Original Article | Breast Imaging

eISSN 2005-8330
https://doi.org/10.3348/kjr.2024.0664
Korean J Radiol 2025;26(3):217-229

Korean Journal of Radiology
KJR

Performance of Digital Mammography-Based Artificial Intelligence Computer-Aided Diagnosis on Synthetic Mammography From Digital Breast Tomosynthesis

Kyung Eun Lee[1], Sung Eun Song[2], Kyu Ran Cho[2], Min Sun Bae[3], Bo Kyoung Seo[3], Soo-Yeon Kim[1]*, Ok Hee Woo[1]*

[1]Department of Radiology, Korea University Guro Hospital, Korea University College of Medicine, Seoul, Republic of Korea
[2]Department of Radiology, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Republic of Korea
[3]Department of Radiology, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, Republic of Korea

Objective: To test the performance of an artificial intelligence-based computer-aided diagnosis (AI-CAD) designed for full-field digital mammography (FFDM) when applied to synthetic mammography (SM).

Materials and Methods: We analyzed 501 women (mean age, 57 ± 11 years) who underwent preoperative mammography and breast cancer surgery. This cohort consisted of 1002 breasts, comprising 517 with cancer and 485 without. All patients underwent digital breast tomosynthesis (DBT) and FFDM during the preoperative workup. The SM is routinely reconstructed using DBT. Commercial AI-CAD (Lunit Insight MMG, version 1.1.7.2) was retrospectively applied to SM and FFDM to calculate the abnormality scores for each breast. The median abnormality scores were compared for the 517 breasts with cancer using the Wilcoxon signed-rank test. Calibration curves of abnormality scores were evaluated. The discrimination performance was analyzed using the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity using a 10% preset threshold. Sensitivity and specificity were further analyzed according to the mammographic and pathological characteristics. The results of SM and FFDM were compared.

Results: AI-CAD demonstrated a significantly lower median abnormality score (71% vs. 96%, $P < 0.001$) and poorer calibration performance for SM than for FFDM. SM exhibited lower sensitivity (76.2% vs. 82.8%, $P < 0.001$), higher specificity (95.5% vs. 91.8%, $P < 0.001$), and comparable AUC (0.86 vs. 0.87, $P = 0.127$) than FFDM. SM showed lower sensitivity than FFDM in asymptomatic breasts, dense breasts, ductal carcinoma in situ, T1, N0, and hormone receptor-positive/human epidermal growth factor receptor 2-negative cancers but showed higher specificity in non-cancerous dense breasts.

Conclusion: AI-CAD showed lower abnormality scores and reduced calibration performance for SM than for FFDM. Furthermore, the 10% preset threshold resulted in different discrimination performances for the SM. Given these limitations, off-label application of the current AI-CAD to SM should be avoided.

Keywords: Artificial intelligence-based computer-aided diagnosis; Synthetic mammography; Full-field digital mammography; Sensitivity; Specificity

INTRODUCTION

Digital breast tomosynthesis (DBT) provides three-dimensional images of the breast by acquiring multiple projections at different angles [1]. DBT helps reduce recall rates by reducing tissue superimposition [1]. Studies have

Received: July 12, 2024 **Revised:** December 10, 2024 **Accepted:** December 12, 2024
*These authors contributed equally to this work.
Corresponding author: Soo-Yeon Kim, MD, PhD, Department of Radiology, Korea University Guro Hospital, Korea University College of Medicine, 148 Gurodong-ro, Guro-gu, Seoul 08308, Republic of Korea
• E-mail: sooyeonkim41@gmail.com
Corresponding author: Ok Hee Woo, MD, PhD, Department of Radiology, Korea University Guro Hospital, Korea University College of Medicine, 148 Gurodong-ro, Guro-gu, Seoul 08308, Republic of Korea
• E-mail: wokhee@korea.ac.kr

demonstrated that combining DBT with full-field digital mammography (FFDM) leads to higher cancer detection rates and lower recall rates in screening populations than using FFDM alone [2-5]. Since its approval by the U.S. Food and Drug Administration in 2011, DBT use has dramatically increased globally. Initially, the "combo mode," utilizing both DBT and FFDM, was implemented; however, this approach doubled the radiation dose and increased scan time. Synthetic mammography (SM), a two-dimensional (2D) image similar to FFDM synthesized from DBT, has been introduced to address these issues. Previous studies have reported no significant differences in the diagnostic accuracy between DBT plus FFDM and DBT plus SM, and SM does not increase the radiation dose or scan time [6-8]. To date, only a few studies have directly compared FFDM and SM in limited clinical situations such as T1-stage breast cancer or microcalcifications [9,10]. These studies demonstrated that both modalities perform comparably in detecting T1-stage breast cancer [9] and microcalcifications [10]. However, SM may exhibit reduced resolution and conspicuity for asymmetries or small abnormalities compared to FFDM [11-14].

Recently, artificial intelligence-based computer-aided diagnosis (AI-CAD) has been increasingly integrated with mammography in clinical practice [15-19]. AI-CAD has improved diagnostic performance and reduced reading time compared with radiologists working alone [20-24]. The initial AI-CAD systems were developed using FFDM data [25]. Given the increased use of DBT with SM [26], evaluating whether FFDM-based AI-CAD can be applied to DBT with SM with satisfactory performance is important. Because FFDM and SM are 2D images with comparable accuracy [9,10], we hypothesized that FFDM-based AI-CAD could perform well on SM. Despite its importance, the research on this topic is limited [27]. Only one study by Lee et al. [28] addressed this issue. This study applied commercially available AI-CAD to both FFDM and SM in 192 patients with 203 breast cancers and compared the abnormality scores and areas under the receiver operating characteristics curve (AUCs) of AI-CAD across both modalities [28]. They found high agreement and comparable AUCs in the AI-CAD performance for FFDM and SM [28]. However, the study did not evaluate the overall sensitivity and specificity or assess the differences in these metrics based on mammographic and pathological characteristics. Therefore, further research is necessary to determine the applicability of FFDM-based AI CAD.

Therefore, our study aimed to test the performance of an AI-CAD designed for FFDM when applied to SM and to further evaluate its sensitivity and specificity according to mammographic and pathologic characteristics.

MATERIALS AND METHODS

This single-center retrospective study was approved by the Institutional Review Board of Korea University Guro Hospital (IRB No. 2024GR0487), and the requirement for written informed consent was waived.

Patient Selection

The inclusion criteria were patients with biopsy-proven breast cancer who underwent surgery at Korea University Guro Hospital between January 2019 and December 2021. During this period, our institution routinely performed both FFDM and DBT ("combo mode") as part of the preoperative workup for breast cancer patients. DBT was not performed in any setting except for preoperative evaluation. The exclusion criteria included patients who received neoadjuvant chemotherapy, those who underwent excisional or vacuum-assisted biopsy before mammography, those with prior breast surgery, and those with missing images from either FFDM or DBT.

Clinicopathologic Data Collection

We retrospectively collected clinicopathological data from electronic medical records. Clinical data included age, family history of breast cancer, menopausal status, and presence of cancer-related symptoms. Pathological data were based on surgical pathology reports, including T-stage, N-stage, and molecular subtype. Hormone receptor (HR) positivity was defined as estrogen or progesterone receptor positive with an Allred score of at least 3. Human epidermal growth factor receptor 2 (HER2) positivity was defined as an HER2 score of 3+ or gene amplification by fluorescence in situ hybridization in tumors with an HER2 score of 2+. Molecular subtypes were categorized as HR-positive/HER2-negative, HR-positive/HER2-positive, HR-negative/HER2-positive, or triple-negative.

Image Evaluation by Radiologists—Clinical Reading

FFDM and DBT were obtained simultaneously using the same equipment (Selenia Dimensions version 1.8.3.4; Hologic, Bedford, MA, USA) with craniocaudal (CC) and medio-lateral-oblique (MLO) views for each breast. Lee et al. [28] used Selenia Dimensions version 1.8.3.6, with no differences in DBT and SM images between the two versions. SM images were automatically reconstructed using DBT. Two radiologists, one with 20 and another with 11 years of

https://doi.org/10.3348/kjr.2024.0664

Korean Journal of Radiology
KJR

breast imaging experience, independently interpreted the mammograms in clinical practice. No consensus was reached. Breast density and mammographic features of known breast cancers were assessed using the Breast Imaging Reporting and Data System (BI-RADS) atlas. Clinical radiology reports were collected without retrospective review. The density was categorized as entirely fatty, scattered fibroglandular, heterogeneously dense, or extremely dense. The mammographic features were categorized as mass, asymmetry, architectural distortion, associated microcalcifications, microcalcifications alone, and occult (not observed).

AI-CAD Application

This study retrospectively applied a commercial AI-CAD (Lunit Inc., INSIGHT MMG, version 1.1.7.2, Seoul, Republic of Korea) to both FFDM and SM. This version of AI-CAD was developed using FFDM data and was validated in multinational studies [21]. We evaluated whether FFDM-based AI-CAD could effectively interpret SM. It calculates abnormality scores ranging from 0% to 100% for the CC and MLO views of each breast as approximate indicators of breast cancer probability [29]. The program was set up such that, for cases with scores exceeding 10%, the specific score and corresponding location were marked on the heatmap. Conversely, for cases with scores not exceeding 10%, the specific score and location were not marked on the heatmap. Instead, "low" is reported without any markings on the heatmap. The commercial version of the current AI-CAD system does not provide abnormal scores below the 10% threshold. However, we obtained abnormality scores from the vendor across all ranges regardless of the 10% threshold.

The AI-CAD generated abnormality scores ranging from 0% to 100% in the CC and MLO views of each breast. A higher score between the CC and MLO views was selected to represent each breast. With the cutoff threshold of this program set at 10%, abnormality scores lower than 10% were considered test-negative, whereas scores of 10% or greater were considered test-positive. This 10% threshold was independently selected in another study [21] and has been widely used to evaluate AI-CAD [17,18,28,30].

Statistical Analysis

Analyses were performed per breast. The median abnormality scores generated by AI-CAD were compared between the SM and FFDM in 517 breasts with cancer using the Wilcoxon signed-rank test. The calibration performance of the abnormality scores of FFDM and SM was evaluated using the calibration curves generated using the val.prob.ci.2 function from the rms package in R [31]. The discrimination performance using AUC, sensitivity, specificity, and accuracy using the 10% threshold were evaluated, and the results were compared between SM and FFDM. The sensitivity, specificity, and accuracy were compared using McNemar's test. AUCs

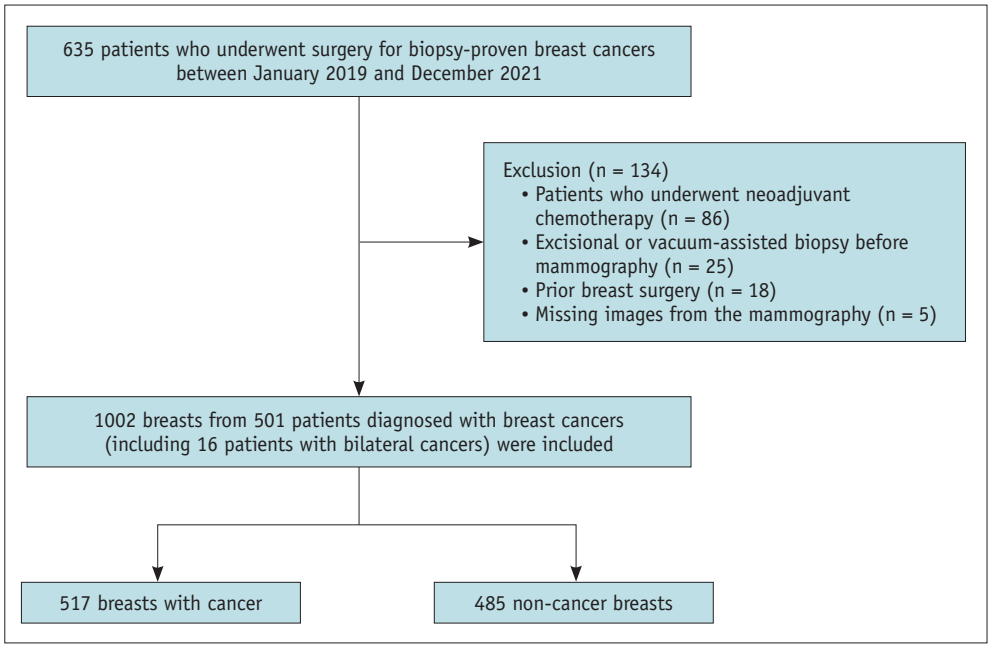

Fig. 1. Study flow chart.

https://doi.org/10.3348/kjr.2024.0664

were compared using DeLong's method. The sensitivity of AI-CAD was compared between SM and FFDM with the 10% threshold according to symptoms, mammographic density (nondense [defined as entirely fatty or scattered fibroglandular], dense [defined as heterogeneously dense or extremely dense]), and other features and pathological characteristics using the McNemar test. The specificity of AI-CAD was compared between SM and FFDM with a 10% threshold according to the mammographic density (nondense and dense) and the presence of biopsy-confirmed atypical lesions using McNemar's test. In addition to the preset 10% threshold, we identified the cutoff value that maximized the Youden index for the SM. Statistical analyses were performed using SPSS software (version 27; IBM Corp., Armonk, NY, USA), MedCalc software (version 20; MedCalc Software Ltd., Ostend, Belgium), and R version 4.4.1 (R Foundation for Statistical Computing, Vienna, Austria). $P < 0.05$ was considered significant.

RESULTS

Patient and Breast Characteristics

A total of 635 patients with biopsy-proven breast cancer underwent surgery at our institution between January 2019 and December 2021 (Fig. 1). Of these, 86 patients were excluded for receiving neoadjuvant chemotherapy, 25 for having undergone excisional or vacuum-assisted biopsies, 18 for having had prior breast surgery, and five due to missing images from either FFDM or DBT. Consequently, 501 women (mean age, 57 ± 11 years) were included in this study. Of these patients, 16 had bilateral breast cancer. Therefore, our analysis included data from 1002 breasts, 517 with and 485 without cancer. The clinicopathological and mammographic characteristics of the patients are shown in Table 1.

Abnormality Score of AI-CAD in Breasts With Cancer

The median SM score was significantly lower than FFDM in all 517 breasts with cancer (Fig. 2A, 71% for SM vs. 96% for FFDM, $P < 0.001$). SM exhibited significantly lower median scores than FFDM across subgroups categorized by symptom, mammographic, and pathological characteristics (Fig. 2B-H, all $P < 0.05$), except for asymmetry or architectural distortion with microcalcifications ($P = 0.082$) and T3-4 stage cancers ($P = 0.096$).

Calibration Performance

The calibration curve for SM demonstrated a greater

Table 1. Clinicopathologic and mammographic characteristics of patients

Parameters	Number
Patients (n = 501)	
Age, yr, mean ± standard deviation (ranges)	57 ± 11 (26–85)
Age, yrs (categorical)	
<50	139 (28)
≥50	362 (72)
Family history	
No	441 (88)
Yes	60 (12)
Menopausal status	
Pre	215 (43)
Post	286 (57)
Breasts with cancer (n = 517)	
Symptoms caused by cancer	
Absent	140 (27)
Present	377 (73)
Mammographic density	
Nondense	126 (24)
Dense	391 (76)
Mammographic features	
Mass with calcifications	99 (19)
Mass	315 (61)
Asymmetry or architectural distortion with calcifications	5 (1)
Asymmetry or architectural distortion	13 (3)
Calcifications only	43 (8)
Occult	42 (8)
T stage	
0 (DCIS)	62 (12)
1	268 (52)
2	173 (33)
3–4	14 (3)
Lymph node metastasis	
Absent	410 (79)
Present	107 (21)
Molecular subtype of invasive cancers	
HR+/HER2-	342 (75)
HR+/HER2+	69 (15)
HER2-enriched	15 (3)
Triple-negative	29 (6)
Breasts without cancer (n = 485)	
Mammographic density	
Nondense	122 (25)
Dense	363 (75)
Biopsy-confirmed lesions	
None	467 (96)
Benign*	11 (2)
Atypia†	7 (1)

Unless otherwise indicated, data are number of patients or breasts with percentages in parentheses.
*Four fibrocystic changes, three usual ductal hyperplasia, two adenosis, and two intraductal papillomas, †Four atypical ductal hyperplasia, two flat epithelial atypia, and one lobular carcinoma in situ.
DCIS = ductal carcinoma in situ, HR = hormone receptor, HER2 = human epidermal growth factor receptor 2

https://doi.org/10.3348/kjr.2024.0664

Korean Journal of Radiology
KJR

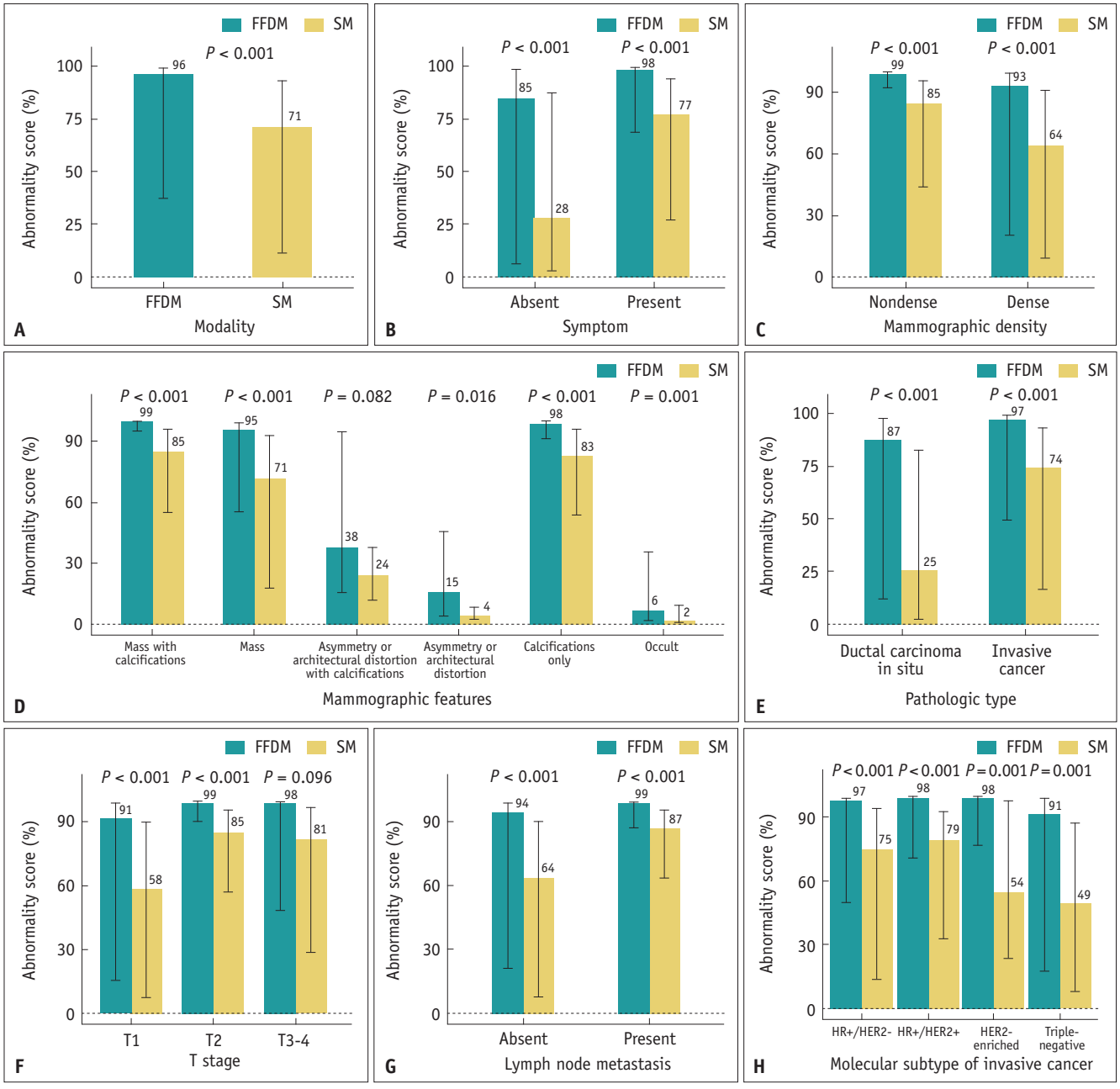

Fig. 2. Comparison of abnormality scores of artificial intelligence-based computer-aided diagnosis system between FFDM and SM in breasts with cancer, presented for all breasts and stratified by symptoms, mammographic, and pathologic characteristics. **A:** All breasts. **B:** Presence or absence of symptoms. **C:** Mammographic density. **D:** Mammographic features of cancer. **E:** Pathologic type. **F:** T stage. **G:** Lymph node metastasis. **H:** Molecular subtype of invasive cancer. The numbers on top of the bar graphs represent the group medians, and the error bars correspond to the 95% confidence intervals. FFDM = full-field digital mammography, SM = synthetic mammography, HR = hormone receptor, HER2 = human epidermal growth factor receptor 2

deviation from the ideal line than that for FFDM, indicating an inferior calibration performance (Fig. 3). The mean absolute error was higher for the SM than for the FFDM (0.053 for the SM and 0.029 for the FFDM).

Discrimination Performance of AI-CAD

Using the 10% preset threshold, AI-CAD applied to SM demonstrated lower sensitivity (Table 2; 76.2% for SM vs. 82.8% for FFDM, $P < 0.001$), higher specificity (95.5% for SM vs. 91.8% for FFDM, $P < 0.001$), and comparable accuracy (85.5% for SM vs. 87.1% for FFDM, $P = 0.121$) and

https://doi.org/10.3348/kjr.2024.0664

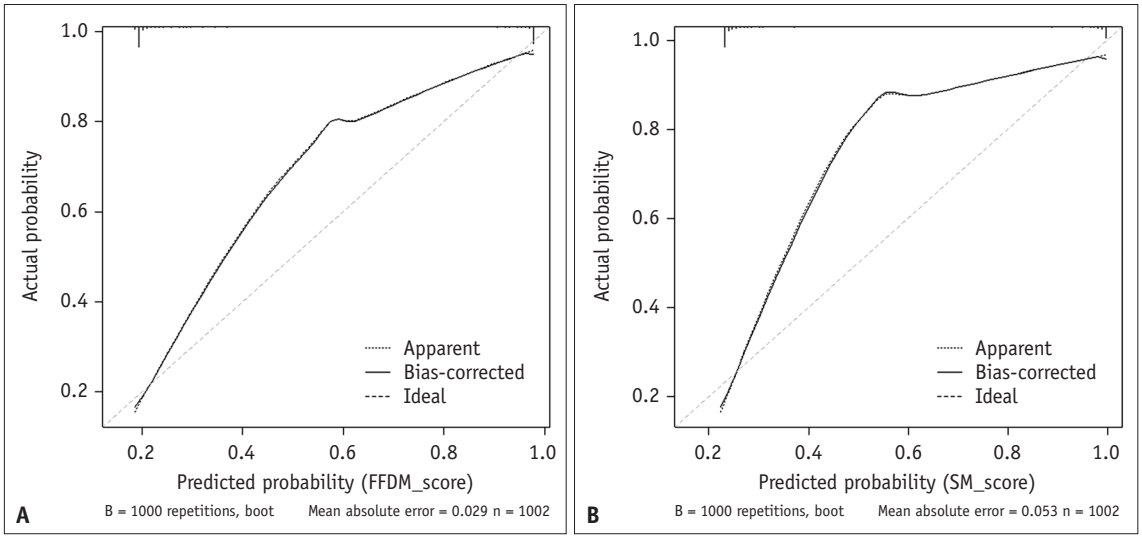

Fig. 3. Calibration curves of abnormality scores from artificial intelligence-based computer-aided diagnosis for **(A)** FFDM and **(B)** SM. A 45-degree ideal line (coarse dashed line) represents perfect calibration, where the predicted probabilities match the actual observed probabilities. FFDM = full-field digital mammography, SM = synthetic mammography

Table 2. Discrimination performance of artificial intelligence-based computer-aided diagnosis on FFDM and SM

	FFDM	SM	*P*
Sensitivity	82.8 (79.3, 85.9) [428/517]	76.2 (72.3, 79.8) [394/517]	<0.001
Specificity	91.8 (88.9, 94.0) [445/485]	95.5 (93.2, 97.1) [463/485]	<0.001
Accuracy	87.1 (84.9, 89.1) [873/1002]	85.5 (83.2, 87.7) [857/1002]	0.121
AUC	0.87 (0.85, 0.89)	0.86 (0.84, 0.88)	0.127

Data are percentages except for AUC; data in parentheses are 95% confidence intervals; data in brackets are numerators/denominators. Results other than AUC were obtained using 10% (preset by the vendor in the commercially available software) as the diagnostic threshold.
FFDM = full-field digital mammography, SM = synthetic mammography, AUC = area under the receiver operating characteristic curve

AUC (0.86 for SM vs. 0.87 for FFDM, $P = 0.127$) than FFDM.

AI-CAD showed a significantly lower sensitivity to SM than to FFDM in asymptomatic breasts (Table 3; 62.1% for SM vs. 72.9% for FFDM, $P = 0.041$) and dense breasts (74.2% for SM vs. 81.1% for FFDM, $P < 0.001$). Sensitivity of AI-CAD was significantly lower on SM compared to FFDM for occult cancers on mammography (Table 3, 21.4% for SM vs. 45.2% for FFDM, $P = 0.021$), ductal carcinoma in situ (61.3% for SM vs. 77.4% for FFDM, $P = 0.006$), T1 cancers (71.6% for SM vs. 79.1% for FFDM, $P = 0.005$), node-negative cancers (72.2% for SM vs. 81.2% for FFDM, $P < 0.001$), and HR+/HER2- subtype (77.5% for SM vs. 83.9% for FFDM, $P = 0.004$). Figures 4 and 5 demonstrate representative cases with mass and asymmetry, respectively, showing decreased sensitivity on SM in patients with early-stage HR+/HER2- breast cancer.

The specificity was significantly higher in dense breasts for SM than for FFDM (Table 4, 96.1% for SM vs. 91.5% for FFDM, $P < 0.001$). Similarly, the specificity was significantly higher in breasts without biopsy-confirmed atypical lesions on SM than on FFDM (Table 4; 95.4% for SM vs. 92.0% for FFDM, $P = 0.031$). The specificity of SM and FFDM was comparable in nondense breasts (92.6% for SM vs. 93.4% for FFDM, $P > 0.999$) and in breasts with atypia (71.4% for SM vs. 100% for FFDM, $P = 0.141$). A representative case demonstrating the higher specificity of AI-CAD for SM compared with FFDM is presented in Figure 6.

The threshold maximizing the Youden index was 4% for SM (the apparent within-sample performance results are provided in Supplementary Table 1).

DISCUSSION

Our study evaluated the feasibility of applying FFDM-based AI-CAD to SM, which was obtained from DBT. The abnormality score and performance of AI-CAD were compared between FFDM and SM in 501 breast cancer patients, comprising 517 breasts with cancer and 485 breasts without cancer. AI-CAD demonstrated significantly lower median abnormality scores and poorer calibration performance on SM than on FFDM, suggesting that, at minimum, threshold adjustments

https://doi.org/10.3348/kjr.2024.0664

Korean Journal of Radiology
KJR

Table 3. Sensitivity of artificial intelligence-based computer-aided diagnosis according to symptoms, mammographic and pathologic characteristics

	FFDM	SM	*P*
Symptoms caused by cancer			
Absent	72.9 (102/140)	62.1 (87/140)	0.041
Present	86.5 (326/377)	81.4 (307/377)	0.067
Mammographic density			
Nondense	88.1 (111/126)	82.5 (104/126)	0.065
Dense	81.1 (317/391)	74.2 (290/391)	<0.001
Mammographic features			
Mass with calcifications	90.9 (90/99)	86.9 (86/99)	0.344
Mass	84.4 (266/315)	80.3 (253/315)	0.055
Asymmetry or architectural distortion with calcifications	100 (5/5)	80 (4/5)	>0.999
Asymmetry or architectural distortion	53.8 (7/13)	23.1 (3/13)	0.125
Calcifications only	95.3 (41/43)	90.7 (39/43)	0.500
Occult	45.2 (19/42)	21.4 (9/42)	0.021
Pathologic type			
DCIS	77.4 (48/62)	61.3 (38/62)	0.006
Invasive cancer	83.5 (380/455)	78.2 (356/455)	0.003
T stage			
T1	79.1 (212/268)	71.6 (192/268)	0.005
T2	90.8 (157/173)	87.9 (152/173)	0.227
T3-4	78.6 (11/14)	85.7 (12/14)	>0.999
Lymph node metastasis			
Absent	81.2 (333/410)	72.2 (296/410)	<0.001
Present	88.8 (95/107)	91.6 (98/107)	0.508
Molecular subtype of invasive cancers			
HR+/HER2-	83.9 (287/342)	77.5 (265/342)	0.004
HR+/HER2+	84.1 (58/69)	82.6 (57/69)	>0.999
HER2-enriched	86.7 (13/15)	86.7 (13/15)	>0.999
Triple-negative	75.9 (22/29)	72.4 (21/29)	>0.999

Data are percentages, with numerators/denominators in parentheses. Results were obtained using 10% (preset by the vendor in the commercially available software) as the diagnostic threshold.
FFDM = full-field digital mammography, SM = synthetic mammography, DCIS = ductal carcinoma in situ, HR = hormone receptor, HER2 = human epidermal growth factor receptor 2

may be necessary when applying AI-CAD to SM. With the preset 10% threshold, AI-CAD showed lower sensitivity but higher specificity and a comparable AUC for SM compared with FFDM. Using adjusted thresholds, such as the Youden index-based 4% for SM, AI-CAD may demonstrate comparable sensitivity, specificity, and AUC to FFDM (Supplementary Table 1). However, it should be noted that abnormality scores below 10% are not routinely available in the commercial version of AI-CAD. Thus, applying the current version of AI-CAD to SM seems infeasible. Furthermore, the Youden index-based threshold was not externally verified. Therefore, the off-label use of AI-CAD in SM should be avoided.

Another study addressed a similar issue [28]. Lee et al. [28] evaluated the applicability of the same AI-CAD software for SM in 192 breast cancer patients, comprising 203 breasts with cancer and 181 breasts without cancer. Contrary to our results, AI-CAD showed significantly higher abnormality scores for SM than for FFDM in breast cancer patients (90.6% for SM vs. 84.8% for FFDM, $P < 0.001$) [28]. The reason for the disparity in abnormality scores between the two studies was difficult to determine because both studies utilized the same brand of AI-CAD software and a similar version of DBT that did not affect the image quality, and they showed comparable baseline clinical and pathological characteristics, except for the mammographic features of the cancers. In contrast to our study, the previous study [28] retrospectively evaluated mammographic features, including the presence or absence of distortion associated with a

 https://doi.org/10.3348/kjr.2024.0664

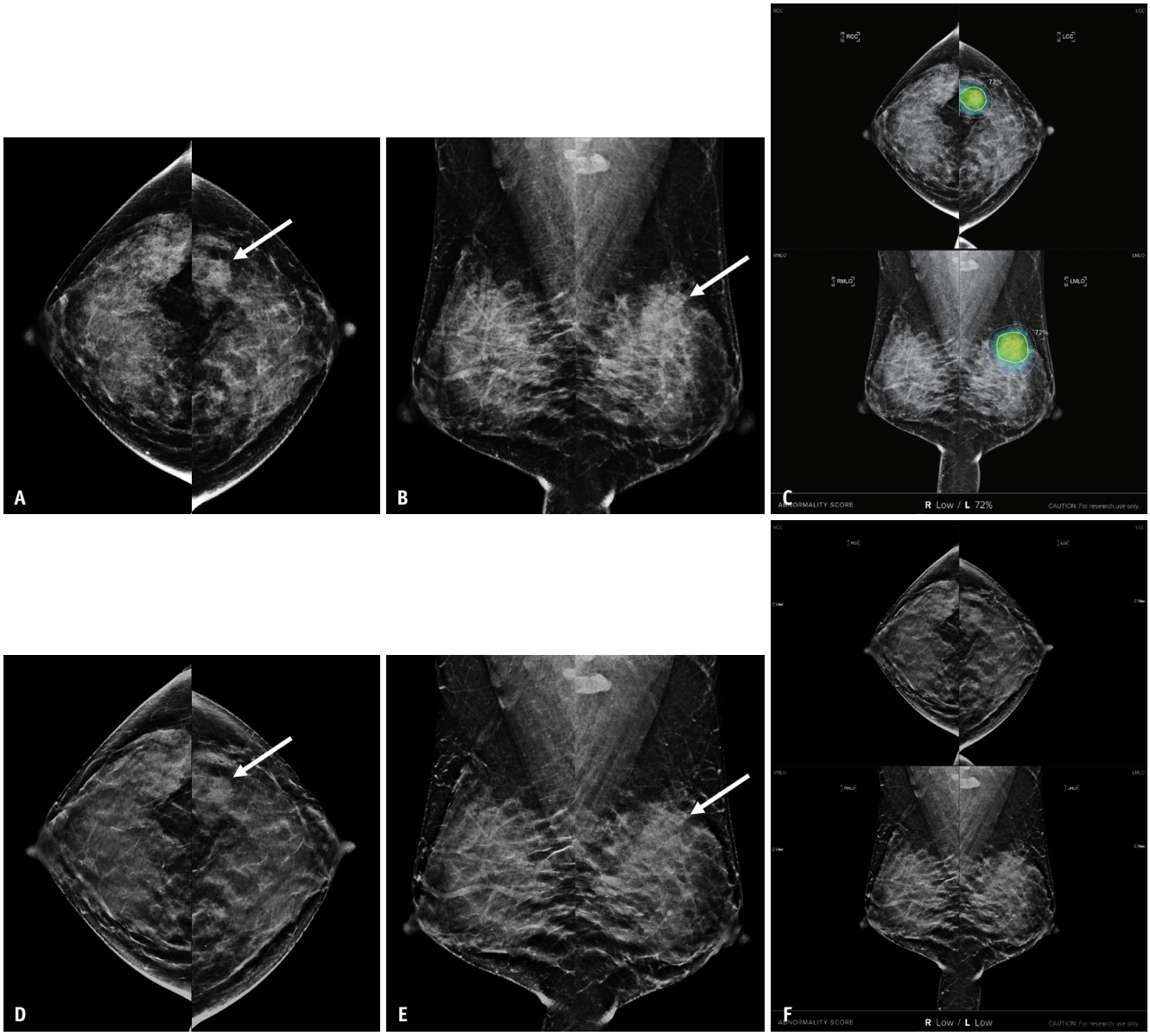

Fig. 4. Images demonstrating decreased abnormality scores and sensitivity of AI-CAD on SM compared to FFDM for a T1, N0 invasive ductal carcinoma with hormone receptor-positive/human epidermal growth factor receptor 2-negative subtype in a 43-year-old woman. **A-F:** The CC **(A)** and MLO **(B)** views of FFDM show a mass (arrows) in the left upper outer breast. AI-CAD **(C)** output screenshot from FFDM detects the mass with abnormality scores of 72% on both CC and MLO views. The CC **(D)** and MLO **(E)** views of SM show a less evident mass (arrows) with abnormality scores of 9% and 6%, respectively. AI-CAD **(F)** output screenshot from SM does not mark the cancer because the scores are below the cutoff of 10%. We obtained all abnormality scores for each view, regardless of the 10% cutoff, directly from the vendor for research purposes. AI-CAD = artificial intelligence-based computer-aided diagnosis, SM = synthetic mammography, FFDM = full-field digital mammography, CC = craniocaudal, MLO = mediolateral oblique

mass. These contrasting results may indicate the limited generalizability of AI-CAD in SM data, further reinforcing the recommendation against its off-label use in SM.

In our study, SM showed significantly lower median abnormality scores than FFDM across all patients and most subgroups categorized by mammographic and pathological features, except for cases with asymmetry or architectural distortion with microcalcifications (n = 5) and T3-4 stage cancers (n = 14), likely because of the small sample sizes in these groups.

Using a 10% preset threshold, SM exhibited significantly lower sensitivity than FFDM in asymptomatic breasts, dense breasts, occult lesions, DCIS, and T1, N0, and HR-positive/HER2-negative breast cancers. In contrast, SM showed a

https://doi.org/10.3348/kjr.2024.0664

Korean Journal of Radiology
KJR

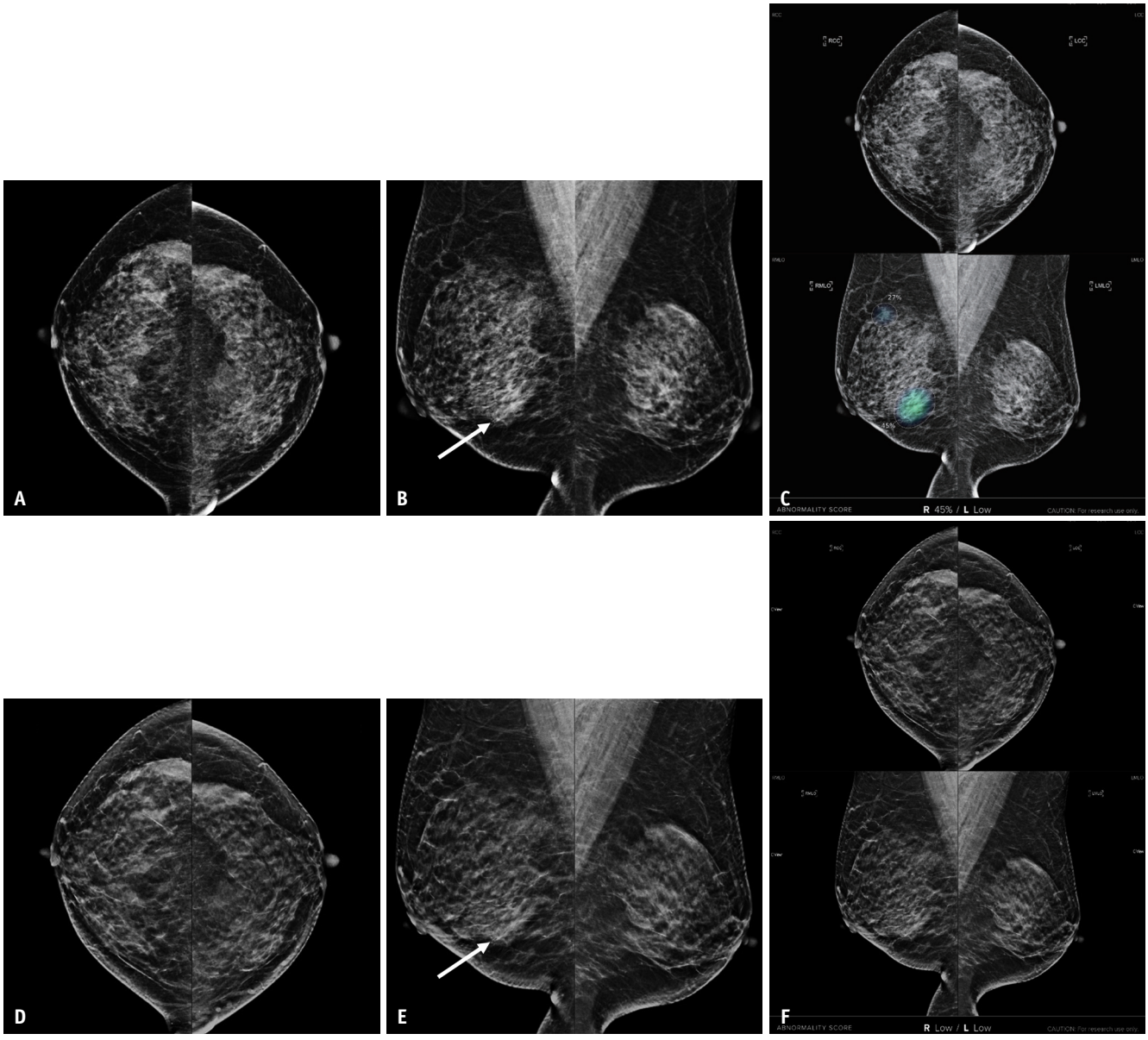

Fig. 5. Images demonstrating decreased abnormality scores and sensitivity of AI-CAD on SM compared to FFDM for a T1, N0 invasive ductal carcinoma with hormone receptor-positive/human epidermal growth factor receptor 2-negative subtype in a 52-year-old woman. **A-F:** The CC **(A)** view of FFDM does not clearly show cancer, but the MLO **(B)** view shows a subtle asymmetry (arrow), which is a biopsy-proven cancer at the 6 o'clock position of the right breast. AI-CAD **(C)** output screenshot from FFDM correctly identifies the cancer on the MLO view with an abnormality score of 45% but misses the cancer on the CC view with an abnormality score of 8%. The 27% marking on the MLO view was a false-positive marking. The CC **(D)** view of SM also does not clearly show cancer with an abnormality score of 3%. The MLO **(E)** view of SM shows a slightly more subtle asymmetry (arrow) compared to FFDM, with an abnormality score of 5%. AI-CAD **(F)** output screenshot from SM does not mark the cancer because the scores are below the cutoff of 10%. We obtained all abnormality scores for each view, regardless of the 10% cutoff, directly from the vendor for research purposes. AI-CAD = artificial intelligence-based computer-aided diagnosis, SM = synthetic mammography, FFDM = full-field digital mammography, CC = craniocaudal, MLO = mediolateral oblique

significantly higher specificity than FFDM, particularly in dense breasts and those without biopsy-confirmed atypical lesions. The higher specificity of SM can be attributed to its enhanced resolution and improved identification of normal superimposed breast tissue on DBT, from which SM images are synthesized [2-5].

The AI-CAD system used in this study was trained exclusively on FFDM data without incorporating SM or DBT data during its development. Our findings indicate that applying FFDM-trained AI-CAD directly to SM images should be avoided, as it compromises the diagnostic performance. To address this limitation, developing a new AI-CAD system

 https://doi.org/10.3348/kjr.2024.0664

Table 4. Specificity of artificial intelligence-based computer-aided diagnosis according to mammographic density and presence of biopsy-confirmed atypical lesions

	FFDM	SM	*P*
Mammographic density			
Nondense	92.6 (113/122)	93.4 (114/122)	>0.999
Dense	91.5 (332/363)	96.1 (349/363)	<0.001
Atypia			
Absence	92.0 (440/478)	95.4 (456/478)	0.031
Presence	71.4 (5/7)	100 (7/7)	0.141

Data are percentages with numerators/denominators in parentheses. Results were obtained using 10% (preset by the vendor in the commercially available software) as the diagnostic threshold. FFDM = full-field digital mammography, SM = synthetic mammography

specifically trained on SM data or fine-tuning the existing FFDM-based AI-CAD system using SM images would be beneficial. Recent advancements have demonstrated success in fine-tuning FFDM-based AI CAD systems using DBT image stacks. For instance, a system initially trained on 60000 FFDM cases was later fine-tuned using 12810 DBT image stacks, enabling section-wise likelihood assessment [32]. This fine-tuned AI-CAD demonstrated a high diagnostic performance as both a standalone tool and an adjunct to radiologists [32]. However, similar fine-tuning with SM data has yet to be explored, underscoring the need for further research.

Our study had several limitations. First, a selection bias may have occurred because of the single-center retrospective design. Second, we utilized a single AI-CAD system. This limits the generalizability of our results to AI-CAD systems from different vendors or versions. Third, patients who underwent neoadjuvant chemotherapy were excluded, leading to a small number of patients with locally advanced, HER2-positive, or triple-negative breast cancers in our cohort. Fourth, we evaluated the standalone performance of AI-CAD applied to SM and FFDM without considering DBT images or how radiologists using this program interpreted mammography in clinical practice. Fifth, our study utilized a cancer-enriched cohort, as the indication for DBT at our institution was a preoperative workup. Therefore, the performance of AI-CAD in other settings, such as screening, may differ from our results. Sixth, mammographic features were analyzed based on clinical radiology reports without retrospective reevaluation, which may have introduced bias due to interobserver variability.

In conclusion, AI-CAD showed lower abnormality scores and reduced calibration performance for SM than for FFDM. Furthermore, the 10% preset threshold resulted in different discrimination performances for SM compared to FFDM. Given these limitations, off-label application of the current AI-CAD to SM should be avoided.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2024.0664.

Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Conflicts of Interest

Sung Eun Song and Min Sun Bae, who hold respective positions as the Editorial Board Members of the *Korean Journal of Radiology*, were not involved in the editorial evaluation or decision to publish this article. The remaining authors have declared no conflicts of interest.

Author Contributions

Conceptualization: Ok Hee Woo. Data curation: Kyung Eun Lee, Ok Hee Woo. Formal analysis: Kyung Eun Lee, Soo-Yeon Kim, Ok Hee Woo. Funding acquisition: Ok Hee Woo. Investigation: Kyung Eun Lee, Soo-Yeon Kim, Ok Hee Woo. Methodology: Kyung Eun Lee, Soo-Yeon Kim, Ok Hee Woo. Project administration: Ok Hee Woo. Resources: Kyung Eun Lee, Ok Hee Woo. Software: Kyung Eun Lee, Ok Hee Woo. Supervision: Soo-Yeon Kim, Ok Hee Woo. Validation: Kyung Eun Lee, Soo-Yeon Kim, Ok Hee Woo. Visualization: Kyung Eun Lee, Soo-Yeon Kim. Writing—original draft: Kyung Eun Lee, Soo-Yeon Kim. Writing—review & editing: all authors.

ORCID IDs

Kyung Eun Lee
https://orcid.org/0000-0002-0352-8228
Sung Eun Song
https://orcid.org/0000-0002-9259-8294
Kyu Ran Cho
https://orcid.org/0000-0002-8936-6468
Min Sun Bae
https://orcid.org/0000-0002-1971-3428
Bo Kyoung Seo
https://orcid.org/0000-0002-9512-5361
Soo-Yeon Kim
https://orcid.org/0000-0001-8915-3924

Korean Journal of Radiology
KJR

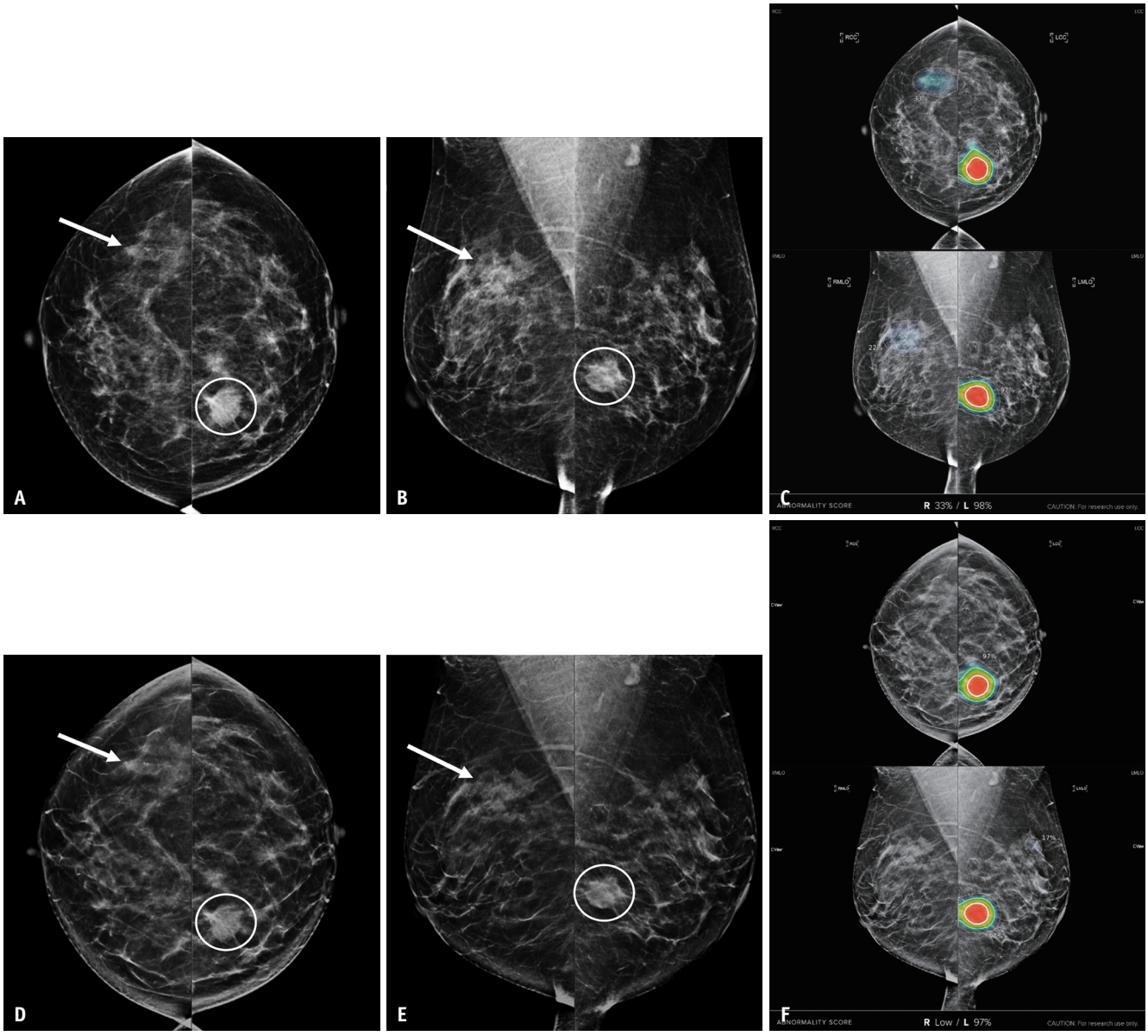

Fig. 6. Images demonstrating higher specificity of AI-CAD on SM compared to FFDM in a 54-year-old woman with left breast cancer and right non-cancerous breast. **A-F:** The CC **(A)** and MLO **(B)** views of FFDM show a mass with microcalcifications (circles) in the left lower inner breast, confirmed as cancer and an incidental asymmetry (arrows) in the right upper outer breast. AI-CAD **(C)** output screenshot from FFDM marks the asymmetry, with abnormality scores of 33% and 22% on the CC and MLO views, respectively. The CC **(D)** and MLO **(E)** views of SM show a less evident asymmetry (arrows), with abnormality scores of 0.62% and 0.46%, respectively. AI-CAD **(F)** output screenshot from SM does not mark the asymmetry in the right breast because the scores are below the cutoff of 10%. We obtained all abnormality scores for each view, regardless of the 10% cutoff, directly from the vendor for research purposes. AI-CAD = artificial intelligence-based computer-aided diagnosis, SM = synthetic mammography, FFDM = full-field digital mammography, CC = craniocaudal, MLO = mediolateral oblique

Ok Hee Woo
https://orcid.org/0000-0003-3953-933X

Funding Statement

This study was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR22C1302).

REFERENCES

1. Vedantham S, Karellas A, Vijayaraghavan GR, Kopans DB. Digital breast tomosynthesis: state of the art. *Radiology*

2015;277:663-684
2. Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013;14:583-589
3. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013;267:47-56
4. Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;311:2499-2507
5. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology* 2013;269:694-700
6. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS. Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. *Radiology* 2017;283:70-76
7. Alabousi M, Wadera A, Kashif Al-Ghita M, Kashef Al-Ghetaa R, Salameh JP, Pozdnyakov A, et al. Performance of digital breast tomosynthesis, synthetic mammography, and digital mammography in breast cancer screening: a systematic review and meta-analysis. *J Natl Cancer Inst* 2021;113:680-690
8. Heywang-Köbrunner SH, Jänsch A, Hacker A, Weinand S, Vogelmann T. Digital breast tomosynthesis (DBT) plus synthesised two-dimensional mammography (s2D) in breast cancer screening is associated with higher cancer detection and lower recalls compared to digital mammography (DM) alone: results of a systematic review and meta-analysis. *Eur Radiol* 2022;32:2301-2312
9. Choi JS, Han BK, Ko EY, Ko ES, Hahn SY, Shin JH, et al. Comparison between two-dimensional synthetic mammography reconstructed from digital breast tomosynthesis and full-field digital mammography for the detection of T1 breast cancer. *Eur Radiol* 2016;26:2538-2546
10. Dodelzon K, Simon K, Dou E, Levy AD, Michaels AY, Askin G, et al. Performance of 2D synthetic mammography versus digital mammography in the detection of microcalcifications at screening. *AJR Am J Roentgenol* 2020;214:1436-1444
11. Ratanaprasatporn L, Chikarmane SA, Giess CS. Strengths and weaknesses of synthetic mammography in screening. *Radiographics* 2017;37:1913-1927
12. Giess CS, Raza S, Denison CM, Yeh ED, Gombos EC, Frost EP, et al. Lesion conspicuity on synthetic screening mammography compared to full field digital screening mammography. *Clin Imaging* 2021;75:90-96
13. Chikarmane S. Synthetic mammography: review of benefits and drawbacks in clinical use. *J Breast Imaging* 2022;4:124-134
14. Chikarmane SA, Offit LR, Giess CS. Synthetic mammography: benefits, drawbacks, and pitfalls. *Radiographics* 2023;43:e230018
15. Geras KJ, Mann RM, Moy L. Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. *Radiology* 2019;293:246-259
16. Lamb LR, Lehman CD, Gastounioti A, Conant EF, Bahl M. Artificial intelligence (AI) for screening mammography, from the AJR special series on AI applications. *AJR Am J Roentgenol* 2022;219:369-380
17. Kim YS, Jang MJ, Lee SH, Kim SY, Ha SM, Kwon BR, et al. Use of artificial intelligence for reducing unnecessary recalls at screening mammography: a simulation study. *Korean J Radiol* 2022;23:1241-1250
18. Yoen H, Jang MJ, Yi A, Moon WK, Chang JM. Artificial intelligence for breast cancer detection on mammography: factors related to cancer detection. *Acad Radiol* 2024;31:2239-2247
19. Yoon JH, Kim EK. Deep learning-based artificial intelligence for mammography. *Korean J Radiol* 2021;22:1225-1239
20. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology* 2019;290:305-314
21. Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health* 2020;2:e138-e148
22. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw Open* 2020;3:e200265
23. Dembrower K, Wåhlin E, Liu Y, Salim M, Smith K, Lindholm P, 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:e468-e474
24. Lee JH, Kim KH, Lee EH, Ahn JS, Ryu JK, Park YM, et al. Improving the performance of radiologists using artificial intelligence-based detection support software for mammography: a multi-reader study. *Korean J Radiol* 2022;23:505-516
25. Goldberg JE, Reig B, Lewin AA, Gao Y, Heacock L, Heller SL, et al. New horizons: artificial intelligence for digital breast tomosynthesis. *Radiographics* 2023;43:e220060
26. Alsheik NH, Dabbous F, Pohlman SK, Troeger KM, Gliklich RE, Donadio GM, et al. Comparison of resource utilization and clinical outcomes following screening with digital breast tomosynthesis versus digital mammography: findings from a learning health system. *Acad Radiol* 2019;26:597-605
27. Yoon JH, Strand F, Baltzer PAT, Conant EF, Gilbert FJ, Lehman CD, 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:e222639
28. Lee SE, Han K, Kim EK. Application of artificial intelligence-based computer-assisted diagnosis on synthetic mammograms from breast tomosynthesis: comparison with digital

https://doi.org/10.3348/kjr.2024.0664

mammograms. *Eur Radiol* 2021;31:6929-6937
29. Park SH, Hwang EJ. Caveats in using abnormality/probability scores from artificial intelligence algorithms: neither true probability nor level of trustworthiness. *Korean J Radiol* 2024;25:328-330
30. Lee SE, Hong H, Kim EK. Positive predictive values of abnormality scores from a commercial artificial intelligence-based computer-aided diagnosis for mammography. *Korean J Radiol* 2024;25:343-350
31. R Core Team. val.prob.ci.2 {CalibrationCurves} [accessed on December 9, 2024]. Available at: https://search.r-project.org/CRAN/refmans/CalibrationCurves/html/val.prob.ci.2.html
32. Park EK, Kwak S, Lee W, Choi JS, Kooi T, Kim EK. Impact of AI for digital breast tomosynthesis on breast cancer detection and interpretation time. *Radiol Artif Intell* 2024;6:e230318

 https://doi.org/10.3348/kjr.2024.0664