


# Added value of deep learning-based computer-aided diagnosis and shear wave elastography to b-mode ultrasound for evaluation of breast masses detected by screening ultrasound

Min Young Kim, MD<sup>a,b,c</sup>, Soo-Yeon Kim, MD, PhD<sup>a,b,c,\*</sup> , Yeon Soo Kim, MD<sup>a,b,c</sup>, Eun Sil Kim, MD<sup>a,b,c</sup>, Jung Min Chang, MD, PhD<sup>a,b,c</sup>

## Abstract

Low specificity and operator dependency are the main problems of breast ultrasound (US) screening. We investigated the added value of deep learning-based computer-aided diagnosis (S-Detect) and shear wave elastography (SWE) to B-mode US for evaluation of breast masses detected by screening US.

Between February 2018 and June 2019, B-mode US, S-Detect, and SWE were prospectively obtained for 156 screening US-detected breast masses in 146 women before undergoing US-guided biopsy. S-Detect was applied for the representative B-mode US image, and quantitative elasticity was measured for SWE. Breast Imaging Reporting and Data System final assessment category was assigned for the datasets of B-mode US alone, B-mode US plus S-Detect, and B-mode US plus SWE by 3 radiologists with varied experience in breast imaging. Area under the receiver operator characteristics curve (AUC), sensitivity, and specificity for the 3 datasets were compared using Delong's method and McNemar test.

Of 156 masses, 10 (6%) were malignant and 146 (94%) were benign. Compared to B-mode US alone, the addition of S-Detect increased the specificity from 8%–9% to 31%–71% and the AUC from 0.541–0.545 to 0.658–0.803 in all radiologists (All  $P < .001$ ). The addition of SWE to B-mode US also increased the specificity from 8%–9% to 41%–75% and the AUC from 0.541–0.545 to 0.709–0.823 in all radiologists (All  $P < .001$ ). There was no significant loss in sensitivity when either S-Detect or SWE were added to B-mode US.

Adding S-Detect or SWE to B-mode US improved the specificity and AUC without loss of sensitivity.

**Abbreviations:** AUC = area under the receiver operator characteristics curve, BI-RADS = breast imaging reporting and data systems, CAD = computer-aided diagnosis, DCIS = ductal carcinoma in situ, IDC = invasive ductal carcinoma, ROI = region of interest, SWE = shear wave elastography, US = ultrasound.

**Keywords:** breast, computer-aided diagnosis, deep learning, screening, shear wave elastography, ultrasound

Editor: Meliha Ekinci.

This study has been supported by the research fund of Samsung Medison and Seoul National University Hospital Grant No. 06-2018-0030.

The authors have no conflicts of interests to disclose.

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

<sup>a</sup> Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea, <sup>b</sup> Department of Radiology, Seoul National College of Medicine, Seoul, Republic of Korea, <sup>c</sup> Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Republic of Korea.

\* Correspondence: Soo-Yeon Kim, Department of Radiology, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea (e-mail: sooyeonkim41@gmail.com).

Copyright © 2021 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Kim MY, Kim SY, Kim YS, Kim ES, Chang JM. Added value of deep learning-based computer-aided diagnosis and shear wave elastography to b-mode ultrasound for evaluation of breast masses detected by screening ultrasound. *Medicine* 2021;100:31(e26823).

Received: 25 March 2021 / Received in final form: 25 June 2021 / Accepted: 15 July 2021

<http://dx.doi.org/10.1097/MD.00000000000026823>

## 1. Introduction

Breast ultrasound (US) can detect early stage breast cancers in asymptomatic women with dense breast.<sup>[1–4]</sup> However, low specificity, low positive predictive value, and operator dependency are the main problems of breast US.<sup>[1,5–7]</sup> To overcome the limitations of conventional B-mode US, additional US techniques measuring quantitative features of stiffness values or morphologic characteristics have been developed. For instance, US elastography, which measures and visualizes the intrinsic strain of a target mass, provides additional information for mass characterization, and increases the accuracy of breast US by providing tissue stiffness information.<sup>[8,9]</sup> Among elastography techniques, shear wave elastography (SWE) evaluates the propagation speed of shear-wave through tissue and provides the quantitative measurements, and this is more reproducible than strain elastography.<sup>[10,11]</sup> Prior studies have reported that addition of SWE features to B-mode US improved specificity without losing sensitivity.<sup>[12–14]</sup>

As another quantitative technique, computer-aided diagnosis (CAD) has been applied to breast US interpretation, providing assistance in the morphologic analysis of breast masses.<sup>[15–17]</sup> S-Detect is a recently developed deep learning-based CAD software (S-Detect; Samsung Medison Co., Seongnam, Korea).<sup>[18–25]</sup> It analyzes the morphological features of breast

masses according to the breast imaging reporting and data systems (BI-RADS) lexicon and then provides a dichotomized final assessment (“possibly benign” or “possibly malignant”) on the possibility of malignancy of a breast mass.<sup>[18–25]</sup> Prior studies have reported that use of S-Detect in assessment of breast masses has improved diagnostic accuracy and specificity.<sup>[18–25]</sup>

Screening US examinations have detected average 3 to 4 breast cancers per 1000 US examinations in women with dense breast.<sup>[1–4]</sup> Efforts to detect more cancers using screening US have led to a lot of recalls and false-positive biopsies. As described above, the additional use of SWE or S-Detect has shown improved performances in assessing breast masses compared to using conventional B-mode US alone.<sup>[12–14,18–25]</sup> However, prior studies using SWE or S-Detect have included patients with variable conditions regardless of patients’ symptoms or mammographic findings. To our knowledge, no study has evaluated the added value of SWE and S-Detect focusing on breast masses detected by screening US. Breast cancers detected by screening US tend to be small in size and lack of typical malignant features, compared to palpable breast cancers or screening mammography-detected breast cancers.<sup>[26,27]</sup> Thus, the additional information obtained from SWE and S-Detect could be more useful for evaluating breast masses detected by screening US. Furthermore, using the complementary tools would be more helpful to less-experienced radiologists as it provides additional information to aid diagnosis.<sup>[23–25]</sup>

Therefore, the purpose of our study was to evaluate the added value of S-Detect and SWE to B-mode US for evaluating breast masses detected by screening US among radiologists with different level of experiences.

## 2. Materials and methods

### 2.1. Study participants

This single center prospective study was approved by an institutional review board and the recruited participants with written informed consent from February 2018 to June 2019 in a tertiary university hospital. The inclusion criteria were asymptomatic women scheduled for US-guided core needle biopsy for breast masses detected by screening US. The exclusion criteria were breast symptoms including lumps or nipple discharge, positive findings on mammography (BI-RADS final assessment category 0, 3, 4, or 5), a personal history of breast cancer, a recently diagnosed breast cancer, and refusal to participate in this study.

### 2.2. Ultrasound examination

All US examinations were performed before US-guided biopsy using 3 to 12 MHz linear probes and a real-time US system (RS85 Prestige; Samsung Medison Co., Seongnam, Korea) equipped with a deep learning-based CAD software (S-Detect). S-Detect employs a deep convolutional neural network that can process large-scale high-dimensional data and determine how to weight parameters through multiple layers of abstraction.<sup>[28]</sup> Weights are used to compute the representation of the data at each layer, and the complex features of multiple levels of representation can be learned. In the final layer, the inputs from the previous layers are computed as probabilities and the outputs from the other networks are combined to make a final decision.

B-mode US, S-Detect, and SWE images for each breast mass were obtained before undergoing biopsy by a dedicated breast

radiologist (J.M.C. with 13 years of experience). The most representative images were selected on B-mode US, followed by application of S-Detect software. To obtain S-Detect images, when the center of mass was indicated by the radiologist, a region of interest (ROI) was automatically drawn along the margin of the mass by the S-Detect software. An analysis button was pressed after the ROI was drawn, and then ROI-based US features and the possibility of malignancy (the final assessment) were automatically evaluated and displayed on the US monitor. The automatic analysis took less than 5 seconds. The ROI-based US features included shape, orientation, margin, posterior features, and echo pattern characteristics according to the 5th edition of BI-RADS lexicon. The final assessment result was provided in the dichotomized form of “possibly benign” or “possibly malignant”.

After B-mode US with S-Detect image acquisition, SWE images were generated by the same radiologists. On SWE, a color-coded image representing the elasticity (kilopascals, kPa) for each pixel was obtained with a default color scale ranging from dark blue, green, yellow, orange, to red. The closer it is to the red scale, the harder it is, and the closer it is to the blue scale, the softer it is. Quantitative elasticity values ranged from 0 to 180 kPa. Once the color-coded SWE image was displayed on the screen, 2 circular ROIs were placed on the mass and normal fat region, respectively for the quantitative assessment.<sup>[12,13]</sup> The ROI for mass was placed on the stiffest part of the mass based on the color scale, and the ROI for fat was placed on the normal fat region showing a uniform dark blue color. The circular ROI with 0.9-mm diameter was a default mode recommended by the vendor. For each ROI, the elasticity parameters including the maximum, minimum, mean, and standard deviation were calculated and displayed on the US monitor as kPa. The lesion-to-fat elasticity ratio was defined as mean elasticity of the lesion ROI divided by mean elasticity of the fat ROI and displayed on the US monitor as percentages.

### 2.3. Image analysis

Prospective image analysis was performed by the radiologist during image acquisition. The radiologist assessed the BI-RADS final assessment category 3 times as follows. First, BI-RADS category was evaluated based on B-mode US features alone (B-mode US). Second, BI-RADS category was evaluated by combining S-Detect results to B-mode US (B-mode US plus S-Detect). Third, BI-RADS category was evaluated by combining SWE results to B-mode US (B-mode US plus SWE). All breast masses underwent biopsy regardless of the image review results.

After completion of participants’ enrollment, additional retrospective image analysis was independently performed by 2 radiologists with different experiences in breast imaging (S.Y.K., a dedicated breast radiologist with 6 years of experience and M.Y.K., a 4th-year resident with 3 months of breast imaging training). The participant’s clinical information and pathology were blinded during image analysis. There was a total of 3 review sessions with two-week intervals between each review. In the first review session (B-mode US), the reviewers analyzed only the B-mode US features and recorded the BI-RADS final assessment category. In the second review session (B-mode US plus S-Detect), the reviewers examined the B-mode US and S-Detect results together and recorded the BI-RADS final assessment category. SWE results were blinded in the second session. In the third review session (B-mode US plus SWE), SWE results were united

with the B-mode US to evaluate the BI-RADS final assessment category. S-Detect results were blinded in this session.

The change of category after combining S-Detect or SWE to B-mode US depended on the subjective judgment of the radiologists. It could be an upgrade, a downgrade, or maintain the original chosen category. For example, if a lesion was assigned as BI-RADS category 2 or 3 on B-mode US, but showed “possibly malignant” on S-Detect or hard elasticity (maximum stiffness color of lesion: orange to red, maximum elasticity of lesion greater than 135kPa) on SWE, the category could be upgraded to 4 or higher or kept at 3. Contrariwise, if a lesion was assigned as BI-RADS category 4 on B-mode US, but showed “possibly benign” on S-Detect or soft elasticity (maximum stiffness color of lesion: blue, maximum elasticity of lesion less than 30kPa) on SWE, the category could be downgraded to 2 or 3 or kept at 4.

#### 2.4. Data and statistical analysis

Clinical data including age, menopausal status, family history of breast cancer, mammographic density, and US data including size at US, the dichotomized assessment of S-Detect, maximum elasticity, mean elasticity, and lesion-to-fat elasticity ratio of SWE and final pathologic data were collected from medical records.

To differentiate benign and malignant breast lesions, area under the receiver operator characteristics (ROC) curve (AUC), sensitivity and specificity were calculated for S-Detect and SWE parameters. For quantitative SWE parameters, the optimal cut-off value was determined by maximizing the sum of sensitivity and specificity on the estimated ROC curve (the Youden index). SWE parameters were also compared between benign and malignant breast lesions by using the independent t-test.

The final assessment categories of each mass after the radiologist’s review were dichotomized for the statistical analysis. The cutoff for separating benign and malignant masses was set at category 4a, that is, category 2 and 3 were considered as benign, and categories 4a to 5 were considered as malignant. AUC, sensitivity and specificity of B-mode US, B-mode US plus S-Detect and B-mode US plus SWE were calculated for each reviewer data and compared using Delong’s method for AUC and McNemar test for sensitivity and specificity. Fleiss kappa statistics were used to analyze the interobserver agreement on each review session. Kappa value <0 indicated poor agreement; 0 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1.00, almost perfect agreement.<sup>[29]</sup> Statistical analyzes were performed using MedCalc Statistical Software version 19.1.3 (MedCalc Software bv, Ostend, Belgium; <https://www.medcalc.org>; 2019) and STATA software version 14.0 (StataCorp, College Station, TX). *P* value of less than .05 indicated statistical significance.

### 3. Results

#### 3.1. Demographics

A total of 156 breast lesions in 146 women (mean age, 46 ± 10 years; ranges, 23–74) who met the inclusion and exclusion criteria were analyzed in this study (Table 1). From these participants, 10 (7%) had 2 breast lesions and the other 136 (93%) had one breast lesion. Of total 146 women, 130 (89%) had dense breast and 16 (11%) had fatty breast on mammography. All women had BI-RADS 1 or 2 on mammography. The

**Table 1**

#### Patients characteristics.

Characteristics	Value
Number of patients	146
Number of breast lesions	156
Age (yr)	
Mean ± standard deviation	46 ± 10
Ranges	23–74
Menopause	
Pre-	104 (71)
Post-	42 (29)
Family history	
No	130 (89)
Yes	16 (11)
Mammography density	
Fatty	16 (11)
Dense	130 (89)
Size on ultrasound (cm)	
Mean ± standard deviation	1.1 ± 0.5
Ranges	0.3 - 3.4
Biopsy results	
Benign	146 (94)
Malignancy	10 (6)

— Data in parentheses are percentages.

mean size of breast lesions at US was 1.1 cm with ranges from 0.3 cm to 3.4 cm. The lesions consisted of 10 malignant lesions (8 invasive ductal carcinomas (IDC) and 2 ductal carcinomas in situ [DCIS]) and 146 benign lesions. The benign lesions were characterized as follows: 58 fibroadenomas, 21 fibrocystic changes, 14 intraductal papillomas, 13 fibroadenomatoid changes, 13 sclerosing adenoses, 5 stromal fibroses, 4 atypical ductal hyperplasias, 4 usual ductal hyperplasias, 4 duct ectasias, 3 benign phyllodes tumors, 2 atypical ductal hyperplasias involving intraductal papillomas, 2 fat necroses, 1 nodular adenosis, 1 cholesterol granuloma, and 1 adenomyoepithelioma.

#### 3.2. S-Detect and shear wave elastography to predict malignancy

The individual performances of S-Detect and SWE parameters to predict malignancy are provided in Table 2. Of 10 malignancies, only 3 had a “possibly malignant” assessment on S-Detect, which yielded 30% sensitivity, 84.9% specificity and 0.575 of AUC. Orange or red maximum stiffness color on SWE had 50% sensitivity, 90.4% specificity, and 0.702 of AUC. Maximum elasticity (mean, 83.9 kPa vs 44.6 kPa, *P* < .001), mean elasticity (mean, 79.7 kPa vs 41.3 kPa, *P* < .001), and lesion-to-fat elasticity ratio (mean, 881.3% vs 481.5%, *P* = .029) of malignant breast lesions were significantly higher than those of benign breast lesions. Maximum elasticity with a cut-off value of 44.5 kPa had 90% sensitivity, 61.6% specificity, and 0.784 of AUC. Mean elasticity with a cut-off value of 39.8 kPa had 90% sensitivity, 61.6% specificity, and 0.783 of AUC. Lesion-to-fat elasticity ratio with a cut-off value of 304.5% had 90% sensitivity, 48.6% specificity, and 0.675 of AUC.

#### 3.3. Added value of S-Detect or shear wave elastography to B-mode ultrasound

AUC, sensitivity, and specificity of B-mode US, B-mode US plus S-Detect, and B-mode US plus SWE are provided in Table 3. ROC

**Table 2**  
**S-Detect and SWE parameters to predict malignancy.**

	Benign (n = 146)	Malignant (n = 10)	AUC	Sensitivity (%)	Specificity (%)
S-Detect final assessment, n (%)					
Possibly benign	124 (95)	7 (5)	0.575	30 (3/10)	84.9 (124/146)
Possibly malignant	22 (88)	3 (12)			
SWE maximum stiffness color, n (%)					
Blue or Green	132 (96)	5 (3.6)	0.702	50 (5/10)	90.4 (132/146)
Orange or Red	14 (74)	5 (26)			
SWE maximum elasticity (kPa), Mean ± SD	44.6 ± 31.2	83.9 ± 42.5	0.784	90 (9/10)	61.6 (90/146)
SWE mean elasticity (kPa), Mean ± SD	41.3 ± 30.1	79.7 ± 41.4	0.783	90 (9/10)	61.6 (90/146)
SWE lesion-to-fat elasticity ratio (%), Mean ± SD	481.5 ± 533.7	881.3 ± 833.3	0.675	90 (9/10)	48.6 (71/146)

— For sensitivity and specificity, data in parentheses are numerator and denominator. AUC = area under the receiver operating characteristic curve, SD = standard deviation, SWE = shear-wave elastography.

**Table 3**  
**Added value of S-Detect or SWE to B-mode US.**

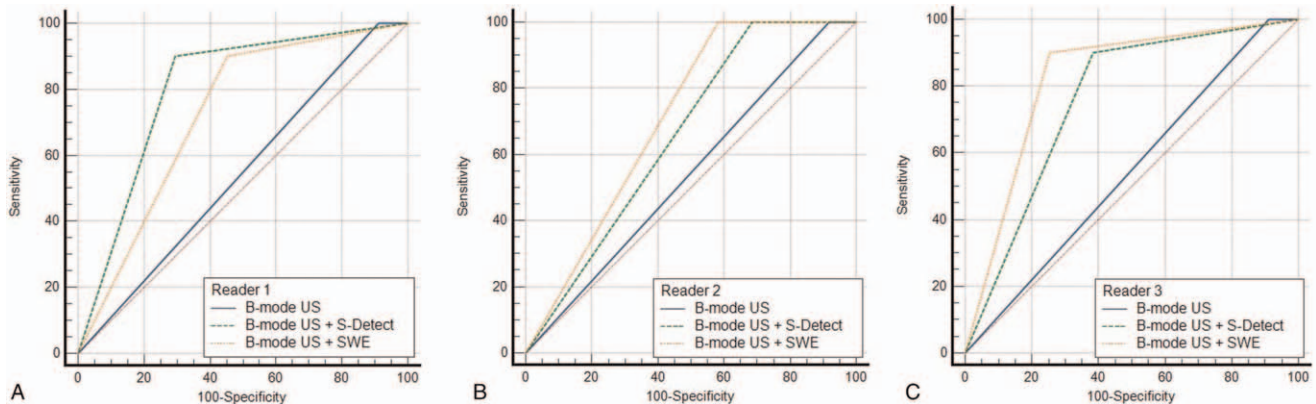
	Reader 1			Reader 2			Reader 3		
	AUC	Sensitivity (%)	Specificity (%)	AUC	Sensitivity (%)	Specificity (%)	AUC	Sensitivity (%)	Specificity (%)
B-mode US	0.545	100 (10/10)	8.9 (13/146)	0.541	100 (10/10)	8.2 (12/146)	0.545	100 (10/10)	8.9 (13/146)
B-mode US plus S-Detect	0.803	90 (9/10)	70.5 (103/146)	0.658	100 (10/10)	31.5 (46/146)	0.758	90 (9/10)	61.6 (90/146)
B-mode US plus SWE	0.724	90 (9/10)	54.8 (80/146)	0.709	100 (10/10)	41.8 (61/146)	0.823	90 (9/10)	74.7 (109/146)
<i>P</i> value <sup>1</sup>	<.001	>.999	<.001	<.001	NA	<.001	<.001	>.999	<.001
<i>P</i> value <sup>2</sup>	<.001	>.999	<.001	<.001	NA	<.001	<.001	>.999	<.001
<i>P</i> value <sup>3</sup>	.313	>.999	.001	.004	NA	.008	<.001	NA	<.001

— Data in parentheses are numerator and denominator. AUC = area under the receiver operating characteristic curve, NA = not applicable, SWE = shear-wave elastography. *P* value<sup>1</sup>: comparison between B-mode US and B-mode US plus S-Detect. *P* value<sup>2</sup>: comparison between B-mode US and B-mode US plus SWE. *P* value<sup>3</sup>: comparison between B-mode US plus S-Detect and B-mode US plus SWE.

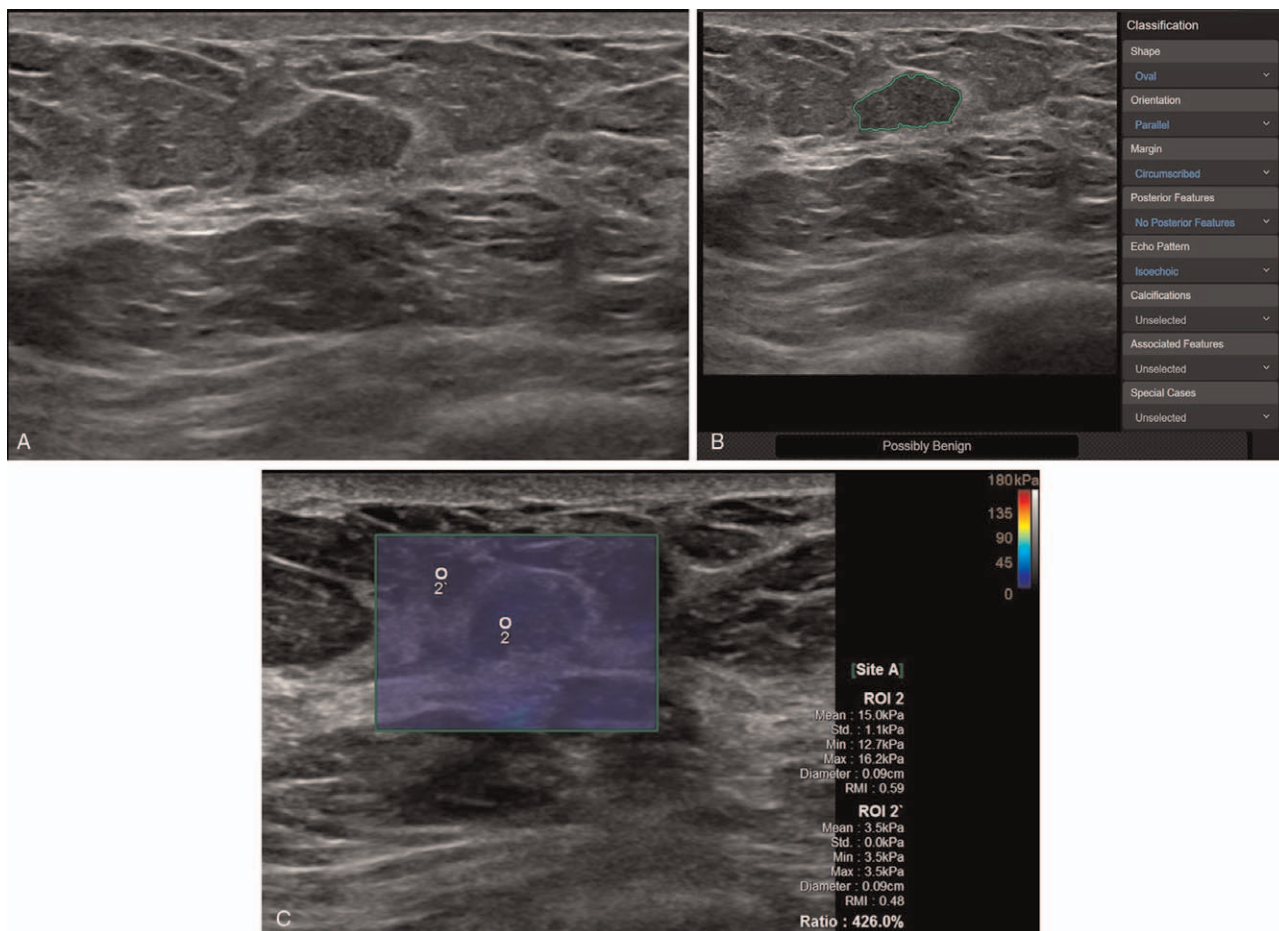
curves are compared in Fig. 1. With B-mode US alone, all readers showed 100% (10 out of 10) sensitivity. However, specificity was 8.9% (13 out of 146) in reader 1 and reader 3 and 8.2% (12 out of 146) in reader 2. AUC was 0.545 in reader 1 and reader 3 and 0.541 in reader 2.

When adding S-Detect to B-mode US, the specificity significantly increased in all readers compared to B-mode US alone

(8.9% [13 of 146] vs 70.5% [103 of 146], *P* < .001 in reader 1; 8.2% [12 of 146] vs 31.5% [46 of 146], *P* < .001 in reader 2; 8.9% [13 of 146] vs 61.6% [90 of 146], *P* < .001 in reader 3) (Fig. 2). Reader 2 did not miss cancer whereas reader 1 and reader 3 missed 1 malignancy, albeit without a statistically significant difference in sensitivity (100% vs 90%, *P* > .999). The AUC significantly increased in all readers (0.545 vs 0.803, *P* < .001 in



**Figure 1.** The receiver operator characteristics (ROC) curves comparing B-mode US, B-mode US plus S-Detect, and B-mode US plus shear wave elastography (SWE) in reader 1 (A), reader 2 (B), and reader 3 (C), respectively.



**Figure 2.** A fibroadenoma correctly downgraded by adding S-Detect and shear wave elastography (SWE). (A) B-mode US in a 52-year-old woman shows a 1.1 cm breast mass with Breast Imaging Reporting and Data Systems (BI-RADS) final assessment category 4A by all 3 radiologists. (B) The mass is assessed as “possibly benign” on S-Detect. (C) The mass shows dark blue maximum stiffness color with maximum elasticity of 16.2kPa. All radiologists downgraded this mass to category 3 when S-Detect and SWE were added to B-mode US. The mass was confirmed as fibroadenoma on US-guided 14-gauge core needle biopsy and was stable on the 12-month follow-up ultrasound.

reader 1, 0.541 vs 0.658,  $P < .001$  in reader 2, 0.545 vs 0.758,  $P < .001$  in reader 3).

The addition of SWE to B-mode US, also increased specificity in all readers compared to B-mode US alone (8.9% [13 of 146] vs 54.8% [80 of 146],  $P < .001$  in reader 1; 8.2% [12 of 146] vs 41.8% [61 of 146],  $P < .001$  in reader 2; 8.9% [13 of 146] vs 74.7% [109 of 146],  $P < .001$  in reader 3) (Fig. 2). Reader 2 did not miss cancer whereas reader 1 and reader 3 missed 1 malignancy, without a statistically significant loss in sensitivity (100% vs 90%,  $P > .999$ ). Accordingly, the AUC significantly increased in all readers (0.545 vs 0.724,  $P < .001$  in reader 1, 0.541 vs 0.709,  $P < .001$  in reader 2, 0.545 vs 0.823,  $P < .001$  in reader 3).

When comparing B-mode US plus SWE and B-mode US plus S-Detect, the specificity (41.8% [61 of 146] vs 31.5% [46 of 146],  $P = .008$  in reader 2; 74.7% [109 of 146] vs 61.6% [90 of 146],  $P < .001$  in reader 3) and AUC (0.709 vs 0.658,  $P = .004$  in reader 2; 0.823 vs 0.758,  $P < .001$  in reader 3) of B-mode US plus SWE were significantly higher than those of B-mode US plus S-Detect in readers 2 and 3. However, in reader 1, the specificity of B-mode US plus S-Detect was significantly higher

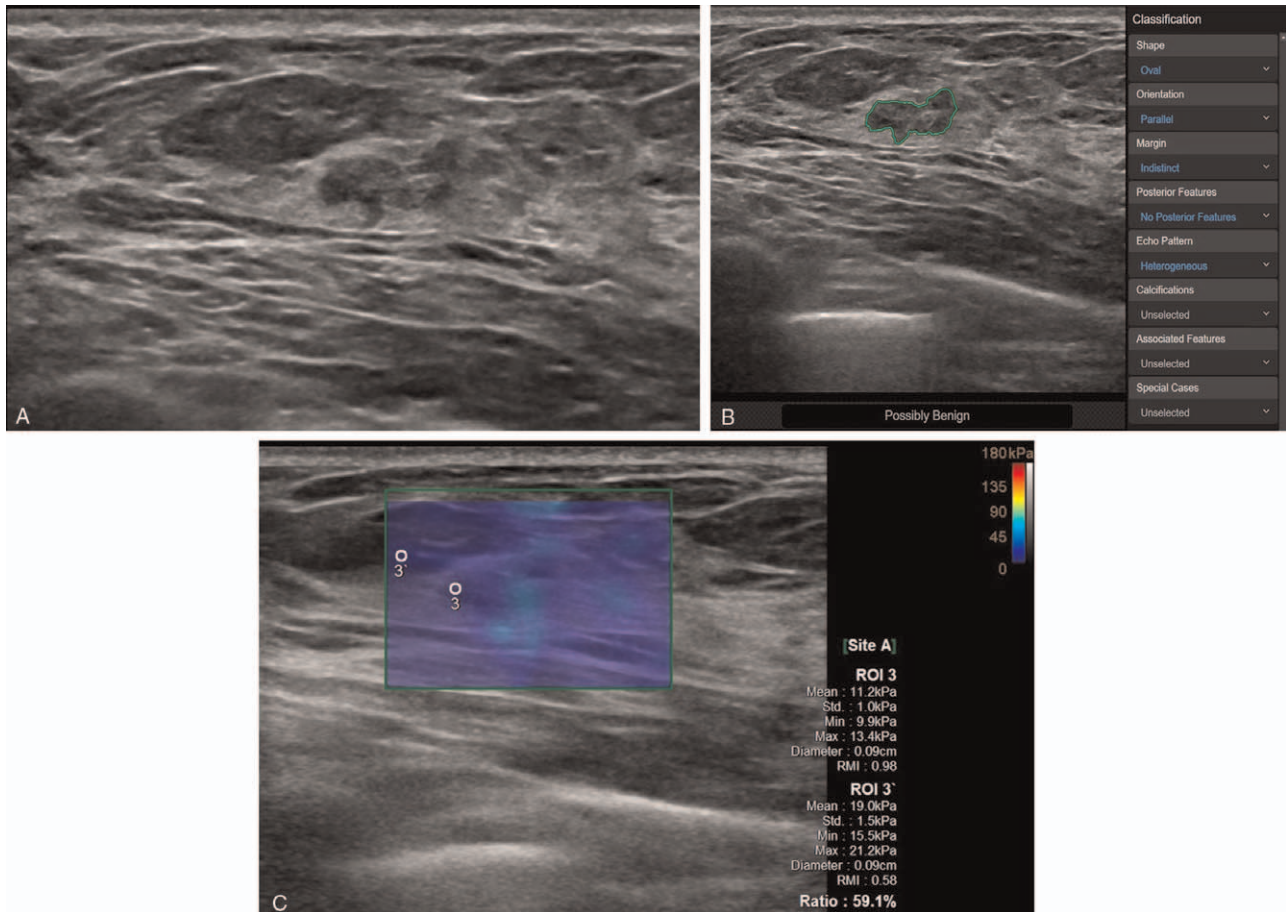
than that of B-mode US plus SWE (70.5% [103 of 146] vs 54.8% [80 of 146],  $P = .001$ ) without significant difference in AUC (0.803 vs 0.724,  $P = .313$ ).

### 3.4. Interobserver variability on each review session

Interobserver agreement among the 3 radiologists were as follows. In the B-mode US session, kappa value was 0.132 (95% confidence interval [CI]: 0.007, 0.214), indicating slight agreement. In the B-mode plus S-Detect session, kappa value was 0.324 (95% CI: 0.233–0.414), indicating fair agreement. In the B-mode plus SWE session, kappa value was 0.355 (95% CI: 0.265–0.446), indicating fair agreement.

### 3.5. Pathological and ultrasound characteristics of malignancy

There were 10 pathologically proven malignancies consisting of 8 IDC and 2 DCIS. All invasive cancers were node-negative, low or intermediate grade, and estrogen receptor-positive/human epidermal growth factor receptor type 2- negative



**Figure 3.** A ductal carcinoma in situ with false-negative assessment by adding shear wave elastography (SWE). (A) B-mode US in a 59-year-old woman shows a 1 cm breast mass with Breast Imaging Reporting and Data Systems (BI-RADS) category 4A by all 3 radiologists. (B) S-Detect interprets this mass as “possibly benign,” but the category was maintained by all radiologists. (C) On SWE, majority of the mass shows dark blue color with a small part of light blue color. One radiologist (Reader 1) downgraded this mass to category 3 according to the SWE results.

subtype. The median size on B-mode US was 1.1 cm (ranges, 0.7–1.8 cm) and the median size on surgical pathology was 1.5 cm (ranges, 1–4 cm).

There were 2 false negative cases when adding S-Detect or SWE to B-mode US. One malignancy (1 cm DCIS with low grade) was falsely downgraded to category 3 when SWE was added to B-mode US in reader 1, due to blue color of SWE (Fig. 3). The other (0.9 cm IDC with low grade) was downgraded to category 3 when S-detect was added to B-mode US in reader 1 and reader 3 due to “probably benign” assessment of S-Detect, and when SWE was added to B-mode US in reader 3 due to green color of SWE (Fig. 4).

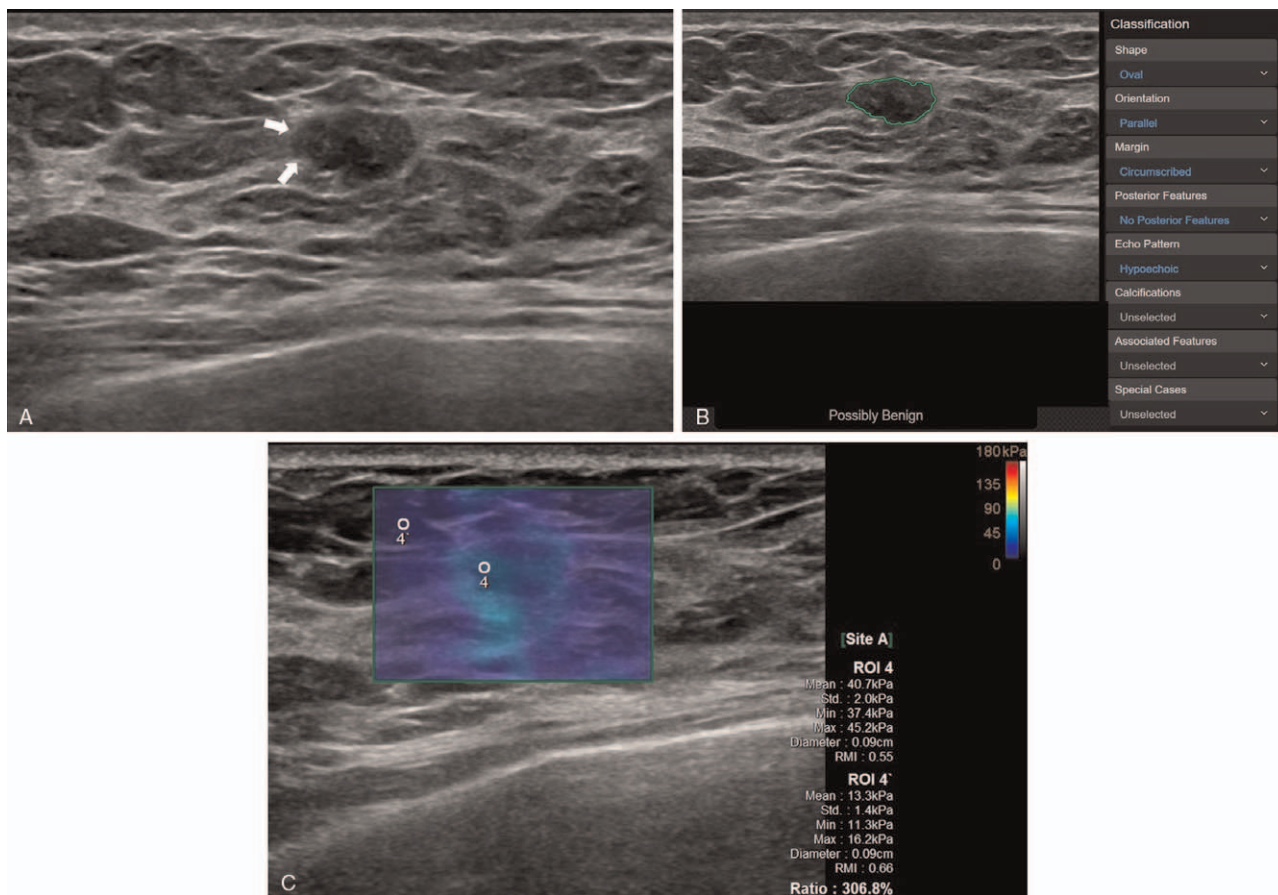
When S-Detect was added to B-mode US, 3 malignancies had upgraded category within category 4 by all 3 readers due to the “possibly malignant” assessment of S-Detect (Figs. 4 and 5). When SWE was added to B-mode US, 3 malignancies had upgraded category within category 4 by all 3 readers due to orange to red color of SWE (Figs. 5 and 6).

#### 4. Discussion

Our study found that using S-Detect or SWE had added value in the differential diagnosis of breast masses detected by screening

US. Compared to using conventional B-mode US alone, the additional use of S-Detect or SWE significantly improved specificity and AUC without losing the sensitivity of all radiologists. In addition, when adding S-Detect or SWE to B-mode US, interobserver agreement between 3 radiologists with different levels of experience improved from slight to fair agreement. These results were concordant with prior literature which applied S-Detect or SWE in the different study settings.<sup>[8,12,14,20,21,25]</sup>

In our study, the preferred method as a second opinion provider differed depending on the radiologists. Differences in experiences and image review settings may have affected these different opinions. Reader 1 with the highest experiences in breast US and SWE and in the prospective real-time analysis setting preferred S-Detect over SWE: the specificity and AUC of B-mode US plus S-Detect were higher than those of B-mode US plus SWE with statistical significance for specificity in reader 1. Whereas, reader 2 and 3 with relatively lower experience in breast US and SWE and in the retrospective analysis setting preferred SWE over S-Detect: the specificity and AUC of B-mode US plus SWE were significantly higher than those of B-mode US plus S-Detect in readers 2 and 3.



**Figure 4.** An invasive ductal carcinoma with false-negative assessment when adding S-Detect and shear wave elastography (SWE). (A) B-mode US in a 55-year-old woman shows a 0.9cm breast mass with partial microlobulations and Breast Imaging Reporting and Data Systems (BI-RADS) category 4A by all 3 radiologists. (B) S-Detect misinterprets this mass as “possibly benign”. Two radiologists (Reader 1 and Reader 3) downgraded this mass to category 3 according to the S-Detect results. (C) On SWE, the mass shows green color with maximum elasticity of 45.2kPa. One radiologist (Reader 3) downgraded the mass to category 3 according to the SWE results.

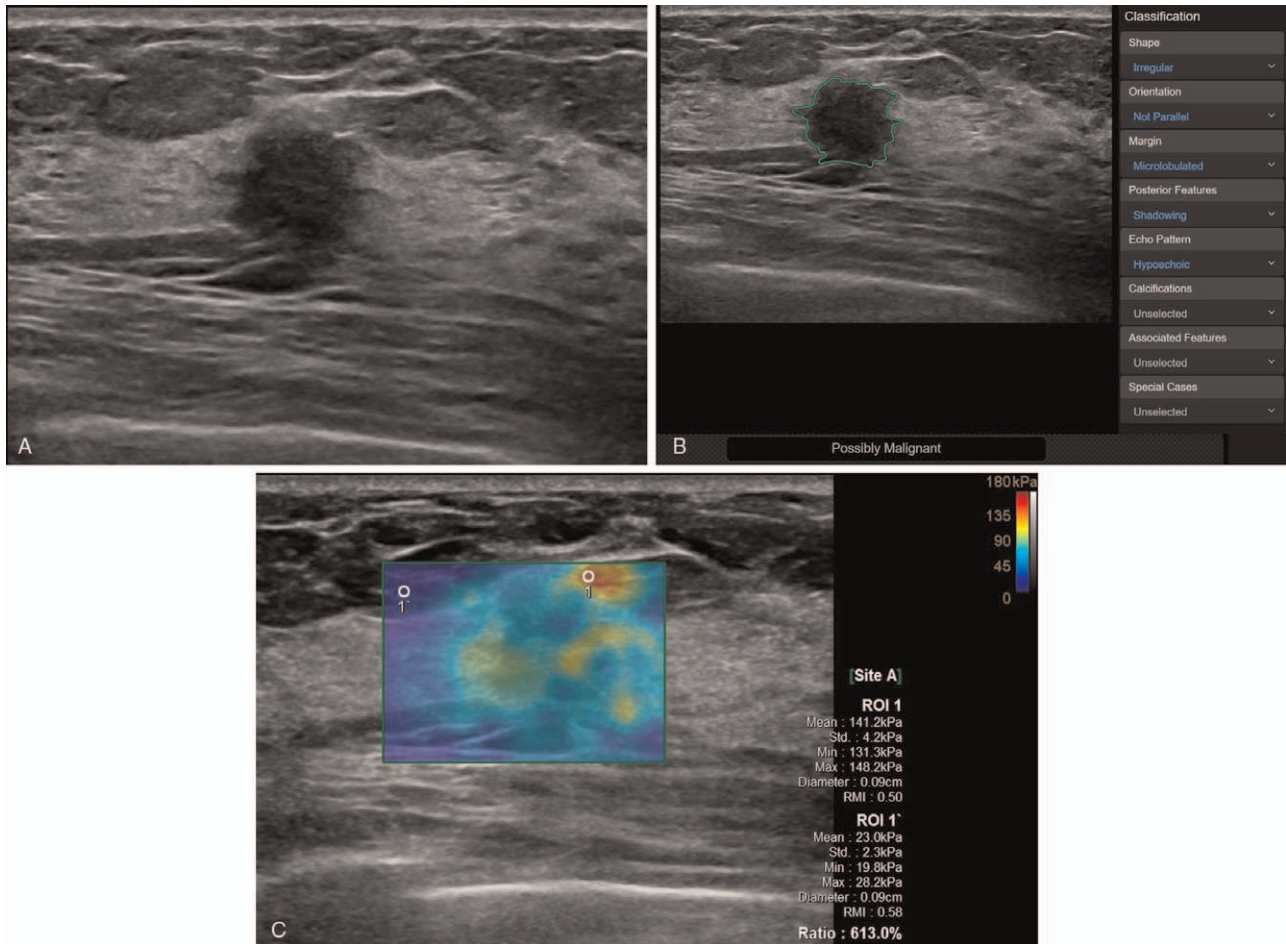
With the recent advance of deep learning technology, the deep learning algorithms have been used in various field of US including image acquisition, processing, and interpretation. Elasticity-based state-space approach was accurate in measuring the two-dimensional motion of the carotid artery.<sup>[30]</sup> A deep learning-based framework was effective for the strain reconstruction of elastography.<sup>[31]</sup> S-Detect, a deep learning-based CAD software, has shown its ability as a useful diagnostic tool that can improve the accuracy and specificity in evaluating breast masses for both novice and experienced readers.<sup>[18–25]</sup>

Compared to previous studies using S-Detect, the different point of our study was that we compared the diagnostic performances of B-mode US alone, B-mode US plus SWE, and B-mode US plus S-Detect focusing on breast masses detected by screening US. Most breast cancers detected by screening US are small in size, without calcification, and lack of typical malignant US features.<sup>[26,27]</sup> Because of these properties, differentiating screening US-detected breast masses as benign or malignant can be challenging and this may reduce the accuracy of screening US. Based on our results, the combined use of S-Detect or SWE to B-

mode US could be helpful in reducing false-positive biopsy without missing cancers.

Of the 10 malignancies in our study, 80% (8 of 10) were invasive and all were node-negative. All invasive cancers were low or intermediate grade and estrogen receptor-positive/human epidermal growth factor receptor type 2- negative subtype. The median size on surgical pathology was 1.5 cm. These favorable histologic findings are consistent with the reported pathologic features of breast cancer detected by screening US.<sup>[1,2]</sup>

When adding S-Detect or SWE to B-mode US, false-negative assessments occurred in 2 readers although there was no statistically significant loss in sensitivity. When S-Detect was added to B-mode US, 1 cancer was missed by 2 readers, because S-Detect misinterpreted the microlobulated margin that exists at a portion of the cancer as a circumscribed margin. It suggests the difficulty of accurate evaluation of margins in small breast masses, and the suspicious margin features identified on B-mode US should not be overlooked.<sup>[32]</sup> When SWE was added to B-mode US, 2 cancers were missed, because the lesions showed blue or green as the maximum stiffness color. These results were



**Figure 5.** An invasive ductal carcinoma with upgraded category when adding S-Detect and shear wave elastography (SWE). (A) B-mode US in a 44-year-old woman shows a 1.2 cm irregular and not-parallel mass with Breast Imaging Reporting and Data Systems (BI-RADS) category 4B by all 3 radiologists. (B) S-Detect correctly interprets this mass as “possibly malignant”. (C) On shear wave elastography (SWE), the mass shows red color with maximum elasticity of 148.2 kPa. When S-Detect and SWE were added to B-mode US, all radiologists had more confidence in malignancy diagnosis, upgrading this mass to BI-RADS category 4C.

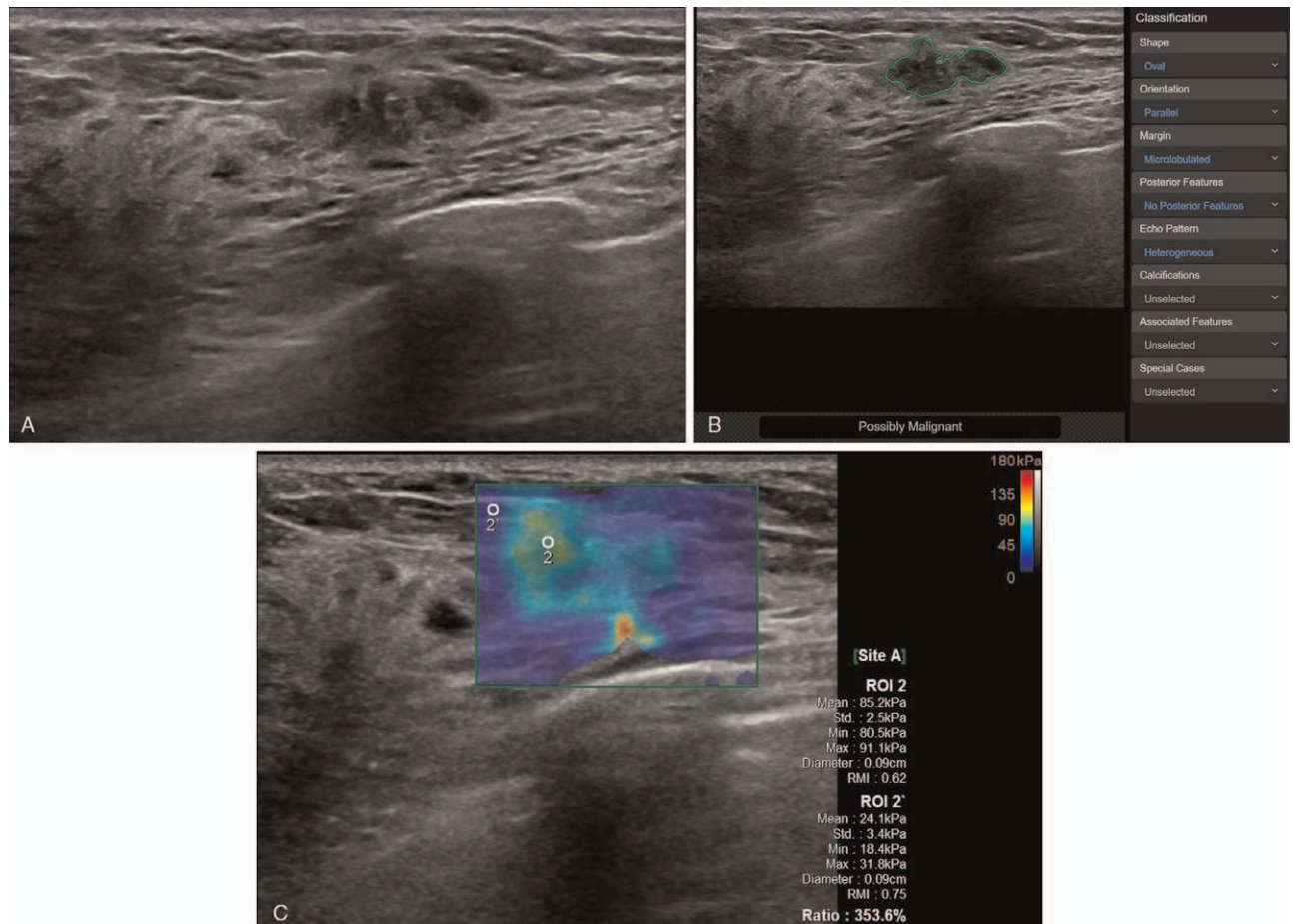
concordant with the previous study in that small, low grade, and screening detected cancers tend to show soft elasticity.<sup>[33]</sup>

Among quantitative SWE parameters, maximum and mean elasticity showed high accuracy with AUC of 0.783 to 0.784 in differentiating benign and malignant breast lesions, concordant with the prior study.<sup>[12,34,35]</sup> With regard to the dichotomized final assessment of S-Detect, S-Detect misinterpreted 7 cancers as “possibly benign”. The sensitivity reported in our study (30%) was lower than that reported in the previous studies (72%–91%).<sup>[18–21]</sup> An explanation for this may be the difference in inclusion criteria of breast masses and the number of malignancies in the studies. Previous studies included symptomatic breast masses as well, while our study included only asymptomatic breast masses detected by screening US. This indicates that breast masses analyzed in this study were challenging cases with a small size and lack of suspicious US features. In addition, the small number of malignancies in our study may account for the low sensitivity of S-Detect. In order to safely use S-Detect without missing cancers in a screening setting, it is necessary to improve the diagnostic ability of S-Detect for small asymptomatic breast masses through deep learning training.

Our study had limitations. First, as a single center study, all of B-mode US, SWE and S-Detect images were acquired by 1 radiologist, thus the reproducibility may be limited. Second, in terms of SWE application, the quantitative SWE parameters could differ according to the ROI location and size. The superiority of a small ROI approach used in our study has been demonstrated in previous studies,<sup>[36,37]</sup> but a further clinical study including various radiologists as readers would be necessary to validate the reproducibility. Third, for reader 2 and reader 3, US images were retrospectively reviewed using the representative still images, not the real-time review. Fourth, the number of cancers included in this study were small (n=10). However, it is known that the cancer detection rate of screening US in women with negative mammography is low as approximately 3 to 4 cancers per 1000 US examinations.<sup>[1–4]</sup>

In conclusion, deep learning-based CAD and SWE may be used as supplemental diagnostic tools to improve the specificity and AUC of screening breast US without a statistically significant loss of sensitivity in asymptomatic women with negative mammography, regardless of radiologists’ experience.





**Figure 6.** An invasive ductal carcinoma with upgraded category when adding S-Detect and shear wave elastography (SWE). (A) B-mode US in a 49-year-old woman shows a 1.5 cm mildly irregular and parallel mass with BI-RADS category 4A by 2 radiologists and 4B by 1 radiologist. (B) S-Detect correctly classifies this mass as “possibly malignant”. (C) The mass shows orange color with maximum elasticity of 91.1 kPa on SWE. All radiologists upgraded BI-RADS category by 1 level when adding S-Detect and SWE.

## Author contributions

**Conceptualization:** Soo-Yeon Kim, Jung Min Chang.

**Funding acquisition:** Jung Min Chang.

**Investigation:** Min Young Kim, Soo-Yeon Kim.

**Methodology:** Min Young Kim, Soo-Yeon Kim, Jung Min Chang.

**Supervision:** Soo-Yeon Kim.

**Writing – original draft:** Min Young Kim, Soo-Yeon Kim.

**Writing – review & editing:** Soo-Yeon Kim, Yeon Soo Kim, Eun Sil Kim, Jung Min Chang.

## References

- [1] Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* 2008;299:2151–63.
- [2] Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09–41. *Radiology* 2012;265:59–69.
- [3] Chang JM, Koo HR, Moon WK. Radiologist-performed hand-held ultrasound screening at average risk of breast cancer: results from a single health screening center. *Acta Radiol* 2015;56:652–8.
- [4] Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. *Am J Obstet Gynecol* 2015;212:9–17.
- [5] Berg WA, Blume JD, Cormack JB, Mendelson EB. Operator dependence of physician-performed whole-breast US: lesion detection and characterization. *Radiology* 2006;241:355–65.
- [6] Bosch AM, Kessels AG, Beets GL, et al. Interexamination variation of whole breast ultrasound. *Br J Radiol* 2003;76:328–31.
- [7] Merry GM, Mendelson EB. Update on screening breast ultrasonography. *Radiol Clin North Am* 2014;52:527–37.
- [8] Cho N, Jang M, Lyou CY, Park JS, Choi HY, Moon WK. Distinguishing benign from malignant masses at breast US: combined US elastography and color doppler US—influence on radiologist accuracy. *Radiology* 2012;262:80–90.
- [9] Lee SH, Chang JM, Cho N, et al. Practice guideline for the performance of breast ultrasound elastography. *Ultrasonography* 2014;33:3–10.
- [10] Mun HS, Choi SH, Kook SH, Choi Y, Jeong WK, Kim Y. Validation of intra- and interobserver reproducibility of shearwave elastography: Phantom study. *Ultrasonics* 2013;53:1039–43.
- [11] Hong MJ, Kim HH. Intra- and interobserver reproducibility of shear wave elastography for evaluation of the breast lesions. *J Korean Soc Radiol* 2017;76:198–205.
- [12] Berg WA, Cosgrove DO, Dore CJ, et al. Shear-wave elastography improves the specificity of breast US: the BE1 multinational study of 939 masses. *Radiology* 2012;262:435–49.
- [13] Au FW, Ghai S, Moshonov H, et al. Diagnostic performance of quantitative shear wave elastography in the evaluation of solid breast

- masses: determination of the most discriminatory parameter. *Am J Roentgenol* 2014;203:W328–336.
- [14] Lee SH, Chang JM, Kim WH, et al. Added value of shear-wave elastography for evaluation of breast masses detected with screening US imaging. *Radiology* 2014;273:61–9.
- [15] Jalalian A, Mashohor SB, Mahmud HR, Saripan MI, Ramli AR, Karasfi B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clin Imaging* 2013;37:420–6.
- [16] Dromain C, Boyer B, Ferre R, Canale S, Delaloue S, Balleyguier C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur J Radiol* 2013;82:417–23.
- [17] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
- [18] Kim K, Song MK, Kim EK, Yoon JH. Clinical application of S-Detect to breast masses on ultrasonography: a study evaluating the diagnostic performance and agreement with a dedicated breast radiologist. *Ultrasonography* 2017;36:3–9.
- [19] Di Segni M, de Soccio V, Cantisani V, et al. Automated classification of focal breast lesions according to S-Detect: validation and role as a clinical and teaching tool. *J Ultrasound* 2018;21:105–18.
- [20] Choi JH, Kang BJ, Baek JE, Lee HS, Kim SH. Application of computer-aided diagnosis in breast ultrasound interpretation: improvements in diagnostic performance according to reader experience. *Ultrasonography* 2018;37:217–25.
- [21] Cho E, Kim EK, Song MK, Yoon JH. Application of computer-aided diagnosis on breast ultrasonography: evaluation of diagnostic performances and agreement of radiologists according to different levels of experience. *J Ultrasound Med* 2018;37:209–16.
- [22] Bartolotta TV, Orlando A, Cantisani V, et al. Focal breast lesion characterization according to the BI-RADS US lexicon: role of a computer-aided decision-making support. *Radiol Med* 2018;123:498–506.
- [23] Lee J, Kim S, Kang BJ, Kim SH, Park GE. Evaluation of the effect of computer aided diagnosis system on breast ultrasound for inexperienced radiologists in describing and determining breast lesions. *Med Ultrason* 2019;21:239–45.
- [24] Yongping L, Zhou P, Juan Z, Yongfeng Z, Liu W, Shi Y. Performance of computer-aided diagnosis in ultrasonography for detection of breast lesions less and more than 2cm: prospective comparative study. *JMIR Med Inform* 2020;8:e16334.
- [25] Park HJ, Kim SM, Yun BL, et al. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of breast masses on ultrasound: added value for the inexperienced breast radiologist. *Medicine (Baltimore)* 2019;98:e14146.
- [26] Cho N, Moon WK, Chang JM, Yi A, Koo HR, Han BK. Sonographic characteristics of breast cancers detected by supplemental screening US: Comparison with breast cancers seen on screening mammography. *Acta Radiol* 2010;51:969–76.
- [27] Bae MS, Han W, Koo HR, et al. Characteristics of breast cancers detected by ultrasound screening in women with negative mammograms. *Cancer Sci* 2011;102:1862–7.
- [28] Han S, Kang HK, Jeong JY, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys Med Biol* 2017;62:7714–28.
- [29] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [30] Gao Z, Xiong H, Liu X, et al. Robust estimation of carotid artery wall motion using the elasticity-based state-space approach. *Med Image Anal* 2017;37:1–21.
- [31] Gao Z, Wu S, Liu Z, et al. Learning the implicit strain reconstruction in ultrasound elastography using privileged information. *Med Image Anal* 2019;58:101534.
- [32] Abdullah N, Mesurole B, El-Khoury M, Kao E. Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses. *Radiology* 2009;252:665–72.
- [33] Vinnicombe SJ, Whelehan P, Thomson K, et al. What are the characteristics of breast cancers misclassified as benign by quantitative ultrasound shear wave elastography? *Eur Radiol* 2014;24:921–6.
- [34] Lee EJ, Jung HK, Ko KH, Lee JT, Yoon JH. Diagnostic performances of shear wave elastography: which parameter to use in differential diagnosis of solid breast masses? *Eur Radiol* 2013;23:1803–11.
- [35] Xue Y, Yao S, Li X, Zhang H. Value of shear wave elastography in discriminating malignant and benign breast lesions: a meta-analysis. *Medicine (Baltimore)* 2017;96:e7412.
- [36] Moon JH, Hwang JY, Park JS, Koh SH, Park SY. Impact of region of interest (ROI) size on the diagnostic performance of shear wave elastography in differentiating solid breast lesions. *Acta Radiol* 2018;59:657–63.
- [37] Skerl K, Vinnicombe S, Giannotti E, Thomson K, Evans A. Influence of region of interest size and ultrasound lesion size on the performance of 2D shear wave elastography (SWE) in solid breast masses. *Clin Radiol* 2015;70:1421–7.