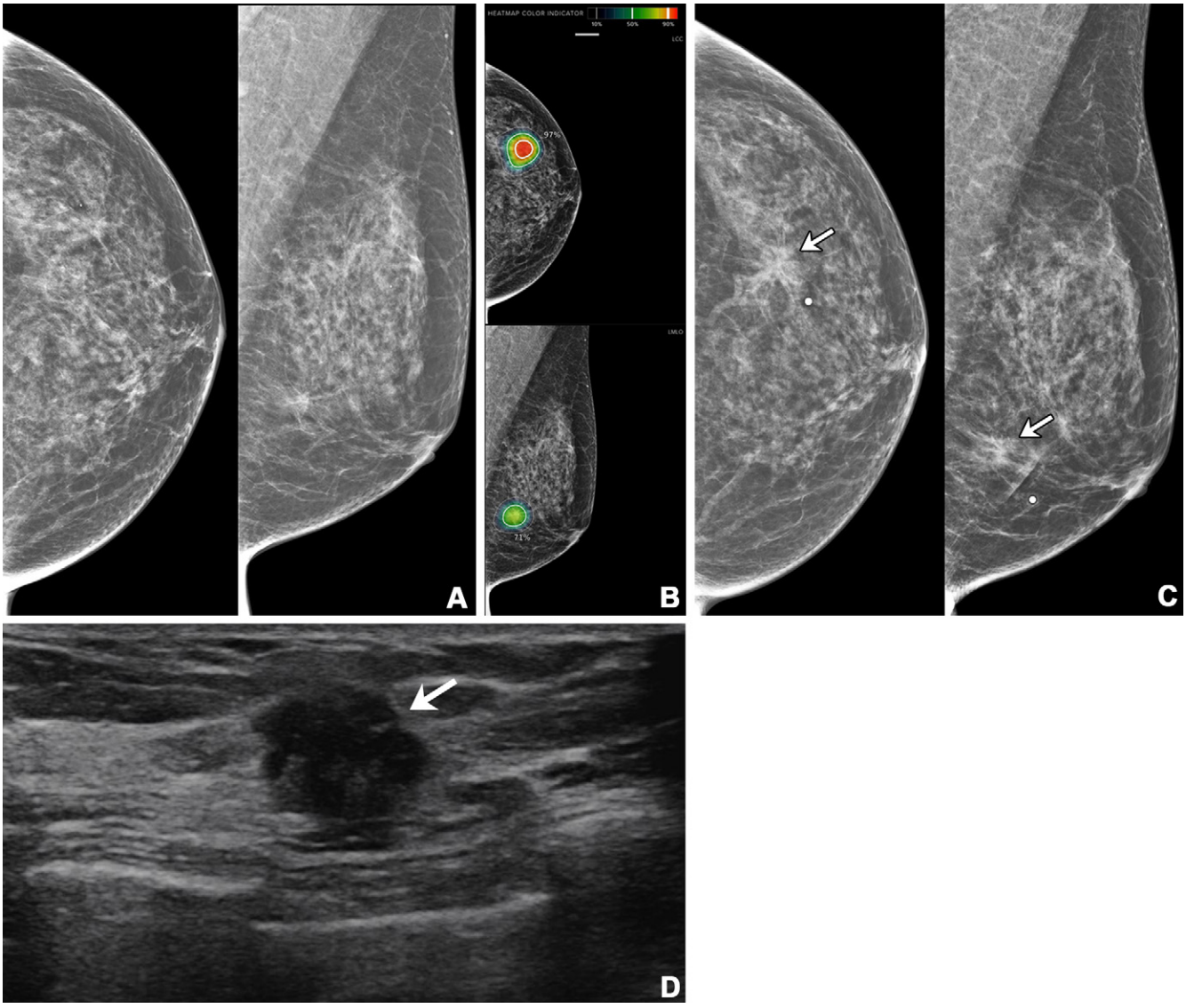

Figure 4: Imaging in a 56-year-old woman with heterogeneously dense breasts (Breast Imaging Reporting and Data System category c). **(A)** Left craniocaudal (left) and mediolateral oblique (right) digital mammograms show no abnormality. **(B)** Left craniocaudal (LCC, top) and left mediolateral oblique (LMLO, bottom) mammograms with retrospectively applied artificial intelligence (AI) (Lunit INSIGHT MMG, version 1.1.7.1; Lunit) show the lesion was given abnormality scores of 97% and 71%, respectively. The color bar indicates pixel-level abnormality scores corresponding to heatmap contour lines. Breast US performed the same day was considered negative (not shown). **(C)** Left craniocaudal (left) and mediolateral oblique (right) digital mammograms acquired 9 months later show an irregular mass (arrows) in the left breast, which correlates with the heatmap areas scored as 97% and 71% by AI in the earlier screening mammogram **(B)**. The radiopaque round marker denotes the site of the palpable abnormality indicated by the patient. **(D)** Transverse US image shows an irregular mass (arrow) in the left breast. The patient was treated with breast-conserving surgery, and the mass was proven to be invasive lobular carcinoma (pT2N1, 2.5 cm in size, estrogen receptor and progesterone receptor negative, human epidermal growth factor receptor 2 positive, histologic grade 3).

ductal carcinoma in situ (Table 3). Of these cancers, 94% (31 of 33) were without lymph node metastasis, 82% (27 of 33) were stage 0 or I, and 70% (23 of 33) were a luminal type. Cancers not seen at mammography by radiologists or AI but detected using supplemental US (n = 12) were invasive, with a median lesion size of 0.3 cm (IQR, 0–1.5 cm) (Fig 2). Of these US-detected cancers, 100% (12 of 12) were without lymph node metastasis, 92% (11 of 12) were stage 0 or I, and 75% (nine of 12) were a luminal type (Fig 3). Standalone AI identified a 2.5-cm invasive lobular carcinoma that was not detected using either mammography or US. In the reader study, the radiologist deemed this a false-positive finding and dismissed the AI results (Fig 4). Further details on the 33 cancers are provided in Table S2.

Discussion

Comparative performance between artificial intelligence (AI) and supplemental breast US for women with dense breasts undergoing screening mammography remains unclear. While early studies suggested AI systems could aid radiologists (17), recent studies have reported a relative decline in the ability of AI to detect malignant lesions (23,24), similar to human readers' limited sensitivity in dense breast mammography interpretation. Thus, in this retrospective study, we compared the

performance of mammography alone, mammography with AI, and mammography plus supplemental US for screening women with dense breasts, and investigated the characteristics of the detected cancers. The results showed that AI improved the specificity of mammography interpretation in women with dense breasts (95.3% vs 94.3%, $P = .003$), most of whom had low to average risk of breast cancer. However, mammography plus supplemental screening breast US outperformed mammography with AI, with a higher cancer detection rate (5.6 vs 3.5 per 1000 examinations, $P = .002$) and sensitivity (97.0% vs 60.6%, $P = .002$), despite lower specificity (77.6% vs 95.3%, $P < .001$) and an increased abnormal interpretation rate (22.9% vs 5.0%, $P < .001$). Supplemental US helped detect 12 additional cancers (six invasive cancers and six ductal carcinomas in situ) that were undetected using mammography and AI. AI may help radiologists interpret screening mammography, but it cannot fully compensate for its low sensitivity in women with dense breasts.

In recent years, numerous mammographic AI systems have been developed and validated, and they are commercially available. Preliminary studies demonstrated that AI as a concurrent decision support tool for mammography interpretation improved reading time, sensitivity, and specificity (16,25,26). In a recent study using screening data sets similar to ours (23), mammography alone showed significantly higher specificity and accuracy, with a lower recall rate and no difference in CDR, compared with standalone AI. In that study (23), mammography with US, compared with mammography with AI and mammography with both US and AI, showed significantly higher specificity and accuracy without significant difference in CDR, suggesting that AI does not improve the performance of screening mammography with or without breast US in women with dense breasts (23). The difference between the study results could be attributed to the small number of cancers in our screening cohort compared with that of previous studies with enriched study samples. Literature investigating the performance of screening mammography with AI compared with other supplemental screening modalities remains scarce. Our study affirms that AI enhances mammography interpretations by radiologists. However, from an early cancer detection standpoint, offering supplemental breast US for women with dense breasts, even after a negative assessment according to AI, could be a method to reduce the number of interval cancers.

Several attempts have been made to integrate AI into clinical practice; however, its optimal use remains to be determined. In one study (27), AI could be used as a single reader, reducing radiologist workload, while simultaneously selecting women for enhanced supplemental screening. The authors suggested that radiologists could consider routing women with high AI scores to targeted US examinations. However, while triage performed by AI could rule out considerable numbers of negative mammograms, it may result in the dismissal of cancers (28). Cancers overlooked by AI were frequently occult in asymptomatic patients with extremely dense breasts. Consistent with our study, cancers missed by AI were often ductal carcinoma in situ; smaller in size, as more than 70% of invasive cancers were less than 1 cm; and the luminal subtype, without axillary lymph node metastasis (24). Thus, negative AI results should be interpreted with caution, as they cannot assure negative outcomes, especially in women with dense breasts.

This study had several limitations. First, this was a retrospective study performed at a single institution, and women without at least 1 year of follow-up were excluded due to insufficient data for the reference standard. Despite the consecutive enrollment of women, selection bias was inevitable due to the retrospective nature and exclusion. Second, a retrospective sequential reader study design was employed to mirror real-world clinical application of an AI system. However, without clinical information or prior imaging examinations, the assessment was challenging for readers. Third, 90% of our study sample had low or average breast cancer risk, while 1% had high or very high risk. Lastly, we used a single commercial AI system, and our results cannot be generalized to other AI systems.

In conclusion, using an artificial intelligence (AI) system to interpret screening mammograms reduced false-positive results and improved specificity in asymptomatic women with dense breasts. However, mammography plus supplemental US still helped detect more node-negative early breast cancers, even after negative results according to mammography and AI. Subsequent study on the clinical application of AI for women undergoing mammography and supplemental screening US, prospectively incorporating more clinical and temporal data, will provide a more definitive answer on the potential of AI to replace supplemental screening US.

Deputy Editor: Kathryn Fowler
Scientific Editor: Sarah Atzen

Author contributions: Guarantors of integrity of entire study, **S.M.H., J.M.C.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **S.M.H., J.M.C.**; clinical studies, **S.M.H., I.Y., H.Y., H.J., S.H.L., A.Y., J.M.C.**; experimental studies, **H.Y., A.Y.**; statistical analysis, **S.M.H., M.J.J., J.M.C.**; and manuscript editing, **S.M.H., J.M.C.**

Disclosures of conflicts of interest: **S.M.H.** No relevant relationships. **M.J.J.** No relevant relationships. **I.Y.** No relevant relationships. **H.Y.** No relevant relationships. **H.J.** No relevant relationships. **S.H.L.** No relevant relationships. **A.Y.** No relevant relationships. **J.M.C.** Research consultant for Hologic; associate editor of *Radiology*.

References

1. Tabar L, Yen MF, Vitak B, Chen HH, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet 2003;361(9367):1405–1410.
2. Hoff SR, Abrahamsen AL, Samset JH, Vigeland E, Klepp O, Hofvind S. Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography-- results from a retrospective review. Radiology 2012;264(2):378–386.
3. Lamb LR, Mohallem Fonseca M, Verma R, Seely JM. Missed Breast Cancer: Effects of subconscious bias and lesion characteristics. RadioGraphics 2020;40(4):941–960.
4. Hovda T, Tsuruda K, Hoff SR, Sahlberg KK, Hofvind S. Radiological review of prior screening mammograms of screen-detected breast cancer. Eur Radiol 2021;31(4):2568–2579.
5. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. Radiology 2002;225(1):165–175.
6. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 2000;92(13):1081–1087.

7. Melnikow J, Fenton JJ, Whitlock EP, et al. Supplemental screening for breast cancer in women with dense breasts: A systematic review for the U.S. preventive services task force. Ann Intern Med 2016;164(4):268–278.
8. Berg WA, Bandos AI, Mendelson EB, Lehrer D, Jong RA, Pisano ED. Ultrasound as the primary screening test for breast cancer: analysis from ACRIN 6666. J Natl Cancer Inst 2015;108(4):djv367.
9. Corsetti V, Houssami N, Ghirardi M, et al. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. Eur J Cancer 2011;47(7):1021–1026.
10. Sprague BL, Stout NK, Schechter C, et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. Ann Intern Med 2015;162(3):157–166.
11. Lehman CD, Wellman RD, Buist DS, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 2015;175(11):1828–1837.
12. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med 2007;356(14):1399–1409.
13. Houssami N, Lee CI, Buist DSM, Tao D. Artificial intelligence for breast cancer screening: Opportunity or hype? Breast 2017;36:31–33.
14. Yoon JH, Kim EK. Deep Learning-Based Artificial Intelligence for Mammography. Korean J Radiol 2021;22(8):1225–1239.
15. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. Nature 2020;577(7788):89–94. [Published correction appears in Nature 2020;586(7829):E19.]
16. Kim HE, Kim HH, Han BK, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. Lancet Digit Health 2020;2(3):e138–e148.
17. Yoon JH, Strand F, Baltzer PAT, et al. Standalone AI for breast cancer detection at screening digital mammography and digital breast tomosynthesis: a systematic review and meta-Analysis. Radiology 2023;307(5):e222639.
18. Kim EK, Kim HE, Han K, et al. Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. Sci Rep 2018;8(1):2762.
19. Yoen H, Chang JM. Artificial intelligence improves detection of supplemental screening ultrasound-detected breast cancers in mammography. J Breast Cancer 2023;26(5):504–513.
20. Ha SM, Yi A, Yim D, et al. Digital breast tomosynthesis plus ultrasound versus digital mammography plus ultrasound for screening breast cancer in women with dense breasts. Korean J Radiol 2023;24(4):274–283.
21. D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BIRADS Atlas, Breast Imaging Reporting and Data System. Reston, Va: American College of Radiology, 2013.
22. Breast Cancer Surveillance Consortium. Breast Cancer Surveillance Consortium Risk Calculator. https://tools.bcsc-scc.ucdavis.edu/BC5yearRisk_V2/calculator.htm. Accessed November 1, 2023.
23. Lee SE, Yoon JH, Son NH, Han K, Moon HJ. Screening in Patients With Dense Breasts: Comparison of Mammography, Artificial Intelligence, and Supplementary Ultrasound. AJR Am J Roentgenol 2024;222(1):e2329655.
24. Lee SE, Han K, Yoon JH, Youk JH, Kim EK. Depiction of breast cancers on digital mammograms by artificial intelligence-based computer-assisted diagnosis according to cancer characteristics. Eur Radiol 2022;32(11):7400–7408.
25. Pacilè S, Lopez J, Chone P, Bertinotti T, Grouin JM, Fillard P. Improving breast cancer detection accuracy of mammography with the concurrent use of an artificial intelligence tool. Radiol Artif Intell 2020;2(6):e190208.
26. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology 2019;290(2):305–314.
27. Dembrower K, Wåhlin E, Liu Y, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. Lancet Digit Health 2020;2(9):e468–e474.
28. Hickman SE, Payne NR, Black RT, et al. Mammography breast cancer screening triage using deep learning: a UK retrospective study. Radiology 2023;309(2):e231173.