



Reliability Studies of Different Readers Using Computer Aided Diagnosis System in Ultrasound for Determining Identical Lesions in a Breast Phantom

Nari Choi¹, Bong Joo Kang², Na Young Jung^{1,*}, Sung Hun Kim², Jung Min Lee² and Ga Eun Park²

¹Department of Radiology, Bucheon St. Mary's Hospital, College of Medicine, Catholic University of Korea, Bucheon, Republic of Korea

²Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, Catholic University of Korea, Seoul, Republic of Korea

*Corresponding author: Department of Radiology, Bucheon St. Mary's Hospital, College of Medicine, Catholic University of Korea, 327, Sosa-ro, Wonmi-gu, Bucheon-si, 14647, Gyeonggi-do, Republic of Korea. Tel: +82-323407369, Email: healmind@catholic.ac.kr

Received 2018 November 29; Revised 2019 June 19; Accepted 2019 June 23.

Abstract

Background: The computer-aided diagnosis (CAD) has been applied in multiple studies conducted recently in order to perform ultrasound (US) of the breast. There are several studies that have indicated that CAD is useful for improving the diagnostic performance in less experienced radiologists. However, there is no study on several readers analyzing the same lesions using a breast phantom and rare reliability studies.

Objectives: To investigate the reliability of different readers using computer aided diagnosis (CAD) system in US for determining identical lesions in a breast phantom.

Patients and Methods: From March 2016 to February 2017, six readers (three senior and three junior residents in the department of radiology) evaluated and analyzed breast phantom including 14 lesions. At the first line study, three senior residents (3rd grade with more than one month of training for breast ultrasound [US]) and three junior residents (1st grade without breast US training) evaluated and analyzed the US and applied CAD for lesions in breast phantom, and they were able to make final decisions by subjective combination. A month later, they conducted the second line study as they did the first line study. We analyzed the inter- and intra-reader reliability and accuracy of US, CAD, and combinations (subjective, conjunctive, and disjunctive).

Results: In the total of first and second line studies of six readers, the kappa value of US (0.609) was significantly higher than CAD (0.411). In the subjective combined conclusion, the kappa value was improved by the junior group. In the whole inter- and intra-reader analysis, the kappa values of final assessment of the senior and junior group were more variable on CAD than on US, especially in the junior group, and this result was statistically significant. The area under the curve (AUC) of US, CAD, subjective, conjunctive, and disjunctive combination in seniors were all better than those of juniors. In all groups, the AUCs, sensitivities, and specificities were improved on conjunctive combined US with CAD.

Conclusion: The combination US with CAD improved the reliability and diagnostic performance. The CAD results of the junior group were variable and inconsistent. Therefore, minimum training and experience for breast US is indispensable for the better use of breast CAD, and, combination US with CAD is useful for all readers.

Keywords: Ultrasonography, Breast Neoplasm, CAD

1. Background

Breast cancer is a leading cause of mortality in women and its incidence continues to increase (1-3). It is known that early detection of breast cancer is important to reduce the mortality rate. Ultrasound (US) is an important diagnostic tool that is widely used for breast cancer screening. It has the advantages of being non-invasive and non-radioactive (2). In addition, it can also be used to detect lesions in dense breast (3, 4).

The main pitfall of breast US is inter- and intra-reader

variability in breast cancer detection and diagnosis (5).

Many recent researches have used computer-aided diagnosis (CAD) with breast US in order to improve the diagnostic performance (4, 6-8). In breast US, CAD plays a role in interpreting lesions that are found by the examiner, rather than detecting lesions.

S-detect™ is a recently developed breast cancer CAD system that is useful for differentiation between malignant and benign breast masses through the morphological analysis based on the American College of Radiology breast imaging reporting and data system (ACR BI-RADS)

(5, 9). The S-detect™ adopted a deep learning algorithm in the processes of lesion segmentation. The S-detect™ for breast module utilizes large data sets collected from numerous breast examination cases and provides the characteristics of displayed lesion as well as a suggestion as to whether a selected lesion is benign or malignant.

Several studies have been conducted in order to improve diagnostic performance in breast US using Breast US CAD system (5, 6, 9), and there are also studies indicating that S-detect™ is useful for improving diagnostic performance in less experienced radiologists (9). However, there have been no studies on several readers analyzing the same lesion, because almost all studies have been conducted on actual patients. In our study, we used breast phantom which enables multi-reader analysis for the same lesion.

2. Objectives

This study aims to identify the merits of the CAD system in the breast US reading by breast senior residents and junior residents, and the aim of this research is to evaluate the reliability of the CAD system for US using breast phantom with multi-reader analysis for identical lesions.

3. Patients and Methods

3.1. Breast Phantom

This study was approved by the Institutional Review Board. Informed consent using phantom was waived.

We used breast US examination phantom “Breast FAN” (Kyoto Kagaku Co. Ltd., Kyoto, Japan). Commercial ‘Breast Ultrasound Examination Phantom’ has lesions that represent a variety of benign and malignant lesions (Figure 1A). It has the same anatomy as the normal breast US and the lesions are embedded in it. However, because of the fact that the original breast phantom only contains a few lesions, we customized our own breast phantom (Customized Breast Ultrasound Examination Phantom) for the study which has 14 lesions: five suspicious lesions, six benign lesions, and three axillary lymph nodes. We sent breast US pictures of typical benign and malignant lesions with specific sizes and depths to the manufacturer and asked them to reproduce these exact lesions (Figure 1B).

3.2. Breast US and CAD System

We used an US machine (Samsung Ultrasound RS80A, Samsung Medison Co. Ltd., Seoul, Korea), which makes it possible in conjunction with breast grayscale US and the new technology of CAD (S-detect™).

When the reader identified the center of the breast lesion and touched the screen, the region of interest (ROI)

was automatically drawn along the lesion boundary. Several drawn borders along the lesion boundary appeared on the screen, and the reader selected the most appropriate border of the lesion. In this system, the US features of the lesion based on the ACR BI-RADS lexicon for US (Shape, orientation, margin, posterior features, echo pattern, calcifications, and associated features) as well as the final assessment classifications were simultaneously analyzed and displayed. In this CAD system (S-detect™), the final assessment was classified as “possibly benign” or “possibly malignant” (Figure 2A and B).

3.3. Analysis by Radiologists

From March 2016 to February 2017, six readers (three senior and three junior residents in the radiology department) evaluated and analyzed breast phantom including 14 lesions.

At the first line of study, three senior residents (third grade with more than one month of training for breast US) and three junior residents (first grade without breast US training) evaluated and analyzed the US and applied CAD for lesions in breast phantom and they were able to make final decisions by subjective combination. The junior group did not have experience with breast US for phantom or patient, but for this study, two hours of lectures were given to the junior group regarding breast US and US BI-RADS.

Each reader evaluated all 14 lesions of breast phantom. They first detected the breast phantom lesion by US. US characteristics of the lesion morphology were defined according to ACR US BI-RADS lexicon and final assessment classification. For statistical analysis, we divided this final assessment into a dichotomized form, and the cut-off was C4 (C2, 3 [benign] vs. C4, 5 [malignancy]). They then simultaneously applied the CAD for the same lesion in breast phantom. When the reader identified the lesion and touched the center of the lesion in the monitor, a ROI was drawn along the border of the mass either automatically by the CAD program. Several drawn borders were presented on the screen, and the reader selected the most appropriate one. BI-RADS lexicon and final assessment classifications were automatically analyzed and displayed by the CAD system. They then made a diagnostic decision subjectively based on US with CAD.

A month later, they did the second line study as they did the first line study. Later analysis includes the conjunctive and disjunctive combination (Figure 2C).

3.4. Statistical Analyses

We analyzed the inter- and intra-reader reliability and diagnostic performance of breast US, CAD, and combinations between junior and senior residents.

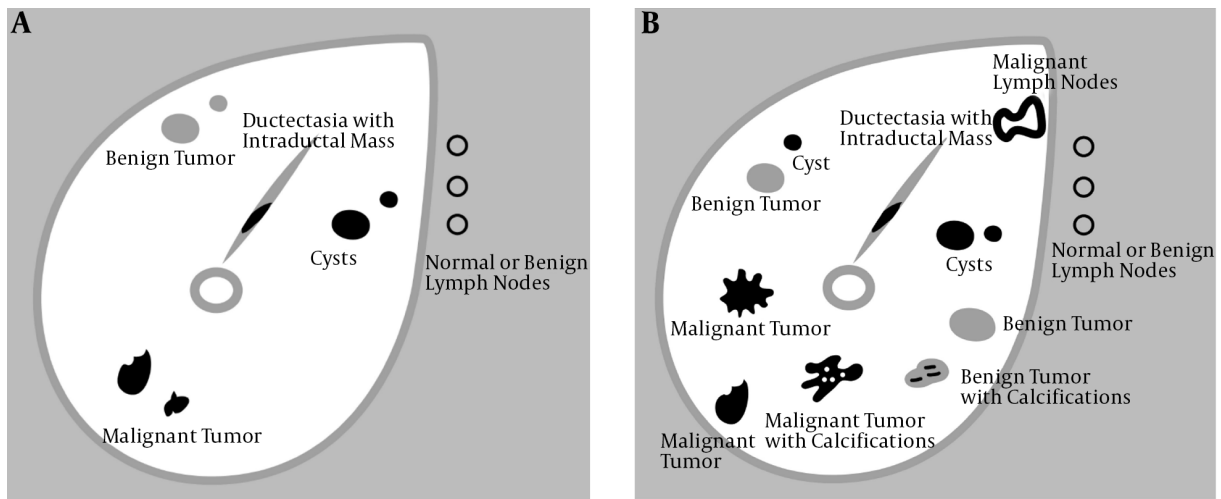


Figure 1. Diagram of breast ultrasound examination phantom. A, Commercial breast ultrasound examination phantom. B, Customized breast ultrasound examination phantom including five suspicious lesions (two malignant tumors, a malignant tumor with calcifications, a ductectasia with intraductal mass, and malignant lymph node), six benign lesions (three cysts, two benign tumors, and a benign tumor with calcifications), and three normal or benign lymph nodes.

In this study, the inter-reader reliability was for comparison of the final assessment in CAD of senior or junior to gold standard. The intra-reader reliability was for comparison between the primary and secondary assessments. The reliability was analyzed using kappa statistics (Cohen's kappa). The kappa value indicates the degree of accidental exceedance. A real number from 0 to 1; a high value indicates a high level of agreement. The level of consent for Cohen's kappa is generally defined as: A value of kappa value = 1.0 corresponds to complete agreement, ≤ 0.20 as poor, 0.21 - 0.40 as fair, 0.41 - 0.60 as moderate, 0.61 - 0.80 as good, and > 0.80 as excellent agreement.

We also evaluated the diagnostic performance of US, CAD, and US with CAD between junior and senior residents. The gold standard of lesions in breast phantom was defined by consensus of two dedicated breast radiologists, because phantom lesions cannot be pathologically confirmed. Since there may be a discordance of judgment on the lesion of the breast phantom we actually designed, two dedicated breast radiologists reaffirmed to reduce the discordance.

A receiver operating characteristic (ROC) curve for each reader was obtained for overall diagnostic performance evaluation. We analyzed the area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

We analyzed the results of combination of US with CAD (subjective, conjunctive, and disjunctive combinations). The proper combined decision was chosen subjectively based on the US with CAD results, which was defined as the subjective combination. Readers subjectively chose based

on the US (totally a task related to their personal knowledge and experience) with CAD results. In a conjunctive combination, "not suspicious" finding on both US (category 3) and CAD (possibly benign) was defined as negative, while "suspicious" finding on either the US (category 4 or above) or CAD (possibly malignant) was defined as positive. In a disjunctive combination, "not suspicious" finding on either US or CAD was defined as negative, while "suspicious" finding on both US and CAD was defined as positive.

Comparison of AUC was performed by Dejong method. For sensitivity, specificity, PPV, and NPV comparison, Chi-square exact test was used.

Statistical analyses were performed by SAS version 9.3 (Cary, NC, USA) and P value below 0.05 was considered statistically significant.

4. Results

4.1. Inter- and Intra-Reader Reliability Test

The results of inter-reader and intra-reader reliability test for the six readers are represented in Table 1. In Table 1, the meaning of P value in inter-reader is the result of comparison of senior or junior to gold standard, and the meaning of P value in intra-reader is the result of comparison between the primary and secondary assessments. In the total of first and second line studies of six readers, the kappa value of US ($\kappa = 0.609$) was significantly higher than CAD ($\kappa = 0.441$). In the subjective combined conclusion, the kappa value was improved on junior group ($\kappa = 0.557$ to 0.625). The kappa values of final decision on senior and junior group were more variable on CAD than US. Especially,

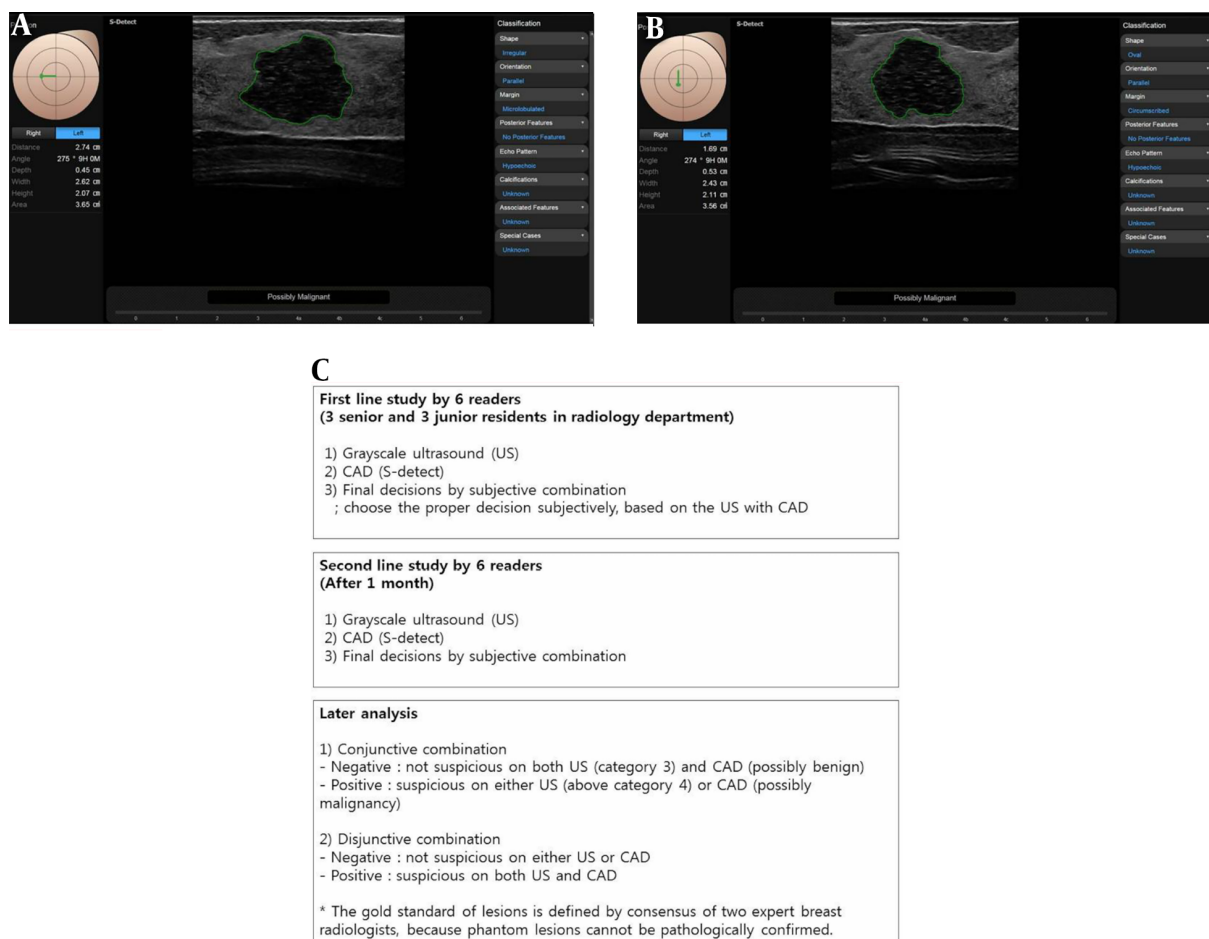


Figure 2. Computer aided diagnosis (CAD) system and flow chart of analysis process of this study. A and B, These figures depict how a breast lesion was classified automatically using the S-detect™ program and how a final assessment was produced. Several drawn borders were presented on the screen, and the reader selected the most appropriate border of the lesion. C, This flow chart shows the analysis process of this study.

in the junior group, kappa value of the first line study of the CAD was 0.694, and that of second line study was 0.278, indicating greater inconsistency of CAD than US, and this result was statistically significant.

The inter-reader and intra-reader reliability for the six readers of each baseline BI-RADS lexicon are represented in Table 2. When analyzing each lexicon, the kappa value of shape, orientation, and margin on US ($\kappa = 0.376, 0.617, \text{ and } 0.769$, respectively) were significantly higher than those on CAD ($\kappa = 0.238, 0.461, 0.562$, respectively ($P \leq 0.001$). The kappa value of echogenicity on CAD ($\kappa = 0.709$) was higher only than that of US ($\kappa = 0.632$).

4.2. Diagnostic Performance

Table 3 shows diagnostic performance of US, CAD, and each combination when the gold standard was specified as

a reference. The gold standard of lesions in breast phantom is defined by consensus of two dedicated breast radiologists, because phantom lesions cannot be pathologically confirmed. P value was provided between the senior and junior groups.

The AUC of gray scale US, CAD, subjective, conjunctive, and disjunctive combination in seniors (0.779, 0.808, 0.769, 0.917 and 0.778, respectively) were all better than those of juniors (0.756, 0.766, 0.769, 0.882, and 0.672, respectively).

In both junior and senior groups, the specificities were higher on CAD (0.997 and 0.976) than those on US (0.882 and 0.792). In the junior group, the AUC and sensitivity were slightly improved on subjective combined US with CAD. The sample size was too small to show the significant difference of diagnostic performances between senior and

Table 1. Reliability Test of Final Assessment Based on Modalities and Readers' Experiences^a

	Inter-reader reliability									
	Total (1st + 2nd line)				1st line study			2nd line study		
	Senior + Junior	Senior	Junior	P value	Senior	Junior	P value	Senior	Junior	P value
Ultrasound	0.609	0.652	0.557	0.279	0.633	0.505	0.308	0.667	0.591	0.537
CAD	0.441	0.425	0.462	0.667	0.318	0.694	0.003	0.518	0.278	0.042
Combination^b	0.603	0.594	0.625	0.718	0.515	0.694	0.154	0.667	0.567	0.397
	Intra-reader reliability									
	Total (Senior + Junior)			Senior			Junior			
	1st line	2nd line	P value	1st line	2nd line	P value	1st line	2nd line	P value	
Ultrasound	0.575	0.639	0.279	0.633	0.667	0.782	0.505	0.591	0.493	
CAD	0.430	0.449	0.738	0.318	0.518	0.104	0.694	0.278	0.001	
Combination	0.591	0.608	0.764	0.515	0.667	0.217	0.694	0.567	0.292	

Abbreviation: CAD, computer aided diagnosis

^aP value meaning: (1) inter-reader: comparison senior or junior to gold standard. (2) Intra-reader: comparison between the primary and secondary assessments.

^bCombination: subjective combination result.

junior groups.

In total, the AUC (0.882 and 0.770, P value = 0.009) and specificity (0.987 and 0.838, P value = 0.001) were significantly improved on conjunctive combined US with CAD than US only.

5. Discussion

The CAD system is used to assist radiologists in breast mass discrimination. There have been many studies on the CAD system investigating the benefit of its output for radiologist diagnosis. Some studies have shown that CAD enhances the diagnostic performance of radiologists (4, 5, 9-13).

S-detect™ is a recently developed breast cancer CAD system using deep-learning algorithm with big data which providing assistance in morphological analysis by the ACR BI-RADS (4, 5, 9-12). Several studies have reported that S-detect™ could enhance the diagnostic performance of radiologists (5, 6, 9-12, 14). In addition, several studies have published that S-detect™ is a useful diagnostic tool that could be clinically used to enhance the specificity, PPV, and accuracy of breast US, regardless of the degree or radiologist's experience (9, 12). However, it is known that combining CAD with breast US is more useful than CAD alone (9, 10). Our study aimed to determine the merits of the CAD system in breast US reading using breast phantom. In one previous study, five readers including residents retrospectively reviewed the CAD images and gave their assessments; inter-rater agreement was measured with Cohen's kappa value (11). The conclusion of that study was that S-detect™ is a feasible tool for characterization of breast le-

sions; it has a potential as a teaching tool for the less experienced operators. However, in that study, the residents only reviewed and assessed images taken by a radiologist with 32 years of experience in breast imaging (11). In our study, we used breast phantom, not real patients. This is the greatest strength of our research. Junior residents who are trained for 2 hours can not perform ultrasound on real patients. We would like to demonstrate that breast phantom enables breast ultrasound education and training for starters.

First, we studied the reliability of CAD system for breast US. We used breast phantom, which enables multi-reader analysis for the same lesion. There have been several papers discussing breast phantom as a tool for breast US training (15-18), but this is the first study to directly apply breast phantom for reliability studies for breast CAD system. In our conclusion, there was better agreement of lexicons and final assessment in US than in CAD. The kappa values of the final decision on senior and junior groups were more variable on CAD than US, especially, in the junior group, there was greater inconsistency of CAD than US. Similar to the breast US, the inter- and intra-readers variability exists in CAD. In one previous study, moderate agreement ($\kappa = 0.58$) was seen in the final assessment between the CAD and dedicated breast radiologist (5). The kappa value ($\kappa = 0.44$) between residents' CAD result and dedicated breast radiologists in our study was lower than that ($\kappa = 0.58$) of the previous study. In order for CAD to be used properly as a dedicated breast radiologist, radiologists must get a proper US shot and then apply CAD. In our study, the CAD results for the junior group (beginner or starter) varied and were inconsistent. The kappa

Table 2. Reliability Test of Baseline Lexicons Based on Modalities and Readers' Experiences

	Inter-reader reliability									
	Total (1st + 2nd line)				1st line study			2nd line study		
	Senior + junior	Senior	Junior	P value	Senior	Junior	P value	Senior	Junior	P value
Ultrasound										
Shape	0.376	0.596	0.147	< 0.001	0.764	0.185	< 0.001	0.468	(-)0.065	< 0.001
Orientation	0.617	0.659	0.533	0.151	0.857	0.619	0.058	0.303	0.426	0.318
Margin	0.769	0.824	0.607	0.013	0.879	0.570	0.014	0.771	0.628	0.245
Echo pattern	0.632	0.303	0.878	< 0.001	0.144	1.000	< 0.001	0.444	0.757	0.011
Posterior shadowing	0.158	0.138	(-)0.015	0.081	0.147	(-)0.031	0.156	(-)0.029	0.000	0.814
CAD										
Shape	0.238	0.346	0.419	0.395	0.139	0.593	< 0.001	0.556	0.164	0.001
Orientation	0.461	0.430	0.437	0.935	0.455	0.389	0.599	0.303	0.478	0.138
Margin	0.562	0.477	0.600	0.152	0.393	0.492	0.430	0.550	0.687	0.246
Echo pattern	0.709	0.592	0.776	0.032	0.389	0.636	0.049	0.777	0.894	0.321
Posterior shadowing	0.095	0.087	(-)0.029	0.177	0.147	(-)0.065	0.091	(-)0.059	0.000	0.617
Intra-reader reliability										
	Total (senior + junior)			Senior			Junior			
	1st line	2nd line	P value	1st line	2nd line	P value	1st line	2nd line	P value	
Ultrasound										
Shape	0.488	0.242	< 0.001	0.764	0.468	0.016	0.185	(-)0.065	0.047	
Orientation	0.732	0.441	< 0.001	0.857	0.303	< 0.001	0.619	0.426	0.124	
Margin	0.712	0.825	0.056	0.879	0.771	0.380	0.570	0.628	0.644	
Echo pattern	0.593	0.666	0.217	0.144	0.444	0.015	1.000	0.757	0.053	
Posterior shadowing	0.171	(-)0.017	0.002	0.147	(-)0.029	0.153	(-)0.031	0.000	0.805	
CAD										
Shape	0.236	0.230	0.916	0.139	0.556	0.001	0.593	0.164	< 0.001	
Orientation	0.425	0.480	0.332	0.455	0.303	0.217	0.389	0.478	0.461	
Margin	0.506	0.608	0.072	0.393	0.550	0.202	0.492	0.687	0.106	
Echo pattern	0.583	0.811	< 0.001	0.389	0.777	0.002	0.636	0.894	0.032	
Posterior shadowing	0.111	(-)0.029	0.014	0.147	(-)0.059	0.094	(-)0.065	0.000	0.590	

Abbreviation: CAD, computer aided diagnosis

value of CAD was lower than that of the US. The statistical value was limited because of the small number of lesions in the breast phantom. However, the variability and inconsistency of the junior group were difficult to ignore. Therefore, we suggest that minimum training and experience for breast US is indispensable for better use of breast CAD.

In the subjective combined conclusion, the kappa value was improved in the junior group. When analyzing each lexicon, the kappa value of shape, orientation, and margin on US were significantly higher than those on CAD. The kappa value of echogenicity on CAD was higher only

than that of US. So, we found that combining with breast US could improve the reliability in this study.

We also evaluated the diagnostic performance of the CAD system with breast US, by the junior and senior readers. The AUC was higher in CAD than US, while conjunctive combination result was the best. In addition, the diagnostic performance of CAD in the senior group was better than that of the junior group similar to US. We also found that combining CAD system with breast US could improve the diagnostic performance in this study. In one previous study, AUC improved for both the experienced and inexpe-

Table 3. Diagnostic Performance of US, CAD, and Combinations

	AUC	Sensitivity	Specificity	PPV	NPV
Ultrasound					
Total	0.770	0.702	0.838	0.714	0.830
Junior	0.756	0.630	0.882	0.739	0.818
Senior	0.779	0.767	0.792	0.697	0.844
P value	0.780	0.385	0.279	0.773	0.794
CAD					
Total	0.787	0.597	0.977	0.956	0.743
Junior	0.766	0.556	0.977	0.952	0.729
Senior	0.808	0.639	0.976	0.958	0.759
P value	0.581	0.631	> 0.999	> 0.999	0.830
Subjective combination					
Total	0.770	0.712	0.828	0.712	0.828
Junior	0.769	0.655	0.882	0.760	0.818
Senior	0.769	0.767	0.771	0.677	0.841
P value	> 0.999	0.399	0.185	0.568	0.796
Conjunctive combination					
Total	0.882	0.778	0.987	0.984	0.804
Junior	0.846	0.718	0.974	0.966	0.776
Senior	0.917	0.833	1.000	1.000	0.837
P value	0.203	0.286	> 0.999	0.453	0.600
Disjunctive combination					
Total	0.726	0.553	0.899	0.703	0.824
Junior	0.672	0.435	0.909	0.667	0.794
Senior	0.778	0.667	0.889	0.727	0.857
P value	0.256	0.147	0.761	0.728	0.472

Abbreviations: AUC, area under the curve; CAD, computer aided diagnosis; NPV, negative predictive value; PPV, positive predictive value; US, ultrasound

rienced readers (0.84 to 0.86 and 0.73 to 0.80) after the addition of CAD (9). In our study, AUC improved for both senior and junior resident groups (0.779 to 0.917 and 0.756 to 0.846) after conjunctive combination. In another study, CAD was a useful additional diagnostic tool for breast US in all radiologists, with benefits differing depending on the radiologist's level of experience. Compared with the experienced radiologists, the less experienced radiologists had significantly improved NPV (0.867 to 0.94 and 0.533 to 0.762) and AUC (0.823 to 0.839 and 0.623 to 0.759) with CAD assistance. In contrast, experienced radiologists had significantly improved specificity (0.525 to 0.542 and 0.661 to 0.661) and PPV (0.556 to 0.585 and 0.649 to 0.649) with CAD assistance. Interobserver variability of US features and final assessment by categories were significantly improved and moderate agreement was seen in the final assessment

after CAD combination regardless of the radiologist's experience (10). In our study, combination of US with CAD improved the reliability and diagnostic performance, especially in the junior group.

There are limitations in our study. First, the data used in this study were derived from too few lesions ($n=14$). The sample volume is very low which could decrease the accuracy of the study. This is probably the major cause of why data from our study did not yield statistically significant results. Specifically, the sample size is too small to show the significant difference of diagnostic performances between senior and junior groups. Since our study was based on breast phantom, the small number of lesions was inevitable. However, using breast phantom has several advantages. It can result in more reproducible results, it is objective, and studies can be repeated many times. In the

future, various studies, especially the reliability test could be applied using various phantoms. Second, in our study, the CAD system did not analyze calcification because the number of lesions including calcification in our phantom was insufficient for reliability analysis. For the same reason, associated features such as duct change were not analyzed in detail. Third, since our study was a study using breast phantom, pathologic confirmation was not possible and the gold standard was reaffirmed by dedicated breast radiologists. Therefore, there is a limit in deriving the diagnostic performance from this. Finally, when the reader selected the representative image, which could be differed in CAD depends on the readers. When the reader identified the lesion and touched the center of the lesion in the monitor, a ROI was drawn along the border of the mass either automatically by the CAD program. Several drawn borders were presented on the screen, and the reader selected the most appropriate one. BI-RADS lexicon and final assessment classifications were automatically analyzed and displayed by the CAD system. However, the readers selected the representative US image, touched the center of the lesion, and selected the most appropriate CAD image. Any change in any of these steps could make a different effect on CAD result based on the readers.

In conclusion, the combination of US with CAD improved the reliability and diagnostic performance, especially in the junior group. As mentioned earlier, several studies have shown that CAD is useful for inexperienced radiologists. However, in our study, the junior group actually meant beginners, and the CAD results of the junior group (beginner or starter) were variable and inconsistent. Therefore, minimum training and experience for breast US is indispensable for the better use of breast CAD, and combination of US with CAD is useful for all readers.

Footnotes

Authors' Contributions: Study concept and design: Bong Joo Kang and Na Young Jung. Acquisition of data: Nari Choi, Bong Joo Kang and Jung Min Lee. Analysis and interpretation of data: Nari Choi, Bong Joo Kang, Sung Hun Kim and Ga Eun Park. Drafting of the manuscript: Nari Choi, Bong Joo Kang and Na Young Jung. Critical revision of the manuscript for important intellectual content: Nari Choi, Bong Joo Kang, Na Young Jung, Jung Min Lee, Sung Hun Kim and Ga Eun Park. Statistical analysis: Bong Joo Kang and Ga Eun Park. Administrative, technical, and material support: Bong Joo Kang, Jung Min Lee and Sung Hun Kim. Study supervision: Bong Joo Kang and Na Young Jung

Conflict of Interests: The authors declare that there are no conflict of interests.

Ethical Approval: This manuscript involves no conflict of interest or ethical adherence issues. This study was performed at Seoul St. Mary's Hospital, College of Medicine, Catholic University of Korea. Ethical Approval code: KC15DNSI0027 (KCMC IRB, <http://eirb.cmcnu.or.kr/Main.aspx>).

Financial Disclosure: Dr. Bong Joo Kang reported receiving research grant from the Korean Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI15C0833). Dr. Bong Joo Kang reported that the statistical consultation was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), which was funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI14C1062).

Funding/Support: This research was supported by a grant from the Korean Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI15C0833). The statistical consultation was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), which was funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI14C1062).

References

1. Wang Y, Wang H, Guo Y, Ning C, Liu B, Cheng HD, et al. Novel computer-aided diagnosis algorithms on ultrasound image: Effects on solid breast masses discrimination. *J Digit Imaging*. 2010;23(5):581-91. doi: [10.1007/s10278-009-9245-1](https://doi.org/10.1007/s10278-009-9245-1). [PubMed: [19902300](https://pubmed.ncbi.nlm.nih.gov/19902300/)]. [PubMed Central: [PMC3046684](https://pubmed.ncbi.nlm.nih.gov/PMC3046684/)].
2. Wang X, Guo Y, Wang Y. Automatic detection of regions of interest in breast ultrasound images based on local phase information. *Biomed Mater Eng*. 2015;26 Suppl 1:S1265-73. doi: [10.3233/BME-151424](https://doi.org/10.3233/BME-151424). [PubMed: [26405886](https://pubmed.ncbi.nlm.nih.gov/26405886/)].
3. Yang HC, Chang CH, Huang SW, Chou YH, Li PC. Correlations among acoustic, texture and morphological features for breast ultrasound CAD. *Ultrasound Imaging*. 2008;30(4):228-36. doi: [10.1177/016173460803000404](https://doi.org/10.1177/016173460803000404). [PubMed: [19507676](https://pubmed.ncbi.nlm.nih.gov/19507676/)].
4. Wang Y, Jiang S, Wang H, Guo YH, Liu B, Hou Y, et al. CAD algorithms for solid breast masses discrimination: Evaluation of the accuracy and interobserver variability. *Ultrasound Med Biol*. 2010;36(8):1273-81. doi: [10.1016/j.ultrasmedbio.2010.05.010](https://doi.org/10.1016/j.ultrasmedbio.2010.05.010). [PubMed: [20691917](https://pubmed.ncbi.nlm.nih.gov/20691917/)].
5. 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(1):3-9. doi: [10.14366/iusg.16012](https://doi.org/10.14366/iusg.16012). [PubMed: [27184656](https://pubmed.ncbi.nlm.nih.gov/27184656/)]. [PubMed Central: [PMC5207353](https://pubmed.ncbi.nlm.nih.gov/PMC5207353/)].
6. Lee SE, Moon JE, Rho YH, Kim EK, Yoon JH. Which supplementary imaging modality should be used for breast ultrasonography? Comparison of the diagnostic performance of elastography and computer-aided diagnosis. *Ultrasonography*. 2017;36(2):153-

9. doi: [10.14366/usg.16033](https://doi.org/10.14366/usg.16033). [PubMed: [27764908](https://pubmed.ncbi.nlm.nih.gov/27764908/)]. [PubMed Central: [PMC5381849](https://pubmed.ncbi.nlm.nih.gov/PMC5381849/)].
7. Shen WC, Chang RF, Moon WK, Chou YH, Huang CS. Breast ultrasound computer-aided diagnosis using BI-RADS features. *Acad Radiol*. 2007;**14**(8):928-39. doi: [10.1016/j.acra.2007.04.016](https://doi.org/10.1016/j.acra.2007.04.016). [PubMed: [17659238](https://pubmed.ncbi.nlm.nih.gov/17659238/)].
8. Chabi ML, Borget I, Ardiles R, Aboud G, Boussouar S, Vilar V, et al. Evaluation of the accuracy of a computer-aided diagnosis (CAD) system in breast ultrasound according to the radiologist's experience. *Acad Radiol*. 2012;**19**(3):311-9. doi: [10.1016/j.acra.2011.10.023](https://doi.org/10.1016/j.acra.2011.10.023). [PubMed: [22310523](https://pubmed.ncbi.nlm.nih.gov/22310523/)].
9. 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**(3):217-25. doi: [10.14366/usg.17046](https://doi.org/10.14366/usg.17046). [PubMed: [28992680](https://pubmed.ncbi.nlm.nih.gov/28992680/)]. [PubMed Central: [PMC6044219](https://pubmed.ncbi.nlm.nih.gov/PMC6044219/)].
10. Park HJ, Kim SM, La Yun B, Jang M, Kim B, Jang JY, 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**(3). e14146. doi: [10.1097/MD.00000000000014146](https://doi.org/10.1097/MD.00000000000014146). [PubMed: [30653149](https://pubmed.ncbi.nlm.nih.gov/30653149/)]. [PubMed Central: [PMC6370030](https://pubmed.ncbi.nlm.nih.gov/PMC6370030/)].
11. Di Segni M, de Soccio V, Cantisani V, Bonito G, Rubini A, Di Segni G, 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**(2):105-18. doi: [10.1007/s40477-018-0297-2](https://doi.org/10.1007/s40477-018-0297-2). [PubMed: [29681007](https://pubmed.ncbi.nlm.nih.gov/29681007/)]. [PubMed Central: [PMC5972107](https://pubmed.ncbi.nlm.nih.gov/PMC5972107/)].
12. 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**(1):209-16. doi: [10.1002/jum.14332](https://doi.org/10.1002/jum.14332). [PubMed: [28762552](https://pubmed.ncbi.nlm.nih.gov/28762552/)].
13. Sahiner B, Chan HP, Roubidoux MA, Hadjiiski LM, Helvie MA, Paramagul C, et al. Malignant and benign breast masses on 3D US volumetric images: Effect of computer-aided diagnosis on radiologist accuracy. *Radiology*. 2007;**242**(3):716-24. doi: [10.1148/radiol.2423051464](https://doi.org/10.1148/radiol.2423051464). [PubMed: [17244717](https://pubmed.ncbi.nlm.nih.gov/17244717/)]. [PubMed Central: [PMC2800986](https://pubmed.ncbi.nlm.nih.gov/PMC2800986/)].
14. Dromain C, Boyer B, Ferre R, Canale S, Delaloge S, Balleyguier C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur J Radiol*. 2013;**82**(3):417-23. doi: [10.1016/j.ejrad.2012.03.005](https://doi.org/10.1016/j.ejrad.2012.03.005). [PubMed: [22939365](https://pubmed.ncbi.nlm.nih.gov/22939365/)].
15. Gresens AA, Britt RC, Feliberti EC, Britt LD. Ultrasound-guided breast biopsy for surgical residents: Evaluation of a phantom model. *J Surg Educ*. 2012;**69**(3):411-5. doi: [10.1016/j.jsurg.2011.10.015](https://doi.org/10.1016/j.jsurg.2011.10.015). [PubMed: [22483146](https://pubmed.ncbi.nlm.nih.gov/22483146/)].
16. De Matheo LL, Geremia J, Calas MJG, Costa-Junior JFS, da Silva FFF, von Kruger MA, et al. PVCP-based anthropomorphic breast phantoms containing structures similar to lactiferous ducts for ultrasound imaging: A comparison with human breasts. *Ultrasonics*. 2018;**90**:144-52. doi: [10.1016/j.ultras.2018.06.013](https://doi.org/10.1016/j.ultras.2018.06.013). [PubMed: [29966842](https://pubmed.ncbi.nlm.nih.gov/29966842/)].
17. Groenhuis V, Visentin F, Siepel FJ, Maris BM, Dall'alba D, Fiorini P, et al. Analytical derivation of elasticity in breast phantoms for deformation tracking. *Int J Comput Assist Radiol Surg*. 2018;**13**(10):1641-50. doi: [10.1007/s11548-018-1803-x](https://doi.org/10.1007/s11548-018-1803-x). [PubMed: [29869320](https://pubmed.ncbi.nlm.nih.gov/29869320/)]. [PubMed Central: [PMC6153655](https://pubmed.ncbi.nlm.nih.gov/PMC6153655/)].
18. Ustbas B, Kilic D, Bozkurt A, Aribal ME, Akbulut O. Silicone-based composite materials simulate breast tissue to be used as ultrasonography training phantoms. *Ultrasonics*. 2018;**88**:9-15. doi: [10.1016/j.ultras.2018.03.001](https://doi.org/10.1016/j.ultras.2018.03.001). [PubMed: [29525227](https://pubmed.ncbi.nlm.nih.gov/29525227/)].