Original Article | Breast Imaging

eISSN 2005-8330
https://doi.org/10.3348/kjr.2023.0528
Korean J Radiol 2024;25(1):11-23

Korean Journal of Radiology
KJR

Unenhanced Breast MRI With Diffusion-Weighted Imaging for Breast Cancer Detection: Effects of Training on Performance and Agreement of Subspecialty Radiologists

Yeon Soo Kim[1,2], Su Hyun Lee[1,2], Soo-Yeon Kim[1,2], Eun Sil Kim[1,2], Ah Reum Park[1,2], Jung Min Chang[1,2], Vivian Youngjean Park[3], Jung Hyun Yoon[3], Bong Joo Kang[4], Bo La Yun[2,5], Tae Hee Kim[6], Eun Sook Ko[7], A Jung Chu[8], Jin You Kim[9], Inyoung Youn[10], Eun Young Chae[11], Woo Jung Choi[11], Hee Jeong Kim[11], Soo Hee Kang[12], Su Min Ha[1,2,13], Woo Kyung Moon[1,2,13]

[1]Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea
[2]Department of Radiology, Seoul National University College of Medicine, Seoul, Republic of Korea
[3]Department of Radiology, Severance Hospital, Research Institute of Radiological Science, Yonsei University College of Medicine, Seoul, Republic of Korea
[4]Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[5]Department of Radiology, Seoul National University Bundang Hospital, Seoul, Republic of Korea
[6]Department of Radiology, Ajou University Medical Center, Suwon, Republic of Korea
[7]Department of Radiology and Center for Imaging Science, Samsung Medical Center, Seoul, Republic of Korea
[8]Department of Radiology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Republic of Korea
[9]Department of Radiology, Medical Research Institute, Pusan National University Hospital, Pusan National University School of Medicine, Busan, Republic of Korea
[10]Department of Radiology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[11]Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[12]Medical Research Collaborating Center, Seoul National University Hospital, Seoul, Republic of Korea
[13]Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Republic of Korea

Objective: To investigate whether reader training improves the performance and agreement of radiologists in interpreting unenhanced breast magnetic resonance imaging (MRI) scans using diffusion-weighted imaging (DWI).

Materials and Methods: A study of 96 breasts (35 cancers, 24 benign, and 37 negative) in 48 asymptomatic women was performed between June 2019 and October 2020. High-resolution DWI with b-values of 0, 800, and 1200 sec/mm^2 was performed using a 3.0-T system. Sixteen breast radiologists independently reviewed the DWI, apparent diffusion coefficient maps, and T1-weighted MRI scans and recorded the Breast Imaging Reporting and Data System (BI-RADS) category for each breast. After a 2-h training session and a 5-month washout period, they re-evaluated the BI-RADS categories. A BI-RADS category of 4 (lesions with at least two suspicious criteria) or 5 (more than two suspicious criteria) was considered positive. The per-breast diagnostic performance of each reader was compared between the first and second reviews. Inter-reader agreement was evaluated using a multi-rater κ analysis and intraclass correlation coefficient (ICC).

Results: Before training, the mean sensitivity, specificity, and accuracy of the 16 readers were 70.7% (95% confidence interval [CI]: 59.4–79.9), 90.8% (95% CI: 85.6–94.2), and 83.5% (95% CI: 78.6–87.4), respectively. After training, significant improvements in specificity (95.2%; 95% CI: 90.8–97.5; P = 0.001) and accuracy (85.9%; 95% CI: 80.9–89.8; P = 0.01) were observed, but no difference in sensitivity (69.8%; 95% CI: 58.1–79.4; P = 0.58) was observed. Regarding inter-

Received: June 4, 2023 **Revised:** October 26, 2023 **Accepted:** October 30, 2023
Corresponding author: Su Min Ha, MD, PhD, Department of Radiology, Seoul National University Hospital, Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul, 03080, Republic of Korea
• E-mail: su3ha5@hanmail.net

reader agreement, the κ values were 0.57 (95% CI: 0.52–0.63) before training and 0.68 (95% CI: 0.62–0.74) after training, with a difference of 0.11 (95% CI: 0.02–0.18; P = 0.01). The ICC was 0.73 (95% CI: 0.69–0.74) before training and 0.79 (95% CI: 0.76–0.80) after training (P = 0.002).
Conclusion: Brief reader training improved the performance and agreement of interpretations by breast radiologists using unenhanced MRI with DWI.
Keywords: Diffusion-weighted imaging; Magnetic resonance imaging; Training; Diagnosis, Differential; Breast

INTRODUCTION

Diffusion-weighted imaging (DWI) is a fast, unenhanced functional magnetic resonance imaging (MRI) technique that reflects the water diffusion properties in tissues [1-4]. Breast cancers appear hyperintense on DWI, with high b-values and low apparent diffusion coefficient (ADC) values on corresponding ADC maps [5]. There is growing interest in DWI as an alternative to dynamic contrast-enhanced (DCE) MRI because it minimizes costs, reduces examination times, improves patient compliance with MRI, and avoids gadolinium retention, especially in patients with intermediate- or high-risk breast cancer who undergo repeated MRI [6].

Multiple retrospective studies using designs that simulate the clinical screening setting have explored the use of DWI as a standalone tool and have demonstrated its usefulness for detecting small breast cancers in women at intermediate-to-high risk, with a pooled sensitivity of 76% and specificity of 89% [7-15]. The readers in these studies assessed only unenhanced MRI sequences (i.e., DWI with or without anatomically unenhanced T1- or T2-weighted sequences). Based on these encouraging results, a prospective clinical trial involving DWI in women at a high risk of breast cancer is being conducted [16]. An expert consensus of the European Society of Breast Radiology (EUSOBI) [17] recommended standardized parameters for high-quality breast DWI that were specifically applied to the unenhanced DWI screening protocol [16]. High-resolution acquisition strategies for DWI sequences can improve lesion conspicuity and produce sharper images, thereby enabling better assessment.

Standardized DWI acquisition and interpretation are of the utmost importance in promoting DWI as an effective screening tool. To train radiologists and evaluate the reliability of the assessment used in a prospective clinical trial [16], a DWI interpretation algorithm using terms adapted from the Breast Imaging Reporting and Data System (BI-RADS) lexicon and the suggested diffusion-level lexicon of EUSOBI was proposed by the Diffusion-Weighted MRI Screening Trial (DWIST) group (Fig. 1) [5,15]. However, to our knowledge, the effects of training on the performance and agreement of radiologists during the final assessment have not been reported. In a multi-reader study that assessed the inter-reader reproducibility of DWI interpretation for breast cancer detection, good agreement was found only among expert readers, and the authors suggested that a learning curve of at least 3 years of experience was required to appropriately interpret breast MRI based on DWI [18].

We hypothesized that training involving unenhanced MRI with DWI using a standardized interpretation algorithm would improve the diagnostic performance and inter-reader agreement. Therefore, this study assessed the diagnostic performance and inter-reader agreement of radiologists interpreting unenhanced MRI with DWI, using a standardized algorithm for breast cancer detection, before and after reader training.

MATERIALS AND METHODS

MRI Selection

This retrospective study was approved by Institutional Review Board of each participating institution, and the requirement for informed consent was waived. From June 2019 to October 2020, four breast radiologists with 2–30 years of experience with breast MRI (W.K.M., S.H.L., S.M.H., and Y.S.K.) retrospectively reviewed the DWI scans; however, they were not blinded to the clinical history and DCE-MRI results. We selected this study period because Seoul National University Hospital first applied an optimized 3.0-T DWI protocol comprising three b-values with both single-shot and multi-shot echo-planar imaging in June 2019 [7].

To include cases from different clinical scenarios, we searched for breast MRI examinations performed for breast cancer staging and high-risk screening, patients with a personal history of breast cancer, and patients with equivocal findings during conventional breast imaging. We

Korean Journal of Radiology
KJR

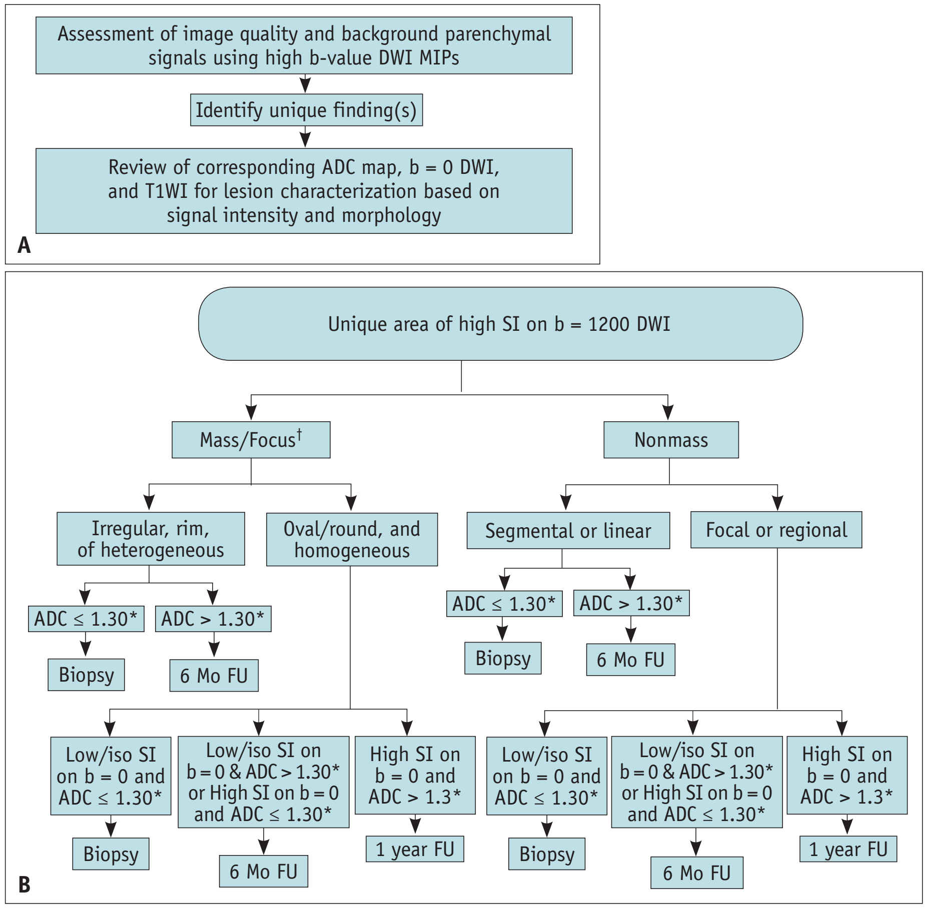

Fig. 1. Interpretation guideline for unenhanced breast MRI with diffusion-weighted imaging. Approach to unenhanced breast MRI **(A)** and interpretation algorithms for unique findings on diffusion-weighted imaging (DWI) **(B)**. Reprinted under a CC BY NC license, from Lee et al. [5], *Korean J Radiol* 2021;22:9-22. *ADC is evaluated at b = 800 mm^2/sec DWI, †Focus is evaluated based on both signal intensity on b = 0 mm^2/sec and ADC value. MRI = magnetic resonance imaging, MIP = maximum intensity projection, ADC = apparent diffusion coefficient, T1WI = T1-weighted imaging, SI = signal intensity, FU = follow-up, Mo = months

excluded patients with poor-quality DWI acquisition and those lacking a standard reference. Our reference standard was based on the pathology results of the core or surgical biopsy and included a follow-up of at least 1 year. We also developed a DWI interpretative task.

We selected 96 breasts from 48 asymptomatic patients with a mean age of 55.5 years (standard deviation, 10.3 years; range, 38–81 years) for the analysis. Malignancies were present in 35 of the 96 breasts (37%), including ductal carcinoma in situ (n = 6, 17%), invasive ductal carcinoma (n = 24, 69%), invasive lobular carcinoma (n = 2, 6%), metaplastic carcinoma (n = 1, 3%), adenoid cystic carcinoma (n = 1, 3%), and microinvasive carcinoma (n = 1, 3%). The median pathological tumor size of invasive cancers was 1.1 cm (interquartile range, 0–4.0 cm). Benign lesions (n = 24; 7 with biopsy results and 17 with no pathological results but validated by a follow-up) and negative cases (n = 37) from a total of 61 breasts showed stability during subsequent clinical and imaging follow-up examinations. Among lesions with pathological results, 34 lesions (all malignant) were biopsied before MRI, while eight (one malignant and seven benign) lesions were biopsied after MRI. The patient and tumor characteristics are summarized in Table 1.

The BI-RADS assessment and lesion types identified using MRI with DWI according to the four unblinded expert consensuses and majority rule are listed in Table 2. Malignancies were observed in 25 (96%) of the 26 lesions with BI-RADS category 4 and 10 (100%) of the 10 lesions with BI-RADS category 5. Malignancies were observed in 28 (76%) of the 37 masses, six (75%) of the eight nonmass lesions, and one (13%) of the eight foci.

MRI Protocol

Breast MRI was performed in the prone position using 3.0-

 https://doi.org/10.3348/kjr.2023.0528

Table 1. Patients and lesion characteristics

Characteristics	Values
Age, yr	55.5 ± 10.3 (38–81)
MRI indications (n = 48)	
Preoperative staging of breast cancer	35 (73)
Surveillance after breast cancer treatment	8 (17)
High-risk screening for breast cancer	4 (8)
Equivocal findings during conventional breast imaging	1 (2)
Lesion size on imaging, cm	
Meanstandard deviation (range)	1.6 ± 1.8 (0.4–9.0)
Median (interquartile range)	1.1 (0.7–1.7)
Negative on imaging	37 (39)
Benign lesions (n = 24)	
Fibroadenoma	1 (4)
Complex fibroadenoma	1 (4)
Usual ductal hyperplasia	1 (4)
Intramammary lymph node	4 (17)
Pathologic results not validated by follow-up	17 (71)
Malignant lesions (n = 35)	
Ductal carcinoma in situ	6 (17)
Invasive ductal carcinoma	24 (69)
Invasive lobular carcinoma	2 (6)
Others*	3 (8)
Pathologic size of invasive cancer, cm, median (interquartile range)	1.1 (0–4.0)

Data are mean ± standard deviation (range) or n (%), unless otherwise specified.
*Others include metaplastic carcinoma (n = 1), microinvasive carcinoma (n = 1), and adenoid cystic carcinoma (n = 1).
MRI = magnetic resonance imaging

T MRI scanners with dedicated 16-channel or 18-channel breast coils (Ingenia CX, Philips Healthcare; Skyra, Siemens Healthineers). Unenhanced breast MRI sequences included DWI and axial non-fat-suppressed and fat-suppressed T1-weighted imaging sequences. For DWI, three b-values of 0, 800, and 1200 sec/mm^2 were chosen [1,17], and single-shot and multi-shot (simultaneous multi-slice readout-segmented) echo-planar imaging sequences were optimized with in-plane resolutions of 1.3 x 1.3 mm and 0.7 x 0.7 mm, slice thicknesses of 2.5–3 mm with no gap, and enough slices to cover both breasts in the axial dimension [16]. The acquisition time was 6–7 min. Additionally, an ADC map was created based on b = 0 sec/mm^2 and b = 800 sec/mm^2 DWI data [1,17]. The DWI was reconstructed as a single-summation image with maximum intensity projections (MIPs) in the sagittal and axial planes.

Table 2. Assessment categories and lesion types of 96 breasts on unenhanced breast MRI with diffusion-weighted imaging

Variable	Total (n = 96*)	Malignant (n = 35)	Benign or Negative (n = 61)
Final BI-RADS category†			
1	37 (39)	0 (0)	37 (100)
2	18 (19)	0 (0)	18 (100)
3	5 (5)	0 (0)	5 (100)
4	26 (27)	25 (96)	1 (4)
5	10 (10)	10 (100)	0 (0)
Lesion type (n = 59)			
Mass	37 (39)	28 (76)	9 (24)
Non-mass	8 (8)	6 (75)	2 (25)
Focus	8 (8)	1 (13)	7 (87)
Others‡	6 (6)	0 (0)	6 (100)

Data are numbers with percentages in parentheses.
*11 were artifacts caused by vessels and 26 were negative according to imaging (n = 37; 39%), †Using the interpretation algorithm, the unblinded readers categorized breast lesions into assessment categories with consensus as follows: BI-RADS category 1 (negative), BI-RADS category 2 (typical benign findings), BI-RADS category 3 (only one suspicion criterion), BI-RADS category 4 (at least two suspicious criteria), BI-RADS category 5 (more than two suspicious criteria), ‡Others include intramammary lymph nodes (n = 4), distortion caused by postoperative changes (n = 1), and chest wall lesions (n = 1).
MRI = magnetic resonance imaging, BI-RADS = Breast Imaging Reporting and Data System

Table 3. Summary of experience in breast MRI and diffusion-weighted imaging for 16 readers

Variables	Values
Years in standard breast MRI	
< 1	0 (0)
1–4	4 (25.0)
5–10	6 (37.5)
> 10	6 (37.5)
Experience in breast DWI	
No	0 (0)
Yes (multi-parametric)	16 (100.0)
Yes (stand-alone)	5 (31.2)
Years in breast DWI	
< 1	0 (0)
1–4	4 (25.0)
5–10	6 (37.5)
> 10	6 (37.5)
No. of breast DWI interpretation per week	
< 5	3 (18.8)
5–9	1 (6.2)
10–19	8 (50.0)
> 20	4 (25.0)

Data are numbers of readers with percentages in parentheses.
MRI = magnetic resonance imaging, DWI = diffusion-weighted imaging

https://doi.org/10.3348/kjr.2023.0528

Characteristics of the Readers

Sixteen breast radiologists from ten academic breast centers with different levels of experience in reading standard breast MRI scans, including DWI, were recruited for this study. All participating institutions consistently applied DWI features during MRI interpretation in clinical practice. Among them, one institution acquired stand-alone DWI for non-contrast breast cancer screening. Four readers (25%) had 1–4 years of experience, six (37.5%) had 5–10 years of experience, and the remaining six (37.5%) had more than 10 years of experience with standard breast MRI interpretation (Table 3).

Image Interpretation

Visualization and interpretation of all DWI images in the picture archiving and communication systems adhered to the imaging standards of Digital Imaging and Communication in Medicine, the imaging insight toolkit and window level were adjusted [19], and the data were anonymized. The readers were blinded to the clinical and pathological information and conventional breast imaging (mammography or breast US) of the study population. Furthermore, DCE-MRI sequences were not provided; therefore, the readers had no access to the DCE-MRI data.

A standardized image interpretation format was used for the reader study (Fig. 2). The interpretation was performed separately for each breast. First, readers were provided with the MIP of DWI to provide a quick overview of the entire breast volume, similar to the use of MIPs in abbreviated contrast-enhanced MRI protocols [3]. Notably, radiologists have recently published DWI interpretation guidelines [5,16] (Fig. 1). These DWI interpretation guidelines include three criteria: morphology, internal characteristics, and diffusion levels [5,17].

Lesion morphology on DWI was categorized as foci, masses, or nonmass lesions. In cases of lesions categorized as masses, the shape (round/oval, irregular) and internal signal pattern (homogeneous, heterogeneous, and rim) could be reported, whereas in nonmass lesions, the distribution (focal, regional, linear, and segmental) and internal signal pattern (homogeneous, heterogeneous) could be reported. Furthermore, lesions detected using DWI with high b-values were correlated with the ADC map. Quantitative ADC values were measured by drawing a region of interest (ROI) on the lesion on the ADC map. According to recently published guidelines, the use of a small ROI (3–10 mm^2) placed on the darkest part of the lesion on the ADC was measured and used as an ADC cut-off value of 1.3 x 10^{-3} mm^2/sec [5,17]. Breast lesions with ADC values lower than the cutoff were considered suspicious, and those with ADC values higher

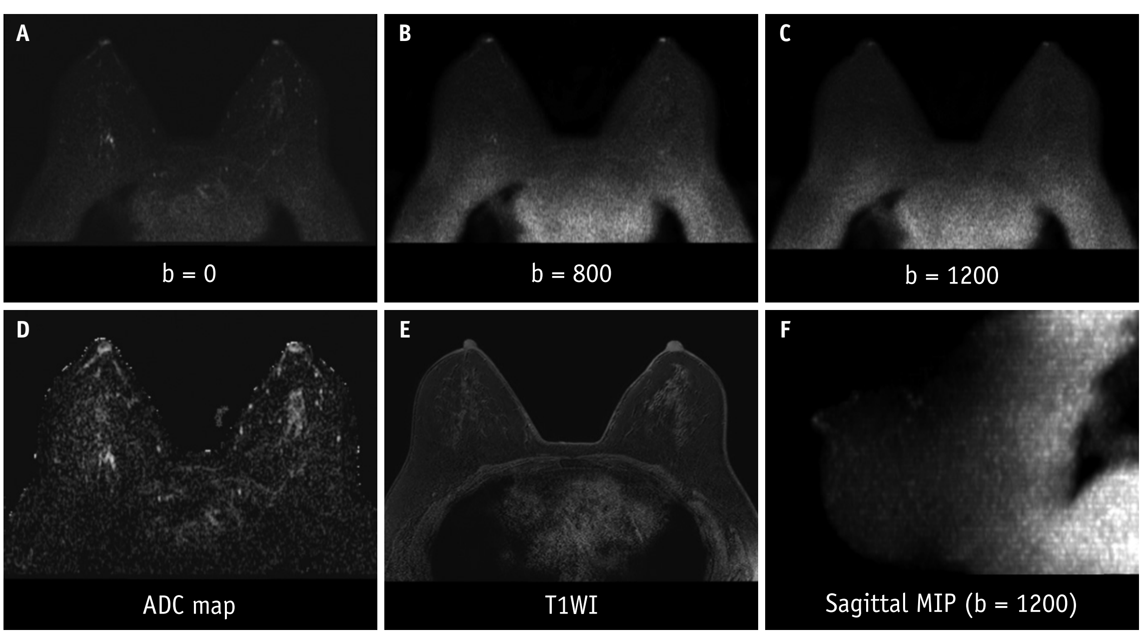

Fig. 2. Standardized image interpretation format used during the reader study. Both breasts were normal. Diffusion-weighted image sets consisting of T2-weighted images obtained without a diffusion gradient **(A)**, with a diffusion gradient b-value of 800 sec/mm^2 **(B)**, with a diffusion gradient b-value of 1200 sec/mm^2 **(C)**, and the parametric apparent diffusion coefficient (ADC) map **(D)**. Pre-contrast T1-weighted image (T1WI) **(E)** and sagittal maximum intensity projection (MIP) images **(F)** reconstructed from a diffusion-weighted b = 1200 sec/mm^2 image.

 https://doi.org/10.3348/kjr.2023.0528

than the cutoff were considered non-suspicious.

T2-weighted imaging was not used for analysis, and DWI images with b = 0 sec/mm^2 were used to provide T2-weighted imaging contrast. Unenhanced T1-weighted or T2-weighted images (b = 0 sec/mm^2) were obtained to provide further insight into lesion morphology and signal intensities. Finally, using the interpretation algorithm, the readers categorized the lesions into the following assessment categories [20]: negative findings, BI-RADS category 1; typical benign findings, BI-RADS category 2; lesions with only one suspicious criterion, BI-RADS category 3; lesions with at least two suspicious criteria, BI-RADS category 4; and lesions with more than two suspicious criteria, BI-RADS category 5. In cases of multiple breast lesions, the most suspicious findings were selected. A BI-RADS category of 0 was not permitted.

All readers were required to attend a 2-h unenhanced MRI interpretation training session, which included lectures and examples of background parenchymal signal intensity on DWI, as well as descriptions of benign and malignant lesions (these were not included in the test set). The training was conducted by an expert with more than 10 years of experience in breast imaging (W.K.M.) [21]. After a 5-month washout period, the same readers were required to reassess the same cases, recording the final BI-RADS categories and the localization of the detected lesions to ensure that they had identified the same target lesion. The data were entered into Microsoft Excel worksheets. Figure 3 and Figure 4 show the representative malignant and benign cases, respectively.

Statistical Analysis

Cases of BI-RADS category 4 or 5 with a cancer diagnosis were considered true positives. Cases with BI-RADS categories 1, 2, and 3 with negative or benign results were considered true negatives. Per-breast sensitivity, specificity, accuracy, positive predictive, and negative predictive values were calculated using 95% confidence intervals (CIs). The performance of interpretative tasks involving unenhanced MRI with DWI before and after training was compared using McNemar's test, and a generalized estimation equation was used to account for multiple readers per scan and the dependence between breasts. The degree of inter-reader agreement for the final assessment was measured using κ statistics. The z-test was used to test differences in multi-rater-weighted κ values [22]. The intraclass correlation coefficient (ICC) among the readers was computed to analyze the inter-reader agreement of BI-RADS categorization. This comparison was made both before and after training (see Supplementary Method for details). We also analyzed the sensitivity, specificity, and accuracy before and after training

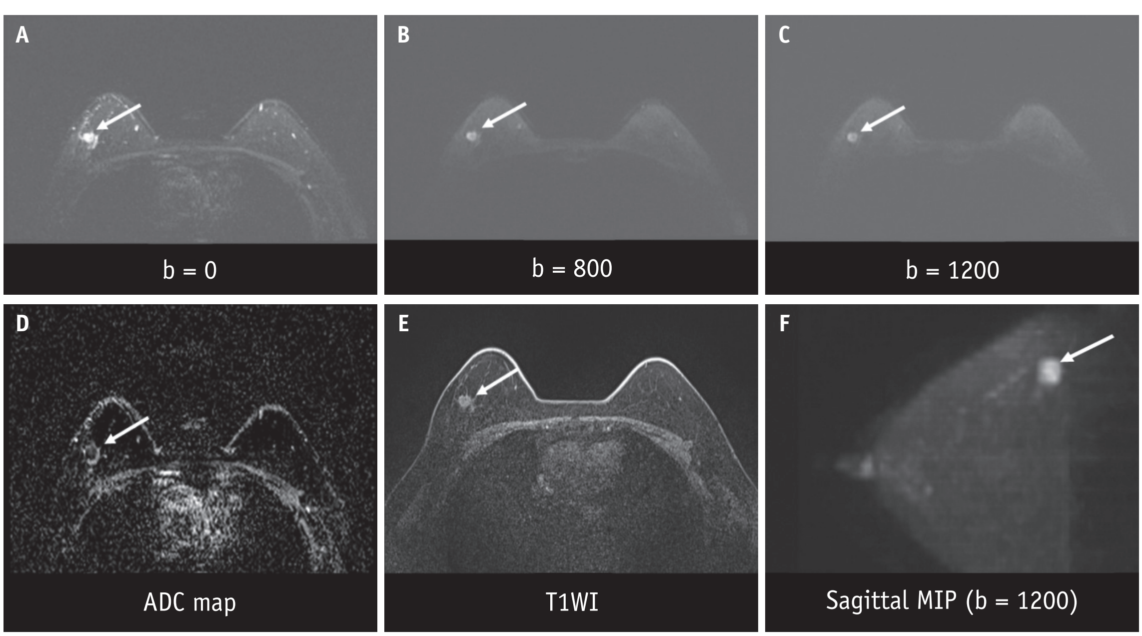

Fig. 3. A 64-year-old woman with invasive ductal carcinoma in the right breast. **A-C:** Axial diffusion-weighted imaging (DWI) shows an irregular mass (arrows) with heterogeneous signal intensity. **D:** The mass shows a low diffusion level with an apparent diffusion coefficient (ADC) value of 0.89 x 10^{-3} mm^2/sec (arrow). **E:** Fat-saturated T1-weighted imaging (T1WI) image showing an irregular mass with spiculated margins (arrow). **F:** Sagittal maximum intensity projection (MIP) showing a mass with high signal intensity (arrow). This is a representative case evaluated as a true positive by all readers.

https://doi.org/10.3348/kjr.2023.0528

Korean Journal of Radiology KJR

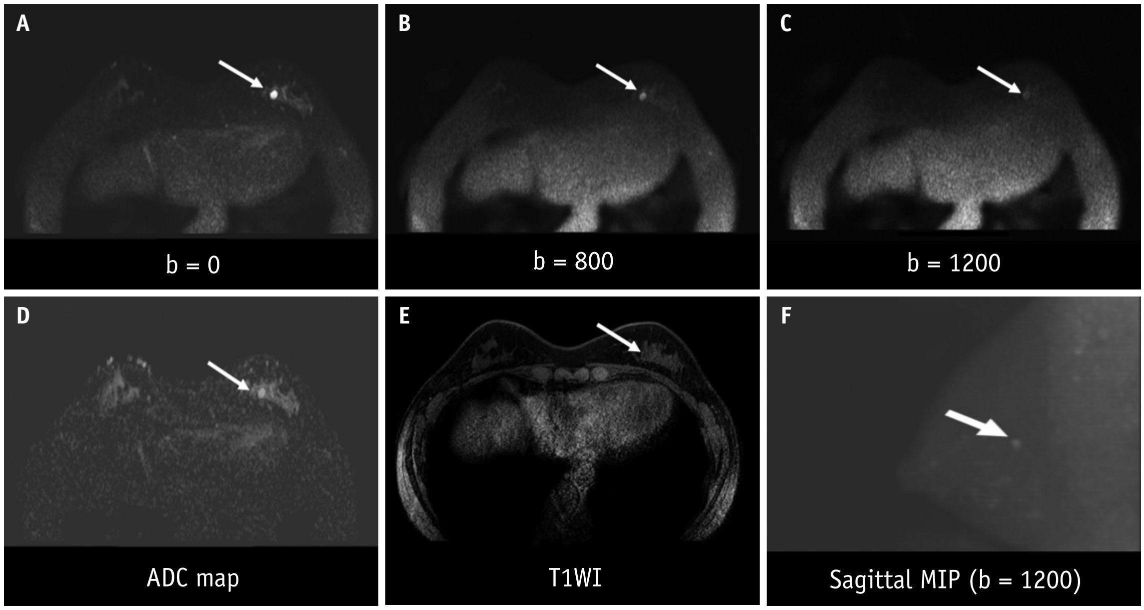

Fig. 4. A 52-year-old woman with a benign mass in the left breast. **A-C:** Axial diffusion-weighted imaging (DWI) images show a circumscribed mass (arrows) with homogeneous signal intensity and gradually decaying signal intensity from b = 0 sec/mm^2 to b = 1200 sec/mm^2. **D:** This mass shows a high diffusion level with an apparent diffusion coefficient (ADC) value of 2.17 x 10^{-3} mm^2/sec (arrow). **E:** Fat-saturated T1-weighted imaging (T1WI) image showing an isointense mass (arrow). **F:** Sagittal maximum intensity projection (MIP) showing a mass with high signal intensity (arrow). This is a representative case evaluated as a true negative by all readers.

Table 4. Diagnostic performance of unenhanced breast MRI with diffusion-weighted imaging per breast before and after training

Parameters	Before training	After training	*P*
Sensitivity	70.7 (59.4–79.9) [396/560]	69.8 (58.1–79.4) [391/560]	0.58
Specificity	90.8 (85.6–94.2) [886/976]	95.2 (90.8–97.5) [929/976]	0.001
PPV	81.5 (70.3–89.1) [396/486]	89.3 (79.4–94.7) [391/438]	< 0.001
NPV	84.4 (77.7–89.3) [886/1050]	84.6 (77.9–89.6) [929/1098]	0.75
Accuracy	83.5 (78.6–87.4) [1282/1536]	85.9 (80.9–89.8) [1320/1536]	0.01

Data are percentages with the 95% confidence interval in parentheses and raw data in brackets.
MRI = magnetic resonance imaging, PPV = positive predictive value, NPV = negative predictive value

according to the readers' level of experience in standard breast MRI interpretation using a generalized estimating equation. All *P*-values were two-sided, and $P < 0.05$ was considered significantly different (SAS 9.4 Version, SAS Institute Inc.).

RESULTS

Effect of Training on Diagnostic Performance

Table 4 and Figure 5 show the changes in the average diagnostic performance of the readers before and after training. Before training, the mean sensitivity, specificity, positive predictive, negative predictive, and accuracy values of the 16 readers in performing unenhanced MRI assessments were 70.7% (95% CI: 59.4–79.9), 90.8% (95% CI: 85.6–94.2), 81.5% (95% CI: 70.3–89.1), 84.4% (95% CI: 77.7–89.3), and 83.5% (95% CI: 78.6–87.4), respectively. After training, significant improvements were observed in the specificity (95.2%; 95% CI: 90.8–97.5; $P = 0.001$), positive predictive value (89.3%; 95% CI: 79.4–94.7; $P < 0.001$), and accuracy (85.9%; 95% CI: 80.9–89.8; $P = 0.01$), but no difference was observed in the sensitivity (69.8%; 95% CI: 58.1–79.4; $P = 0.58$) or negative predictive value (84.6%; 95% CI: 77.9–89.6; $P = 0.75$).

Readers with 1–4 years of experience considered 225 scans to be negative before training and 235 scans to be negative after training (difference of 10). This trend demonstrated a significant improvement in specificity, which improved from 92.2% (95% CI: 85.1–96.1) to 96.3% (95% CI: 90.0–98.7) ($P = 0.01$). Furthermore, readers with more than 10 years of experience considered 320 scans to be negative before training and 348 scans to be negative after training (difference of 28), showing a significant improvement in specificity, which improved from 87.4% (95% CI: 81.9–91.5) to 95.1% (95% CI: 90.7–97.4) ($P < 0.001$). However, readers with 5–10 years of experience showed no significant improvement in specificity (before training: 93.2% [95% CI: 88.1–96.2]; after training: 94.5% [95% CI: 89.9–97.1]; $P = 0.37$). The interaction between the training and experience of the readers was analyzed to determine whether the degree of improvement in specificity, as a result of training, differed according to reader expertise; however, no statistical significance was observed (difference of 4.1, 1.3, and 7.7%; $P = 0.06$) (Supplementary Table 1).

More detailed data on each breast are provided in Supplementary Table 2. Supplementary Table 3 shows the changes in the diagnostic performance of each reader before and after training. After training, the change in accuracy varied. Although five readers showed lower accuracy, 11 showed improved accuracy. Thus, it was evident that the overall accuracy was significantly different before and after training ($P = 0.01$) (Supplementary Table 3).

Two representative false-negative cases were identified: a 0.8 cm ductal carcinoma in situ of intermediate grade that was misclassified by 13 readers, which was measured as 0.9 cm at the periphery of the lower inner breast on DCE MRI (Fig. 6); and another 0.5 cm invasive ductal carcinoma of grade I that was misclassified by 15 readers, which manifested as the focus on DCE MRI (Fig. 7). Supplementary Figure 1 shows a representative false-positive case diagnosed as a complex fibroadenoma.

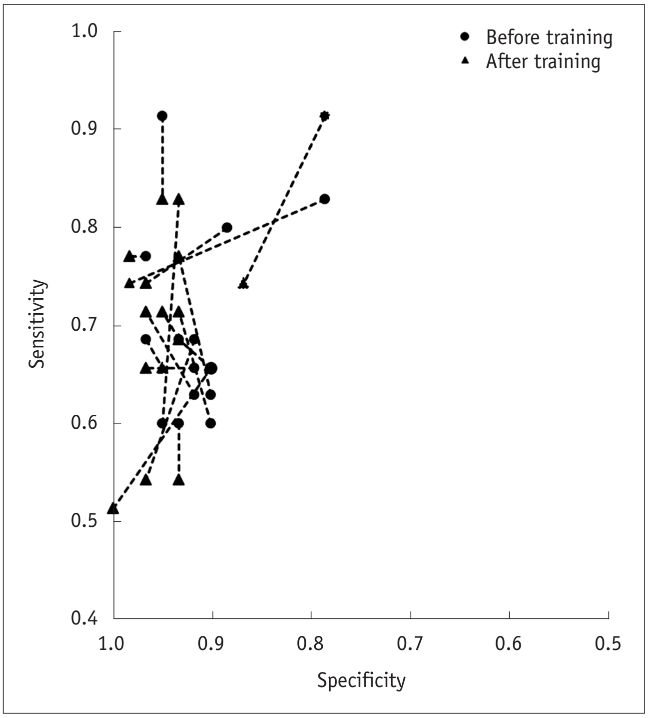

Fig. 5. Sensitivity and specificity of 16 readers before and after training.

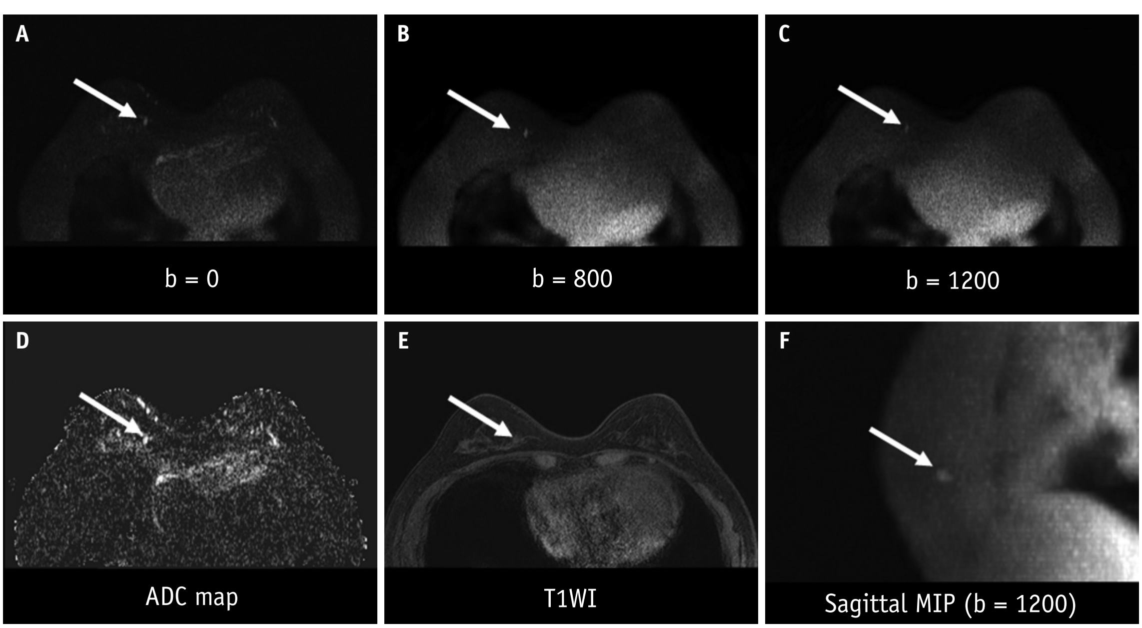

Fig. 6. A 56-year-old woman with ductal carcinoma in situ in the right breast. **A-C:** Axial diffusion-weighted imaging (DWI) shows a mass with hyperintensity (arrows). **D:** The corresponding lesion on the apparent diffusion coefficient (ADC) map shows a high diffusion level with an ADC value of 1.95 x 10^{-3} mm^2/sec (arrow). **E:** Fat-saturated T1-weighted imaging (T1WI) image showing an irregular mass (arrow). **F:** Sagittal maximum intensity projection (MIP) showing a mass with high signal intensity (arrow). This case was missed by 13 readers.

Korean Journal of Radiology
KJR

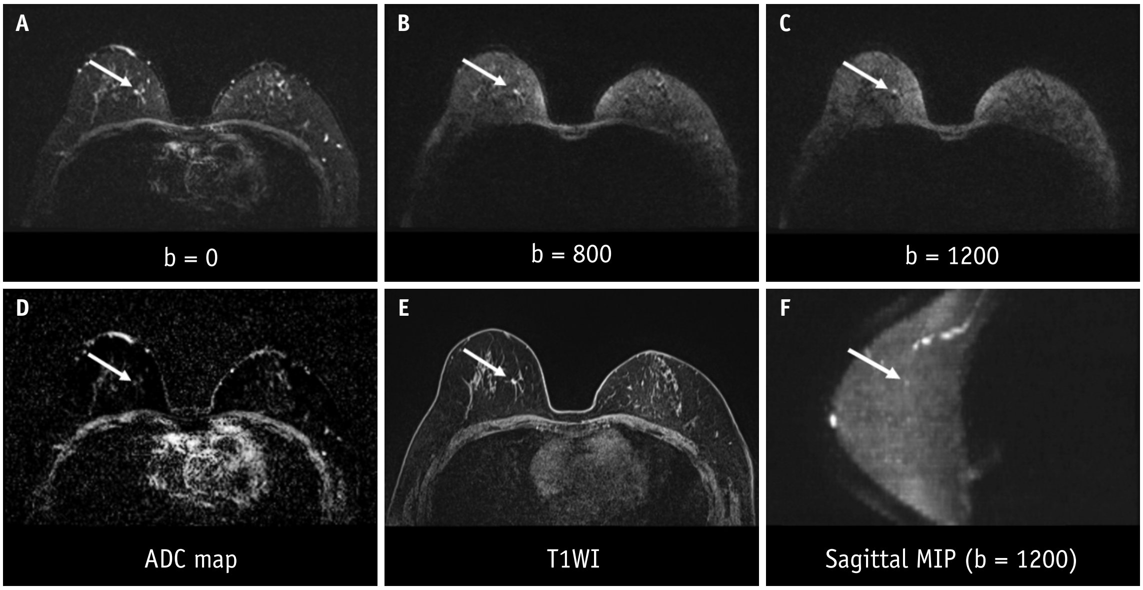

Fig. 7. A 59-year-old woman with invasive ductal carcinoma in the right breast. **A-C:** Axial diffusion-weighted imaging (DWI) shows a focus with subtle hyperintensity (arrows). **D:** The corresponding lesion on the apparent diffusion coefficient (ADC) map shows a low diffusion level with an ADC value of 0.90 x 10^{-3} mm^2/sec (arrow). **E:** Fat-saturated T1-weighted imaging (T1WI) image showing an irregular mass (arrow). **F:** Sagittal maximum intensity projection (MIP) showing a focus with high signal intensity (arrow). This case was missed by 15 readers.

Table 5. Kappa for inter-reader agreement between 16 readers who performed BI-RADS assessments before and after training

BI-RADS category	κ before training (95% CI)	κ after training (95% CI)	Change in κ (95% CI)	*P*
Individual categories	0.57 (0.52–0.63)	0.68 (0.62–0.74)	0.11 (0.02–0.18)	0.01
1 vs. all others	0.73 (0.67–0.80)	0.75 (0.70–0.81)	0.02 (-0.06–0.11)	0.65
2 vs. all others	0.71 (0.65–0.77)	0.78 (0.73–0.84)	0.07 (-0.01–0.15)	0.08
3 vs. all others	0.79 (0.74–0.84)	0.82 (0.77–0.87)	0.03 (-0.03–0.10)	0.38
4 vs. all others	0.74 (0.67–0.80)	0.79 (0.73–0.85)	0.05 (-0.03–0.14)	0.24
5 vs. all others	0.87 (0.82–0.92)	0.89 (0.85–0.94)	0.02 (-0.04–0.08)	0.47

κ values: ≤ 0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–0.99, almost perfect agreement.
BI-RADS = Breast Imaging Reporting and Data System, CI = confidence interval

Table 6. Intraclass correlation coefficient for inter-reader agreement between 16 readers who performed BI-RADS assessments before and after training

ICC before training (95% CI)	ICC after training (95% CI)	Change in ICC (95% CI)	*P*
0.73 (0.69–0.74)	0.79 (0.76–0.80)	0.06 (0.03–0.10)	0.002

The ICC before and after training was computed to analyze agreement of BI-RADS categorization across multiple readers, using absolute agreement measure, single measurement, two-way mixed effect model. ICC values: < 0.4, poor reproducibility; 0.40–0.75, good reproducibility; > 0.75, excellent reproducibility.
BI-RADS = Breast Imaging Reporting and Data System, ICC = intraclass correlation coefficient, CI = confidence interval

Effect of Training on Reader Agreement

Before training, the average inter-reader agreement was moderate for the unenhanced MRI assessments. Improvements were observed after training, with substantial agreement with unenhanced MRI assessments. The κ values for interpreting individual BI-RADS categories were 0.57 (95% CI: 0.52–0.63) before training and 0.68 (95% CI: 0.62–0.74) after training, with a difference of 0.11 (95% CI: 0.02–0.18), which was statistically significant (*P* = 0.01) (Table 5). After training, the changes in κ were 0.07 for BI-RADS category 2 (vs. all others) and 0.02 for BI-RADS category 1 (vs. all others) and BI-RADS category 5 (vs. all others) (Supplementary Fig. 2). The ICC was 0.73 (95% CI: 0.69–0.74) before and 0.79 (95% CI: 0.76–0.80) after training (P = 0.002) (Table 6). Separate results for both breasts are presented in Supplementary Tables 4 and 5.

DISCUSSION

Standardized interpretation and image acquisition are prerequisites for promoting unenhanced breast MRI with DWI as an effective tool for breast cancer detection. In this reader study involving 16 breast radiologists, we assessed the diagnostic performance and inter-reader variability before and after training to perform unenhanced breast MRI interpretation using a standardized algorithm. Our results showed that minimal training improved the diagnostic performance, especially the specificity (90.8%–95.2%, P = 0.001) and accuracy (83.5%–85.9%, P = 0.01). Additionally, it resulted in substantial inter-reader agreement (κ value 0.57 before and 0.68 after training, P = 0.01) and excellent reproducibility (ICC value of 0.73 before and 0.79 after training, P = 0.002) among all readers, despite their diverse backgrounds and experiences with breast imaging, thus highlighting the importance of education in this context.

Most published studies on DWI have involved diagnostic evaluations [2,3,23] or evaluations of women with newly diagnosed breast cancer [24-26]. However, there is growing interest in exploring the use of unenhanced MRI with DWI as a standalone screening tool for breast cancer detection [27]. Multiple retrospective studies have included both positive and negative imaging findings to simulate the screening population and assess the potential of unenhanced MRI [8-10,12-14,28]. However, some of these studies were enriched with high cancer prevalence (25%–67%) [9,10,12-14].

In comparison to previous studies, we simulated a screening setting by including negative and benign cases, and the readers were blinded to the cancer prevalence. In addition, because both lesion detection and accurate ADC quantitation are important for DWI screening, we optimized our high-resolution MRI and acquired three b-value images [1,16]. Readers were required to qualitatively and quantitatively assess lesions detected on DWI with an ADC threshold of 1.30 x 10^{-3} mm^2/sec according to the DWI interpretation algorithm proposed by the EUSOBI and DWIST groups [16,17]. Readers could detect a unique high signal on at least two of the three b-value images and measure the ADC value using the ADC map cross-correlation. However, for small lesions, the partial volume effects of ADC measurements should be considered. Additionally, readers qualitatively evaluated the lesion signal intensity and morphology on unenhanced T1-weighted or T2-weighted (b = 0 sec/mm^2) sequences, which may assist in avoiding misclassification of false-positive benign lesions, such as complicated cysts, as well as in avoiding misdiagnosis of false-negative malignant lesions, including mucinous carcinoma [29].

A recent study [20] involving 108 breast lesions (61 malignant and 47 benign) evaluated the diagnostic performance of an unenhanced MRI protocol combining high-resolution DWI and T1-weighted and T2-weighted images using a decision tree based on the modified BI-RADS lexicon. This is similar to our interpretation algorithm, although we did not define the specific associated features for upgrades between categories and subcategories. The results of that study were similar to ours, with a sensitivity, specificity, and accuracy of 70%, 90%, and 82% [20].

Another study investigated the diagnostic performance and inter-reader agreement between expert and non-expert readers using DWI [18]. In this reader study involving four radiologists, 382 women were analyzed, and the reported overall sensitivity and specificity were 80%–93% and 83%–93%, respectively. In the study, the inter-reader agreement was substantial (κ = 0.74) for two expert readers and poor (κ = 0.37) for two non-expert readers [18]. These results emphasize the importance of experience and training before the clinical implementation of DWI as a standalone sequence.

In our study, the improved inter-reader agreement after training showed that the accurate interpretation of unenhanced MRI with DWI can be learned and is reproducible; therefore, non-expert readers could benefit from training. We also assessed the effect of training and the number of years of experience on standard breast MRI interpretation. There was an improvement in specificity for readers with 1–4 years and > 10 years of experience, and there was an improvement in accuracy for readers with > 10 years of experience. Nonetheless, overall, no association between readers' expertise and the effects of training was observed. This was interesting because we expected to observe greater improvements among readers with less experience [30]. This result may have been caused by individual predispositions and personal skills rather than by different levels of expertise. Alternatively, it may reflect the difficulty of performing unenhanced MRI interpretation, regardless of experience. Although the current BI-RADS atlas does not yet provide a standardized interpretation algorithm that integrates DWI with contrast-enhanced images, we expect that DWI will be implemented, trained, and ultimately increase cancer detection as a standalone modality.

Our study has several limitations. First, the sample size was relatively small, and the study was conducted at a single institution. The inclusion of a larger number of patients would have improved the internal and external validity of the study. Second, our cohort included patients with variable indications for MRI. Furthermore, the prevalence of per-breast cancer was higher than that in previous studies simulating the screening setting [11,13]; thus, the outcomes cannot be directly compared. Third, although the assessments per breast were recorded and the performances were calculated, readers had access to both breasts in one image simultaneously. Therefore, our results should be interpreted carefully, particularly with respect to the specificity and accuracy, which may have been overestimated. Fourth, the complete clinical information was not provided to the readers, as the task was designed for skill assessment rather than the interpretation of clinical cases. This lack of a true impact on the clinical management of patients may have reduced the care of radiologists when interpreting the images, thus contributing to the reduced accuracy. Finally, acknowledging that the diagnostic performance of DWI is limited for smaller breast cancers (1 cm or less in size), mucinous carcinomas with high water content, and breast lesions manifesting as nonmass enhancement is essential [31].

In conclusion, brief training can help improve the diagnostic performance and inter-reader agreement of breast radiologists in interpreting unenhanced breast MRI with DWI. Dedicated training before performing interpretive skill tasks is encouraged for breast radiologists to adopt unenhanced MRI examinations as a standalone method.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2023.0528.

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

Vivian Youngjean Park and Woo Jung Choi, the editor board member of the *Korean Journal of Radiology*, were not involved in the editorial evaluation or decision to publish this article.

All authors have declared no conflicts of interest.

Author Contributions

Conceptualization: Su Hyun Lee, Woo Kyung Moon, Su Min Ha. Data curation: Yeon Soo Kim, Su Hyun Lee, Su Min Ha. Formal analysis: Su Hyun Lee, Woo Kyung Moon, Soo Hee Kang, Su Min Ha. Funding acquisition: Woo Kyung Moon, Su Min Ha. Investigation: Yeon Soo Kim, Su Hyun Lee, Woo Kyung Moon, Soo Hee Kang, Su Min Ha. Methodology: Yeon Soo Kim, Su Hyun Lee, Woo Kyung Moon, Soo Hee Kang, Su Min Ha. Project administration: Woo Kyung Moon, Su Min Ha. Resources: Yeon Soo Kim, Su Hyun Lee, Su Min Ha. Software: Su Hyun Lee, Woo Kyung Moon, Su Min Ha. Supervision: Woo Kyung Moon, Su Min Ha. Validation: Woo Kyung Moon, Su Min Ha. Visualization: Woo Kyung Moon, Su Min Ha. Writing—original draft: Yeon Soo Kim, Su Min Ha. Writing—review & editing: all authors.

ORCID IDs

Yeon Soo Kim
https://orcid.org/0000-0003-1838-202X
Su Hyun Lee
https://orcid.org/0000-0002-0171-8060
Soo-Yeon Kim
https://orcid.org/0000-0001-8915-3924
Eun Sil Kim
https://orcid.org/0000-0002-0632-9902
Ah Reum Park
https://orcid.org/0000-0001-9672-844X
Jung Min Chang
https://orcid.org/0000-0001-5726-9797
Vivian Youngjean Park
https://orcid.org/0000-0002-5135-4058
Jung Hyun Yoon
https://orcid.org/0000-0002-2100-3513
Bong Joo Kang
https://orcid.org/0000-0002-5991-6035
Bo La Yun
https://orcid.org/0000-0002-5457-7847
Tae Hee Kim
https://orcid.org/0000-0002-0520-2395
Eun Sook Ko
https://orcid.org/0000-0002-0399-7956
A Jung Chu
https://orcid.org/0000-0003-2018-6706
Jin You Kim
https://orcid.org/0000-0002-9631-501X

Inyoung Youn
https://orcid.org/0000-0002-3518-3332
Eun Young Chae
https://orcid.org/0000-0002-3478-6721
Woo Jung Choi
https://orcid.org/0000-0003-1341-6470
Hee Jeong Kim
https://orcid.org/0000-0001-9831-542X
Soo Hee Kang
https://orcid.org/0000-0002-7353-8122
Su Min Ha
https://orcid.org/0000-0002-1833-0919
Woo Kyung Moon
https://orcid.org/0000-0001-8931-3772

Funding Statement

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00210238) and the radiological research foundation of Korea (2021_02).

Acknowledgments

We thank Hee Jung Shin, MD, a professor at Asan Medical Center and Min Jung Kim, MD, a professor at Severance Hospital, for helping us with organization.

REFERENCES

1. Amornsiripanitch N, Bickelhaupt S, Shin HJ, Dang M, Rahbar H, Pinker K, et al. Diffusion-weighted MRI for unenhanced breast cancer screening. *Radiology* 2019;293:504-520
2. Baltzer PAT, Bickel H, Spick C, Wengert G, Woitek R, Kapetas P, et al. Potential of noncontrast magnetic resonance imaging with diffusion-weighted imaging in characterization of breast lesions: intraindividual comparison with dynamic contrast-enhanced magnetic resonance imaging. *Invest Radiol* 2018;53:229-235
3. Bickelhaupt S, Laun FB, Tesdorff J, Lederer W, Daniel H, Stieber A, et al. Fast and noninvasive characterization of suspicious lesions detected at breast cancer X-Ray screening: capability of diffusion-weighted MR imaging with MIPs. *Radiology* 2016;278:689-697
4. Rahbar H, Zhang Z, Chenevert TL, Romanoff J, Kitsch AE, Hanna LG, et al. Utility of diffusion-weighted imaging to decrease unnecessary biopsies prompted by breast MRI: a Trial of the ECOG-ACRIN Cancer Research Group (A6702). *Clin Cancer Res* 2019;25:1756-1765
5. Lee SH, Shin HJ, Moon WK. Diffusion-weighted magnetic resonance imaging of the breast: standardization of image acquisition and interpretation. *Korean J Radiol* 2021;22:9-22
6. McDonald RJ, McDonald JS, Kallmes DF, Jentoft ME, Murray DL, Thielen KR, et al. Intracranial gadolinium deposition after contrast-enhanced MR Imaging. *Radiology* 2015;275:772-782
7. Ha SM, Chang JM, Lee SH, Kim ES, Kim SY, Cho N, et al. Diffusion-weighted MRI at 3.0 T for detection of occult disease in the contralateral breast in women with newly diagnosed breast cancer. *Breast Cancer Res Treat* 2020;182:283-297
8. Kang JW, Shin HJ, Shin KC, Chae EY, Choi WJ, Cha JH, et al. Unenhanced magnetic resonance screening using fused diffusion-weighted imaging and maximum-intensity projection in patients with a personal history of breast cancer: role of fused DWI for postoperative screening. *Breast Cancer Res Treat* 2017;165:119-128
9. Kazama T, Kuroki Y, Kikuchi M, Sato Y, Nagashima T, Miyazawa Y, et al. Diffusion-weighted MRI as an adjunct to mammography in women under 50 years of age: an initial study. *J Magn Reson Imaging* 2012;36:139-144
10. McDonald ES, Hammersley JA, Chou SH, Rahbar H, Scheel JR, Lee CI, et al. Performance of DWI as a rapid unenhanced technique for detecting mammographically occult breast cancer in elevated-risk women with dense breasts. *AJR Am J Roentgenol* 2016;207:205-216
11. Rotili A, Trimboli RM, Penco S, Pesapane F, Tantrige P, Cassano E, et al. Double reading of diffusion-weighted magnetic resonance imaging for breast cancer detection. *Breast Cancer Res Treat* 2020;180:111-120
12. Telegrafo M, Rella L, Stabile Ianora AA, Angelelli G, Moschetta M. Unenhanced breast MRI (STIR, T2-weighted TSE, DWIBS): an accurate and alternative strategy for detecting and differentiating breast lesions. *Magn Reson Imaging* 2015;33:951-955
13. Trimboli RM, Verardi N, Cartia F, Carbonaro LA, Sardanelli F. Breast cancer detection using double reading of unenhanced MRI including T1-weighted, T2-weighted STIR, and diffusion-weighted imaging: a proof of concept study. *AJR Am J Roentgenol* 2014;203:674-681
14. Yabuuchi H, Matsuo Y, Sunami S, Kamitani T, Kawanami S, Setoguchi T, et al. Detection of non-palpable breast cancer in asymptomatic women by using unenhanced diffusion-weighted and T2-weighted MR imaging: comparison with mammography and dynamic contrast-enhanced MR imaging. *Eur Radiol* 2011;21:11-17
15. Iima M, Partridge S, Le Bihan D. *Diffusion MRI of the breast*. 1st ed. Philadelphia: Elsevier, 2023:87
16. Shin HJ, Lee SH, Park VY, Yoon JH, Kang BJ, Yun B, et al. Diffusion-weighted magnetic resonance imaging for breast cancer screening in high-risk women: design and imaging protocol of a prospective multicenter study in Korea. *J Breast Cancer* 2021;24:218-228
17. Baltzer P, Mann RM, Iima M, Sigmund EE, Clauser P, Gilbert FJ, et al. Diffusion-weighted imaging of the breast-a consensus and mission statement from the EUSOBI International Breast Diffusion-Weighted Imaging working group. *Eur Radiol*

2020;30:1436-1450
18. Pesapane F, Rotili A, Penco S, Montesano M, Agazzi GM, Dominelli V, et al. Inter-reader agreement of diffusion-weighted magnetic resonance imaging for breast cancer detection: a multi-reader retrospective study. *Cancers (Basel)* 2021;13:1978
19. Malyarenko D, Fedorov A, Bell L, Prah M, Hectors S, Arlinghaus L, et al. Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. *J Med Imaging (Bellingham)* 2018;5:011006
20. Ota R, Kataoka M, Iima M, Honda M, Kishimoto AO, Miyake KK, et al. Evaluation of breast lesions based on modified BI-RADS using high-resolution readout-segmented diffusion-weighted echo-planar imaging and T2/T1-weighted image. *Magn Reson Imaging* 2023;98:132-139
21. Berg WA, Bandos AI, Zuley ML, Waheed UX. Training radiologists to interpret contrast-enhanced mammography: toward a standardized lexicon. *Journal of Breast Imaging* 2021;3:176-189
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174
23. Baltzer PA, Benndorf M, Dietzel M, Gajda M, Camara O, Kaiser WA. Sensitivity and specificity of unenhanced MR mammography (DWI combined with T2-weighted TSE imaging, ueMRM) for the differentiation of mass lesions. *Eur Radiol* 2010;20:1101-1110
24. Kuroki-Suzuki S, Kuroki Y, Nasu K, Nawano S, Moriyama N, Okazaki M. Detecting breast cancer with non-contrast MR imaging: combining diffusion-weighted and STIR imaging. *Magn Reson Med Sci* 2007;6:21-27
25. Belli P, Bufi E, Bonatesta A, Patrolecco F, Padovano F, Giuliani M, et al. Unenhanced breast magnetic resonance imaging: detection of breast cancer. *Eur Rev Med Pharmacol Sci* 2016;20:4220-4229
26. Shin HJ, Chae EY, Choi WJ, Ha SM, Park JY, Shin KC, et al. Diagnostic performance of fused diffusion-weighted imaging using unenhanced or postcontrast T1-Weighted MR imaging in patients with breast cancer. *Medicine (Baltimore)* 2016;95:e3502
27. Mann RM, Kuhl CK, Moy L. Contrast-enhanced MRI for breast cancer screening. *J Magn Reson Imaging* 2019;50:377-390
28. Ha SM, Chang JM, Lee SH, Kim ES, Kim SY, Kim YS, et al. Detection of contralateral breast cancer using diffusion-weighted magnetic resonance imaging in women with newly diagnosed breast cancer: comparison with combined mammography and whole-breast ultrasound. *Korean J Radiol* 2021;22:867-879
29. Radovic N, Ivanac G, Divjak E, Biondic I, Bulum A, Brkljacic B. Evaluation of breast cancer morphology using diffusion-weighted and Dynamic Contrast-Enhanced MRI: intermethod and interobserver agreement. *J Magn Reson Imaging* 2019;49:1381-1390
30. Miglioretti DL, Gard CC, Carney PA, Onega TL, Buist DS, Sickles EA, et al. When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology* 2009;253:632-640
31. Avendano D, Marino MA, Leithner D, Thakur S, Bernard-Davila B, Martinez DF, et al. Limited role of DWI with apparent diffusion coefficient mapping in breast lesions presenting as non-mass enhancement on dynamic contrast-enhanced MRI. *Breast Cancer Res* 2019;21:136