

Evaluation of the effect of computer aided diagnosis system on breast ultrasound for inexperienced radiologists in describing and determining breast lesions

Jeongmin Lee, Sanghee Kim, Bong Joo Kang, Sung Hun Kim, Ga Eun Park

Departments of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea

Abstract

Aim: To investigate the effect of a computer-aided diagnosis (CAD) system on breast ultrasound (US) for inexperienced radiologists in describing and determining breast lesions. **Materials and methods:** Between October 2015 to January 2017, 500 suspicious or probable benign lesions in 413 patients were reviewed. Five experienced readers retrospectively reviewed for each of 100 lesions according to the Breast Imaging Reporting and Data System (BI-RADS) lexicon and category, with CAD system (S-detectTM). The readers then made final decisions by combining CAD results to their US results. Using the nested experiment design, five inexperienced readers were asked to select the appropriate BI-RADS lexicons, categories, CAD results, and combination results for each of the 100 lesions, retrospectively. Diagnostic performance of experienced and inexperienced radiologists and CAD were assessed. For each case, agreements in the lexicons and categories were analyzed among the experienced reader, inexperienced reader and CAD. **Results:** Indicators of the diagnostic performance for breast malignancy of the experienced group (AUC=0.83, 95%CI [0.80, 0.86]) were similar or higher than those of CAD (AUC = 0.79, 95%CI[0.74, 0.83], $p=0.101$), except for specificity. Conversely, indicators of diagnostic performance of inexperienced group (AUC=0.65, 95%CI[0.58, 0.71]) did not differ from or were lower than those of CAD(AUC=0.73, 95%CI[0.67, 0.78], $p=0.013$). Also, the diagnostic performance of the inexperienced group after combination with the CAD result was significantly improved (0.71, 95% CI [0.65, 0.77], $p=0.001$), whereas that of the experienced group did not change after combination with the CAD result, except for specificity and positive predictive value (PPV). Kappa values for the agreement of the categorization between CAD and each radiologist group were increased after applying the CAD result to their result of general US. Especially, the increase of the Kappa value was higher in the inexperienced group than in the experienced group. Also, for all the lexicons, the Kappa values between the experienced group and CAD were higher than those between the inexperienced group and CAD. **Conclusion:** By using the CAD system for classification of breast lesions, diagnostic performance of the inexperienced radiologists for malignancy was significantly improved, and better agreement was observed in lexicons between the experienced group and CAD than between the inexperienced group and CAD. CAD may be beneficial and educational for the inexperienced group.

Keywords: ultrasound; computer-aided diagnosis; breast cancer

Received 20.02.2019 Accepted 04.06.2019

Med Ultrason

2019, Vol. 21, No 3, 239-245

Corresponding author: Bong Joo Kang, MD, PhD, Professor
Department of Radiology, Seoul St. Mary's
Hospital, College of Medicine,
The Catholic University of Korea
222, Banpo-daero, Seocho-gu, Seoul,
06591, Republic of Korea
Phone: 82-2-2258-6253
Fax: 82-2-599-6771
E-mail: gmliionmain@gmail.com

Introduction

In the development of medical technology, new advanced methods are emerging that actively assist in the diagnosis and treatment. The most representative method in the radiological field is the computer-aided detection or diagnosis (CAD) system. The first CAD system commercially approved by the Food and Drug Administration (FDA) in 1998, was to automatically detect and mark calcifications or masses in digital mammography

[1]. While several previous studies have suggested that CAD systems can contribute to increase the early cancer detection rates, the number of breast radiologists actively relying on this CAD system for mammography readings is small [2].

In addition to simplify the lesions' detection, the CAD systems for breast lesion have been introduced in magnetic resonance imaging (MRI) and ultrasound (US) techniques. MRI CAD system automatically analyzes the kinetics of enhancing lesions on breast MRI and it can contribute to improve the specificity and reduce the interpretation time by radiologists [1,3-5].

Breast US is one of the commonly used screening methods in clinical practice as a supplementary tool especially in women with dense breast tissue [6]. However, US is more operator dependent than other methods. For this reason, many previous studies have demonstrated the efficiency and usefulness of CAD systems for improving diagnostic accuracy in US [7-12]. In our previous study, a CAD system newly developed in Korea (S-Detect, Samsung Medison Co. Ltd., Seoul, Korea) was useful for less experienced radiologists and the combination of the CAD system and general US improved the specificity in inexperienced and experienced radiologists [13].

The purpose of this study was to investigate the effect of CAD (S-detect™) on breast US when inexperienced radiologists described and categorized breast lesions, especially in comparison with experienced radiologists.

Materials and methods

This retrospective study was approved by our Institutional Review Board (IRB) and all enrolled participants provided written informed consent. A total of 682 breast lesions from 566 patients examined by US between October 2015-January 2017 were consecutively obtained using the Samsung Ultrasound RS80A, Samsung Medison Co. Ltd., Seoul, Korea machine by five experts in breast radiologists, with more than 5 years' experience. The cases without pathologic confirmation (N = 19) and findings associated with post-operative changes (N = 163) such as post-operative fibrosis or fat necrosis, were excluded. Finally, the data of 500 suspicious or probable benign breast lesions from 413 patients were enrolled. All the lesions were examined using the CAD system (S-detect™) at the same time as general US. When radiologists identified the center of the breast lesion by touching the screen, a region of interest (ROI) was automatically drawn along the border of the lesion. Several drawn borders were presented on the screen and the radiologists saved more than two CAD images as with the general US.

Imaging analysis

All 500 cases were retrospectively analyzed by two radiologist groups. The first radiologist group comprised five experienced radiologists with clinical fellowship in breast imaging and more than five years' experience of breast US. Expert breast radiologists who performed US and extracted cases, were excluded from this group. The second radiologist group was composed with five inexperienced radiologists with less than one year of breast US.

Firstly, experienced and inexperienced groups described and categorized the lesions according to the Breast Imaging Reporting and Data System (BI-RADS) lexicon: shape, orientation, margin, echogenicity and acoustic shadowing. In the first conclusions using general US, description and categorization of lesions by CAD (S-detect™) were disclosed to two radiologist groups. The CAD system described the lesions according to the BI-RADS lexicons and they were classified in to two categories: "possibly benign" or "possibly malignant". Readers selected the most appropriate CAD image as with the general US. After referring to results by CAD, experienced and inexperienced groups were given the opportunity to re-categorize the lesions and they made final subjectively combined conclusion for the lesions (fig 1). The diagnostic performance of both radiologist groups and CAD were compared, and the gold standard of diagnosis was the pathologic report presented after surgery or the follow-up breast US realized more than one year later.

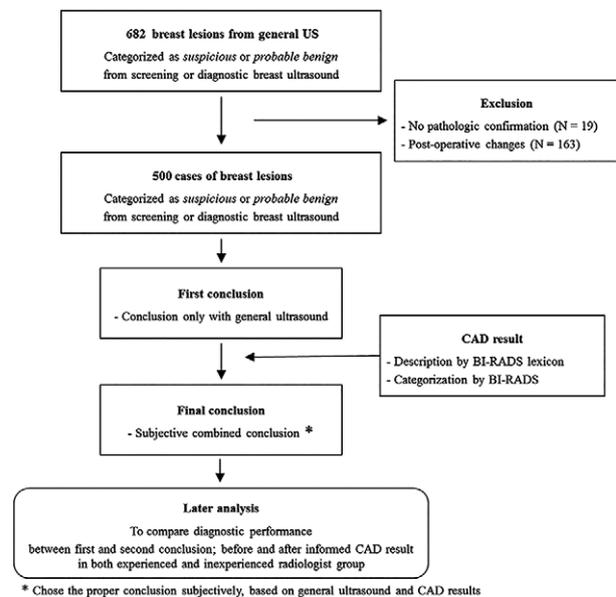


Fig 1. Flow chart of exclusion and decision-making for breast lesions detected from screening or diagnostic US by both experienced and inexperienced radiologist groups.

Statistical analysis

The diagnostic performance for breast malignancy including the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity and specificity of the two radiologist groups and CAD were calculated. The same parameters were then also calculated for the combination conclusions of each group combining the gray-scale US results and the CAD results.

After evaluation of the diagnostic performance of CAD and each radiologist group, we evaluated the agreement of BI-RADS US lexicon and final categorization between CAD and each radiologist group using Kappa statistics. The estimated Kappa used to estimate the extent of agreement was defined by Landis et al [14] as follows: if Kappa is less than 0, “No agreement”, if 0-0.2, “Slight agreement”, if 0.2-0.4, “Fair agreement”, if 0.4-0.6, “Moderate agreement”, if 0.6-0.8, “Substantial agreement”, and if 0.8-1.0, “Almost perfect agreement”.

The statistical analysis was performed using the Statistical Package for the Social Sciences (SPSS) Statistics software 24.0 (IBM Corp, Armonk, New York). A p value less than 0.05 was statistically significant.

Result

The demographic characteristics of the 413 patients with 500 lesions are detailed in table I.

Diagnostic performance of radiologists and CAD

Firstly, comparing the results of general US obtained by experienced radiologists with those of inexperienced radiologists, all diagnostic performance indicators were significantly higher in the experienced group ($p < 0.001$), except for specificity, that did not differ between the two groups ($p = 0.798$).

The diagnostic accuracy for breast malignancy of the experienced radiologist group was high (AUC=0.83, 95% CI [0.80, 0.86]) and no significant difference was

observed with the CAD result (AUC=0.79, 95%CI [0.74, 0.83], $p = 0.101$). However, the diagnostic accuracy of the inexperienced group was 0.65 (95%CI [0.58, 0.71]), lower than that of the CAD result (AUC=0.73, 95%CI [0.67, 0.78], $p = 0.013$).

The AUC value of the combination conclusion (general US and CAD) of the experienced group was not significantly different from the AUC value of the general US only result by the experienced group (Table II). However, in the inexperienced group, the combination conclusion revealed higher AUC value than the general US only result. Not only the AUC value of the combination conclusion in the inexperienced group, but other factors including sensitivity, specificity, PPV and negative predictive value (NPV) were higher. This means that the inexperienced group made a more accurate judgment for breast lesions when making the conclusion based on the CAD results (fig 2).

No significant differences were revealed in other diagnostic performance indicators of the combination conclusion in the experienced group, compared to the results of general US, except for specificity and PPV. In the experienced group, sensitivity was higher than that of CAD, but CAD revealed higher specificity than specificity of general US in the experienced group.

Agreement of BI-RADS categorization and US lexicon between CAD and radiologists

We compared the categorization by CAD and radiologists for the suspicious or probable benign lesions on US. According to the BI-RADS category, lesions with more than C4A – low suspicion for malignancy – were re-categorized as “suspicious lesions” and the others as “not suspicious lesions” by radiologists.

When comparing the categorizations, the agreement between CAD and general US of the experienced group reveals fair agreement with the Kappa value of 0.38 (95%CI [0.29, 0.47], $p < 0.001$) and moderate agreement

Table I. Patients' characteristics

	Total	Benign	Malignancy	p value
Patients	413	354(85.74)	59(14.29)	
Lesions	500	432(86.4)	68(13.6)	
Age (years)	48.76±11.53	48.13±11.64	52.53±10.18	0.026
US for screening	272(65.9)	243(89.34)	29(10.66)	0.006
US for diagnostic	141(34.1)	111(78.72)	30(21.28)	
Symptoms' presence	63(15.3)	36(57.14)	27(42.86)	
Family history	109(26.4)	96(88.07)	13(11.93)	
Postmenopausal status	177(42.9)	148(83.62)	29(16.38)	
Size (cm)	1.19±0.78	1.15±0.77	1.46±0.76	<0.001

The results are expressed as number, mean±standard deviation or mean (%). US – ultrasonography.

Table II. Diagnostic performance of combination conclusion by experienced and inexperienced radiologist groups

	AUC	p value	Sensitivity	p value	Specificity	p value	PPV	p value	NPV	p value
Experienced group										
General ultrasound	0.83 (0.8-0.86)	reference	0.97 (0.90-1.00)	reference	0.7 (0.65-0.74)	reference	0.34 (0.27-0.41)	reference	0.99 (0.98-1.00)	reference
S-detect	0.79 (0.74-0.84)	0.101	0.81 (0.70-0.89)	<0.001	0.77 (0.73-0.81)	0.004	0.36 (0.28-0.44)	0.351	0.96 (0.94-0.98)	0.001
Subjective combined conclusion	0.84 (0.81-0.87)	0.451	0.96 (0.88-0.99)	0.317	0.72 (0.68-0.77)	0.003	0.35 (0.28-0.43)	0.025	0.99 (0.97-1.00)	0.358
Inexperienced group										
General ultrasound	0.65 (0.58-0.71)	reference	0.59 (0.46-0.71)	reference	0.7 (0.66-0.75)	reference	0.24 (0.18-0.31)	reference	0.92 (0.88-0.94)	reference
S-detect	0.73 (0.67-0.78)	0.013	0.76 (0.65-0.86)	0.007	0.69 (0.65-0.74)	0.710	0.28 (0.22-0.35)	0.069	0.95 (0.92-0.97)	0.009
Subjective combined conclusion	0.71 (0.65-0.77)	0.001	0.69 (0.57-0.8)	0.008	0.73 (0.69-0.77)	0.033	0.29 (0.22-0.37)	<0.001	0.94 (0.91-0.96)	0.003

Table III. Kappa values between CAD and radiologist group for determination (final category) and descriptions (lexicon)

	Kappa statistics	
	Experienced group	Inexperienced group
Determination		
Final Category	0.38 (0.29-0.47)	0.4 (0.32-0.49)
Combined Conclusion†	0.44 (0.53-0.36)	0.57 (0.50-0.65)
Descriptions (Lexicon)		
Shape	0.51 (0.43-0.58)	0.3 (0.22-0.38)
Orientation	0.58 (0.45-0.71)	0.29 (0.16-0.42)
Margin	0.37 (0.3-0.44)	0.25 (0.18-0.32)
Echo pattern	0.43 (0.36-0.51)	0.12 (0.04-0.2)
Posterior features	0.37 (0.27-0.47)	0.22 (0.11-0.33)

† “Combined Conclusion” means subjective combined conclusion that appropriate decision was chosen subjectively, based on final US category with CAD conclusion.

with Kappa value 0.4 (95%CI [0.32, 0.49]) between CAD result and general US of inexperienced group was noted. However, after applying the CAD result to the general US of each group, increases of Kappa value were revealed in both groups, and were more prominent in the inexperienced group (from 0.38 to 0.44 in experienced group, from 0.4 to 0.57 in inexperienced group).

The Kappa values for all US lexicons between CAD and each radiologist group were also evaluated. The Kappa value between CAD and experienced group is higher than that between CAD and inexperienced group in all US lexicons (Table III, fig 3).

Discussions

Breast cancer is the most frequently diagnosed cancer in women worldwide. However, due to its favorable prognosis it ranks as the fifth cause of deaths in Korea

[15]. For this reason, early detection and diagnosis of breast cancer is critical for survival [16-18]. Mammography is the only screening method proven to improve early diagnosis and decrease morbidity and mortality of breast cancer. However, in dense breast which is classified as high-density pattern – C or D – in the BI-RADS lexicon, the sensitivity of mammography to detect cancer is poor. Therefore, breast US was adopted as a complementary method as a screening tool for dense breasts [19].

Breast US is potentially effective for screening, but due to its operator-dependent nature, many additional US techniques have recently been suggested and applied in practice to improve diagnostic accuracy. One of the newly suggested techniques is the CAD system, that can provide efficient interpretation or second opinion for a lesion detected on breast US [10,12].

S-detect™ (Samsung Medison, Co, Ltd, Seoul, Korea) is a recently developed US CAD system that pro-

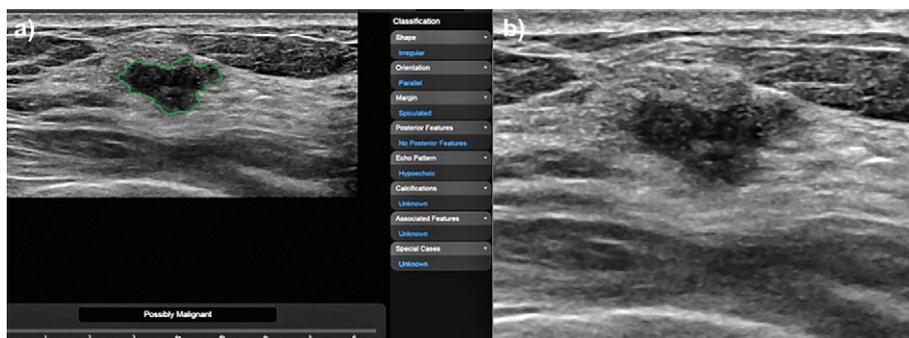


Fig 2. A hypoechoic mass was detected at 10 o'clock direction of left breast on general US (A) for screening of 43-year-old woman. For this lesion, CAD analysis was applied (B) at the same time of general US. An experienced radiologist categorized this lesion as BI-RADS category 4C (high suspicion for malignancy), and the inexperienced radiologist categorized it as BI-RADS category 3 (probably benign lesion). From the CAD result, both groups finally categorized the lesion as a suspicious lesion of BI-RADS category 4 or higher. This lesion was confirmed as an invasive ductal carcinoma after excision.

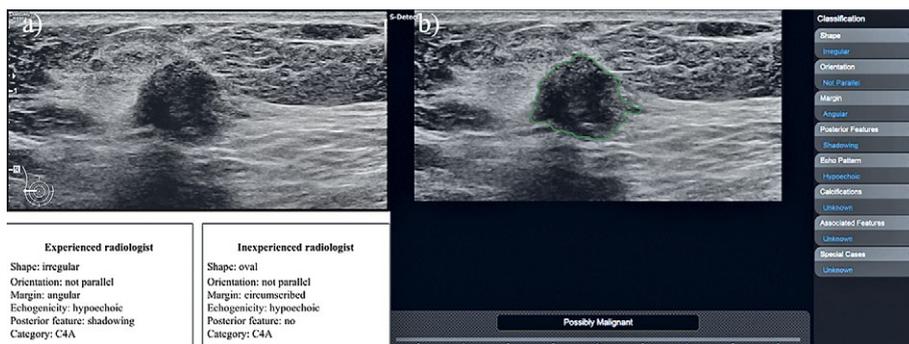


Fig 3. A hypoechoic mass was detected at 9 o'clock direction of the right breast on general US (A) for a diagnostic sonography of 47-year-old woman with palpable mass. Both experienced and inexperienced radiologists described and categorized the mass according to the BI-RADS lexicon. The CAD result (B) was compared with both radiologists' results respectively. While no differences were found in the description between CAD and the experienced radiologist, some differences were found between the CAD result and the inexperienced radiologist's result. Categorization of the mass by radiologists and CAD was the same. After excision, this lesion was confirmed as an invasive ductal carcinoma.

vides additional morphologic analysis of breast masses detected on breast US according to the BI-RADS lexicon and assists in the final assessment for the detected breast masses [20]. According to previous studies using S-detect™, it can be beneficial to improve specificity for the diagnosis of malignant breast mass in breast US regardless of the experience of radiologists [13,21]. In particular, the use of S-detect™ showed increased specificity of the less experienced radiologists compared to the experienced breast radiologists in previous studies done by our group [13] and by Di Segni et al [22].

Similar to previous studies, application of the CAD result to the result of general US of each radiologist group showed significant increase of specificity of both groups in our study. In the inexperienced group, not only specificity, but other indicators of diagnostic performance significantly increased. However, unlike previous studies on S-detect™, this study has a significantly greater number of lesions analyzed (n=500), and each group was composed of similarly experienced radiologists.

The Kappa statistics showed that the agreement between CAD and each radiologist group for categorization of suspicious lesions was increased after each radiologist group applied CAD result to their final categorization. Moreover, increase of Kappa value was higher in the inexperienced group than in the experienced group. This implies that the inexperienced group is more likely to accept the CAD result than the experienced group when making a final decision for the suspicious lesions on general US. Furthermore, not only the increase of Kappa value but also the absolute Kappa value of the inexperienced group was higher than that of the experienced group after applying the CAD result.

Given these results, the US CAD system can serve as a supplementary tool for inexperienced groups when making a final decision about the status of a lesion detected on US. This also means that it can compensate for the disadvantages of operator-dependency of US. Conversely, for the experienced group, CAD serves as a means of providing a second opinion similar to mammography CAD.

Agreement between CAD and each radiologist group for US lexicon was also reviewed in our study. In every lexicon, higher Kappa values were noted between CAD and the experienced group, than CAD and the inexperienced group. Although the result of CAD for US lexicon is not a gold standard, it may be beneficial and educational for inexperienced radiologists to describe and understand the features of a lesion with a CAD result.

Our study has a few limitations. First, due to its retrospective nature, analysis of the features of lesions was limited, because the radiologists did not analyze real-

time images, but captured images. Second, calcifications and non-mass lesions were not included in the analysis due to a lack of detection in general US. This situation can imply different results when applying CAD in clinical practice, because some cases of breast cancers appear as calcifications or non-mass lesions only.

We comparatively investigated the diagnostic performances of an experienced radiologist group and an inexperienced radiologist group to distinguish malignant and benign lesions when applying the CAD system in US. Overall, while no significant changes were observed in the experienced group when using the CAD system, a significant increase of diagnostic performance was revealed in the inexperienced group. Also, we found that the inexperienced group is more likely to accept and apply the CAD result when making final decision for the categorization of the lesions through Kappa values between CAD and each radiologist group.

In **conclusion**, the CAD system can be more beneficial and educational for less experienced radiologists than for experienced radiologists, not only when describing lesions, but also when determining if the lesion is malignant.

Acknowledgement: The statistical consultation was supported by Department of Biostatistics of the Catholic Research Coordinating Center, Catholic Medical Center, South Korea.

Conflict of interest: none

References

1. Dromain C, Boyer B, Ferre R, Canale S, Delaloue S, Balleysguier C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur J Radiol* 2013;82:417-423.
2. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001;220:781-786.
3. Wood C. Computer Aided Detection (CAD) for breast MRI. *Technol Cancer Res Treat* 2005;4:49-53.
4. Meeuwis C, van de Ven SM, Stapper G, et al. Computer-aided detection (CAD) for breast MRI: evaluation of efficacy at 3.0 T. *Eur Radiol* 2010;20:522-528.
5. Meinel LA, Stolpen AH, Berbaum KS, Fajardo LL, Reinhardt JM. Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. *J Magn Reson Imaging* 2007;25:89-95.
6. 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;299:2151-2163.

7. Chen CM, Chou YH, Han KC, et al. Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology* 2003;226:504-514.
8. Kim KG, Cho SW, Min SJ, Kim JH, Min BG, Bae KT. Computerized scheme for assessing ultrasonographic features of breast masses. *Acad Radiol* 2005;12:58-66.
9. Drukker K, Giger ML, Vyborny CJ, Mendelson EB. Computerized detection and classification of cancer on breast ultrasound. *Acad Radiol* 2004;11:526-535.
10. Horsch K, Giger ML, Vyborny CJ, Venta LA. Performance of computer-aided diagnosis in the interpretation of lesions on breast sonography. *Acad Radiol* 2004;11:272-280.
11. Sehgal CM, Cary TW, Kangas SA, et al. Computer-based margin analysis of breast sonography for differentiating malignant and benign masses. *J Ultrasound Med* 2004;23:1201-1209.
12. Horsch K, Giger ML, Vyborny CJ, Lan L, Mendelson EB, Hendrick RE. Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set. *Radiology* 2006;240:357-368.
13. 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;37:217-225.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
15. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359-E386.
16. Jacobs MA, Wolff AC, Macura KJ, et al. Multiparametric and Multimodality Functional Radiological Imaging for Breast Cancer Diagnosis and Early Treatment Response Assessment. *J Natl Cancer Inst Monogr* 2015;2015:40-46.
17. Massat NJ, Dibden A, Parmar D, Cuzick J, Sasieni PD, Duffy SW. Impact of Screening on Breast Cancer Mortality: The UK Program 20 Years On. *Cancer Epidemiol Biomarkers Prev* 2016;25:455-462.
18. Weedon-Fekjaer H, Romundstad PR, Vatten LJ. Modern mammography screening and breast cancer mortality: population study. *BMJ* 2014;348:g3701.
19. Bahl M, Baker JA, Bhargavan-Chatfield M, Brandt EK, Ghate SV. Impact of Breast Density Notification Legislation on Radiologists' Practices of Reporting Breast Density: A Multi-State Study. *Radiology* 2016;280:701-706.
20. 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.
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;37:209-216.
22. 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;21:105-118.