

Future Directions in the Era of Commercially Available AI for Breast US

Jung Hyun Yoon, MD, PhD¹, Jin Chung, MD, PhD², Won Hwa Kim, MD, PhD^{3,4}, Jaeil Kim, PhD^{4,5}

¹Department of Radiology, Severance Hospital, Research Institute of Radiological Science Yonsei University, College of Medicine, ²Department of Radiology, Mokdong Hospital, Ewha Womans University, College of Medicine, ³Department of Radiology, Kyungpook National University Chilgok Hospital, School of Medicine, Kyungpook National University, ⁴BeamWorks Inc., ⁵School of Computer Science and Engineering, Kyungpook National University

Artificial intelligence (AI) systems for breast US are now commercially available for clinical use. AI for breast US has high performances, but both breast US and AI systems are used in a variety of clinical settings that can affect outcomes of using AI for decision support. Among the studies using commercially available AI systems for breast US, studies commonly show that the diagnostic performances are significantly improved by using AI. There are several factors that users should consider for appropriate AI use such as the use case of the AI system, how the AI system generates results, and the differences in the level of suspicion given by different AI systems. For optimal AI integration, differences in workflow (technician vs. radiologist performed examinations) and clinical background (screening vs. diagnostic setting) should be considered. In this review, we will discuss the status of AI systems for breast US, the outcomes of using AI for breast US interpretation and what we need to consider when we use AI in our practice.

Index words: Breast; Ultrasonography; Artificial intelligence

Background

Ultrasonography (US) is widely used today for the diagnosis of breast diseases. Traditionally, breast US has been used to provide additional imaging information for differential diagnosis of symptomatic lesions or abnormalities detected on mammography

(1, 2). Breast US is a safe, easily available imaging tool that provides real-time scanning and can be used for guidance during biopsy procedures, but the major shortcoming of US is that both detection sensitivity and diagnostic performances depend highly on the operator's level of expertise (3, 4). To overcome this limitation, artificial intelligence (AI) based systems for breast US have been developed and are currently commercially available for clinical use (5, 6). Studies show that AI for breast US have high diagnostic performances, similar to the performances of radiologists, providing evidence that these systems can be safely used in our practice

Correspondence to: Jung Hyun Yoon, M.D., Ph.D.
Department of Radiology, Severance Hospital,
Research Institute of Radiological Science,
Yonsei University, College of Medicine 50-1
Yonsei-ro, Seodaemun-gu, Seoul, Korea
Tel. +82-2-2228-7400 Fax. +82-2-2227-8337
E-mail: lvjenny@yuhs.ac

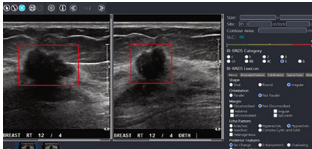
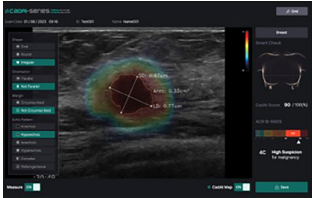
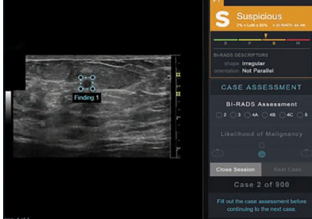
(7–9). Still, unlike mammography, US is used in various situations by different medical professionals that can significantly affect the outcomes of using AI for decision support. In this review, we will discuss the status of AI systems for breast US, the outcomes of using AI for breast US interpretation and what we need to consider when we use AI in our practice.

Commercially Available AI Systems for Breast US

There are several commercially available AI systems for handheld breast US (Table 1). S-Detect for Breast is embedded in the Medison US machine that is available exclusively for Samsung Medison US images while the remaining AI systems are standalone, vendor neutral devices. All AI systems commonly aid in lesion diagnosis, i.e., the

probability of cancer of the region-of-interest (ROI) that is either marked by the user or the AI system. Early systems such as Koios (Koios Medical Inc., USA) or S-Detect for Breast (Samsung Healthcare System, Samsung Medison, Korea) provides analysis of the user-provided area while BU-CAD (Taihao Medical Inc., Taiwan) or CadAI-B for Breast (BeamWorks Inc., Korea) provides autonomous lesion detection by the system (computer aided detection, CADe). Samsung Medison recently launched Live BreastAssist that provides automatic lesion detection, in which the diagnostic procedure works in a two-step procedure first guided by CADe (Live BreastAssist) and then computer aided diagnosis (CADx, S-Detect for Breast) (Fig. 1). CadAI-B for Breast employs a one-step approach for lesion detection and diagnosis, using heatmaps where the color scale ranging from blue (low probability of malignancy) to red (high probability

Table 1. Summary of Commercially Available Artificial Intelligence Systems for Breast Handheld US

AI system	Vendor	AI ability	Image display	AI output
BU-CAD	TaiHao Medical Inc., Taipei, Taiwan	CADe / CADx		- Automated ROIs (square) - Likelihood of malignancy: range 0-100 - BI-RADS descriptors - BI-RADS category (BI-RADS 2, 3, 4, 5)
CadAI-B for Breast	BeamWorks, Inc., Daegu, South Korea	CADe/ CADx		- Automated ROIs (heatmap) - Probability of malignancy: range 0-100 - BI-RADS descriptors - BI-RADS category (BI-RADS 1/2, 3, 4a, 4b, 4c, 5)
Koios DS	Koios Medical Inc., Chicago, IL	CADx		- User-selected ROIs - Four output categories (benign, probably benign, suspicious, probably malignant) - Continuous graphical confidence level within the category
S-Detect	Samsung Healthcare, Seoul, South Korea	CADx		- User-selected ROIs - BI-RADS descriptors - Two-scale assessment (possibly benign, possibly malignant)

CADe, Computer-aided detection; CADx, Computer-aided diagnosis; ROI, Region of interest.

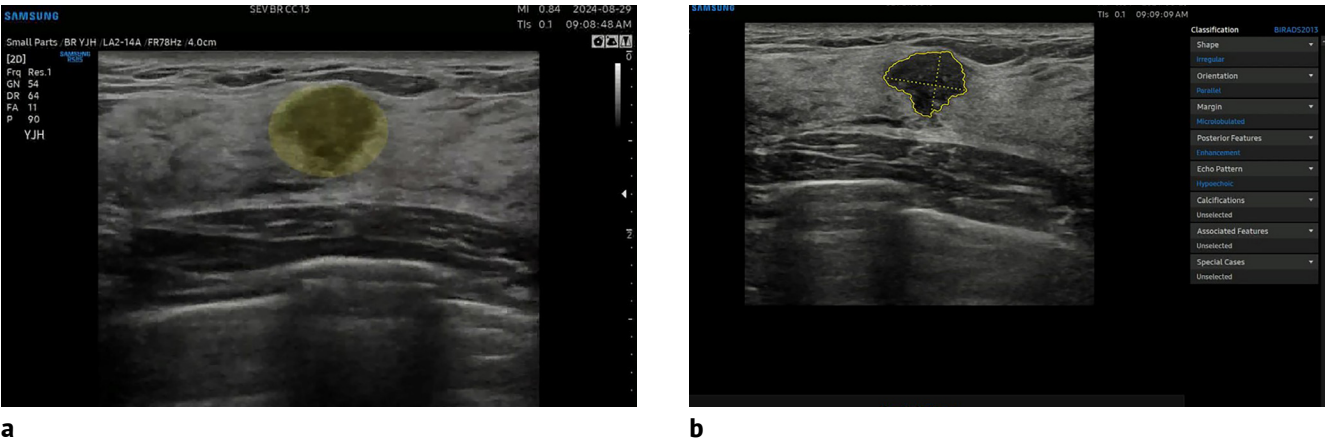


Fig. 1. Representative images of two-step detection and diagnosis using Live BreastAssist and S-Detect for Breast. Live BreastAssist highlights the mass area for the user (a, yellow circle) and S-Detect for Breast (b) analyzes this result sequentially for diagnostic support. ‘Possibly Malignant’ result was given by S-Detect for breast, and this mass was pathologically confirmed as invasive ductal carcinoma.

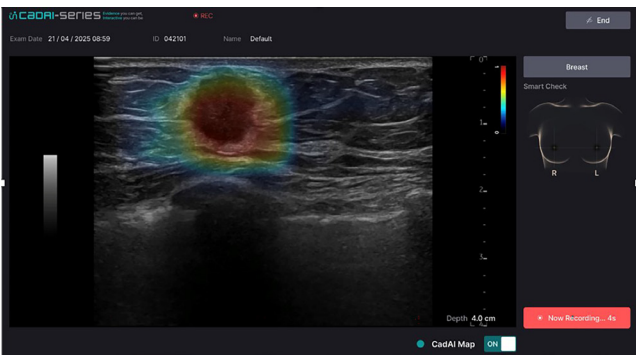


Fig. 2. Representative image of one-step detection and diagnosis using CadAI-B for Breast. Detection of the breast mass is highlighted using heatmaps where the color scale ranges from blue (low probability) to red (high probability) providing intuitive cancer probability simultaneously to lesion detection. CadAI-B BI-RADS 4c was given to this mass which was pathologically confirmed as invasive ductal carcinoma.

of malignancy) gives intuitive information of cancer probability during real-time scans (Fig. 2). AI outcomes differ according to each system, ranging from 2-scale ‘possibly benign’ vs. ‘possibly malignant’ to 6-scale BI-RADS categories including subcategories for BI-RADS 4.

Reported Outcomes of Using AI for Breast US

Table 2 summarizes the published studies evaluating commercially available AI systems for breast lesions. Most of the studies commonly report improved area under the receiving operator characteristics curve (AUC) after using AI for breast US interpretation (8, 10). Adding to this, studies report that using AI leads to improved specificity, PPV, and accuracy compared to grayscale US alone (7, 11–14). Reader-average area under the localization ROC was significantly improved after using CadAI-B for Breast in a multireader, multicase study, similar to studies using other AI systems (15). However, sensitivity and accuracy were improved after using AI for reader assistance with a tradeoff of decreased specificity in less experienced reader group.

In addition to improved diagnostic performances, using AI for US interpretation brought reduced variability among readers (8, 12, 13) that would have influenced performance changes. This finding is important in particular deciding upon the next step for the patient, that is, whether to recommend conservative imaging follow-up or biopsy. Less

Table 2. Summary of Publications Using Commercial AI Systems for Handheld Breast US

Study	AI system	Study design	Test data origin (Ethnicity if provided)	Number of patients or lesions	Readers/Operators	Changes in reader performance after AI application
Cho et al. (2018) (11)	S-Detect for Breast (Samsung Medison)	Retrospective, multi-reader	Korea	119 lesions Malignant, 54 (45.4%) Benign, 65 (54.6%)	2 breast radiologists	- ↑ Specificity, ↑ PPV, and ↑ accuracy of readers after AI
Choi et al. (2018) (12)	S-Detect for Breast (Samsung Medison)	Retrospective, multi-reader	Korea	200 patients Malignant, 12 (6.0%) Benign, 188 (94.0%)	4 readers - 2 breast radiologists - 2 first-year radiology residents	- ↑ Specificity and accuracy of readers - ↑ Interobserver agreement
Park et al. (2019) (13)	S-Detect for Breast (Samsung Medison)	Retrospective, multi-reader	Korea	100 lesions Malignant, 41 (41%) Benign, 59 (59%)	5 readers - 3 first-year fellowship trainees - 2 breast radiologists	- ↑ NPV and ↑ AUC in inexperienced readers - ↑ Specificity and ↑ PPV in experienced readers - ↑ Interobserver agreement
Mango et al. (2019) (8)	Koios DS for Breast (Koios Medical)	Retrospective, multi-center, multi-reader	United States (592 white (65.8%), 77 black (8.6%), 73 Hispanic (8.1%), 133 Asian (14.8%) and 25 other (2.8%))	900 lesions Malignant, 430 (47.8%) Benign, 470 (52.2%)	15 readers - 11 radiologists - 2 breast surgeons - 2 obstetricians/gynecologists	- ↑ Mean AUC of readers - ↑ Interobserver agreement - ↓ Intrareader variability
Berg et al. (2021) (18)	Koios DS for Breast (Koios Medical)	Retrospective, multi-reader	United States	319 lesions Malignant, 88 (27.6%) Benign, 222 (69.6%) Atypical, 9 (2.8%)	9 breast radiologists	- No significant difference in mean AUC in the original mode - ↑ Mean AUC in the high-sensitivity/specificity mode
Wei et al. (2022) (14)	S-Detect for Breast (Samsung Medison)	Retrospective, multi-reader	China	266 lesions Malignant, 69 (25.9%) Benign, 197 (74.1%)	4 radiologists - 1, 4, 9, 20 years of experience in breast imaging	- ↑ Accuracy, ↑ specificity, ↑ PPV, ↑ NPV ↑ AUC in less experienced readers - ↓ Unnecessary biopsy rate of less experienced readers

Continued

Table 2. Continued

Study	AI system	Study design	Test data origin (Ethnicity if provided)	Number of patients or lesions	Readers/Operators	Changes in reader performance after AI application
Lee et al. (2022) (7)	S-Detect for Breast (Samsung Medison)	Retrospective, multi-reader	Korea	492 lesions Malignant, 200 (40.7%) Benign, 292 (59.3%)	6 readers - 2 radiology residents, 1 fellow, 3 breast radiologists	- ↑ Mean specificity, PPV, accuracy of readers - ↑ Mean AUC of inexperienced readers in simultaneous reading - ↑ Interobserver agreement
Lai et al. (2022) (10)	BU-CAD (Taihao Medical Inc.)	Retrospective, multi-reader, multi-case	Taiwan	172 patients Malignant, 65 (37.7%) Benign, 107 (62.3%)	16 readers - 14 radiologists and 2 breast surgeons Average experience: 10.1 years (range 1-24 years)	- ↑ Mean AUC of readers - ↓ Mean interpretation time by 40%
He et al. (2023) (19)	S-Detect for Breast (Samsung Medison)	Prospective, Multi-center	China	313 lesions Malignant, 102 (32.6%) Benign, 211 (67.4%)	Radiologists without breast US expertise at secondary or rural hospitals	- ↑ Accuracy, ↑ specificity, ↑ PPV of readers - no significant difference in sensitivity or NPV - ↓ Frequency of benign biopsies by 79%
Yoon et al. (2024) (15)	CadAI-B for Breast (BeamWorks)	Retrospective, multi-reader, multi-case	Korea	1000 images Malignant, 500 (50%) Benign, 250 (25%) Negative, 250 (25%)	15 readers - 6 Radiologists - 3 Breast surgeons - 2 obstetricians/gynecologists, 1 family medicine physician - 3 Radiographers	- ↑ AULROC after using AI - ↑ Sensitivity, accuracy - ↓ Specificity in less experienced readers - Enhanced performances in less experienced readers such as general radiologists without breast fellowship training, physicians or radiographers

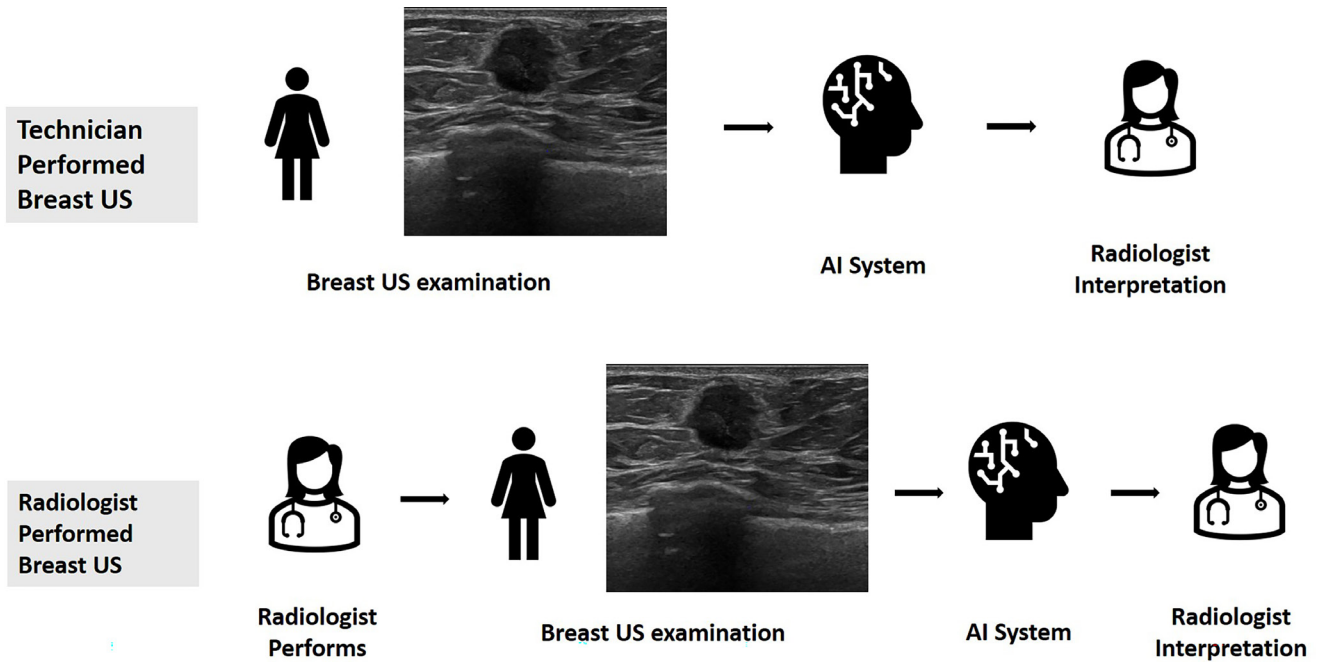


Fig. 3. Schema showing different workflows in breast US examination.

experienced readers tend to lack confidence in their interpretations leading to more benign biopsies, but the reduced variability among readers with AI assistance show that the clinical decision made based on the US examinations reach the level of experienced readers. AI can also aid readers in reducing interpretation time; Lai et al. reported approximately 40% reduction in interpretation time when using AI for assistance (10). Based on these results, we consider AI to have additional roles such as supporting less experienced readers to improve performance in areas of limited medical resources (16, 17) or prioritizing examinations for radiologists among batch examinations.

Needs for Safe AI Integration

Current publications support that using AI systems for breast US are ready for realworld application. Before considering clinical implementation, however, users should acknowledge the following factors for appropriate use. First, users should know what the use case of the AI system, that

is, the specific task for which the AI system was developed. For instance, most of the systems do not specify whether they can be applied on nonmass lesions, a specific finding that we come across not infrequently during US examinations. Second, users should be trained on how the system activates. For AI systems that need user-selected ROI for analysis, results can vary significantly according to where the ROI box is set for analysis. On the contrary, incorrect marking of automatically generated ROIs can mislead readers to false positive/negative interpretations. Third, users should acknowledge the level of suspicion that AI output categories withhold for accurate interpretation. AI outputs generated by the systems are initially continuous values (probability of malignancy) that is calibrated to the categorical outputs of the system. Each system produces categorical outputs that has different probability or likelihood of malignancy as displayed in Table 3. For instance, the BI-RADS 3 output for BU-CAD has probability range of 26–50% while the BI-RADS 3 for CadAI-B for Breast is at less than 2%. As most of the users are familiar with

Table 3. Differences in Probability of Malignancy for Categorical Outcomes According to the Commercially Available AI Systems for Breast US

	Categorical outputs	Probability/Likelihood of Malignancy	Note
S-Detect for Breast	Possibly Benign	0-50%	
	Possibly Malignant	51-100%	
Koios DS	Benign	<0.5%	Aligns to BI-RADS 2
	Probably Benign	<2%	BI-RADS 3
	Suspicious	2-50%	BI-RADS 4A, 4B
	Probably Malignant	>50%	BI-RADS 4C, 5
BU-CAD	BI-RADS 2	0-26%	
	BI-RADS 3	26-51%	
	BI-RADS 4	51-98%	
	BI-RADS 5	98-100%	
CadAI-B for Breast	BI-RADS 1/2	0%	
	BI-RADS 3	<2%	
	BI-RADS 4A	2-10%	
	BI-RADS 4B	11-50%	
	BI-RADS 4C	51-95%	
	BI-RADS 5	96-100%	

the positive predictive values that each BI-RADS category stand for, users can misinterpret AI output categories to align with the American College of Radiology (ACR) recommendations.

Future Investigations and Perspectives

Unlike mammography, various medical professions perform and even interpret breast US images. When considering integration of AI for breast US, we expect the differences in workflow to affect the outcomes for using AI in US interpretation. In the technician-performed breast US workflow, AI results are given simultaneously with the grayscale images to the radiologist for interpretation, i.e., simultaneous reading. In contrast, when the radiologist performs breast US examinations, AI results are given after full assessment of grayscale US features of the target lesion, i.e., sequential reading. In a study comparing the effect of using AI for breast US in two different scenarios, inexperienced readers had significantly increased performances when using AI in simultaneous

reading (17). In addition, the role of AI may differ according to which workflow it is being used, for instance in the technician-performed workflow, AI may be used for triaging images/examinations for the radiologists while in the radiologist-performed workflow it may be used to provide confidence or second opinion. How AI integrated in different workflows would need further investigation to validate the true outcomes with future prospective studies.

The display or performances of AI systems can also affect the outcomes of using AI in breast US. The reader study using CadAI-B for Breast showed significantly improved sensitivity with a trade-off or decreased specificity in less experienced readers (Table 2) (15). Unlike other AI systems, CadAI-B for Breast displays results in colored heatmap that not only alerts the users to lesion detection but gives them simultaneous impression of the suspicion level that can influence the decision of the reader. For the inexperienced readers, lesions highlighted in red heatmaps may not be easy to neglect regardless of the initial impression of the lesion that may

have been the cause for higher sensitivity using this AI system. Standalone performance of AI itself is also important and should be considered when selecting AI systems for clinical use. In the study by Berg et al. results showed that radiologists showed improved performances and responded more when the AI system was set at high-specificity mode (18). The clinical environment where AI is used, i.e., supplemental screening vs. diagnostic examinations, requires different AI performances. Realtime detection ability with high sensitivity would be more beneficial in the screening environment, while AI systems set at high specificity may be more effective in reducing benign biopsies in the diagnostic setting.

Finally, we need consensus among breast radiologists for appropriate use of AI systems. Little is established on how to educate users to get the optimal AI results that should include image quality or AI/user factors that can affect results. Regulations for system maintenance, including regular updates, should be established to keep up with the advances in technology. Based on studies evaluating factors that can affect outcomes of using AI for breast US discussed above, we anticipate guidelines for appropriate AI integration and maintenance programs to provide more efficient and advanced medical services for our patients.

In conclusion, AI systems can aid readers in enhancing performances in breast US examinations. For safe and accurate implementation, users should be aware that differences in how AI systems generate results and AI output categories can significantly affect the outcomes. With the technical advances, the detection and diagnostic ability of AI systems are integrated within a single workflow. How these advances would affect outcomes of using AI for breast US needs to be investigated in near future.

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;299:2151-2163.
2. Dempsey PJ. The history of breast ultrasound. *J Ultrasound Med* 2004;23:887-894.
3. Lee HJ, Kim EK, Kim MJ, et al. Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. *Eur J Radiol* 2008;65:293-298.
4. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology* 2006;239:385-391.
5. Kim J, Kim HJ, Kim C, Kim WH. Artificial intelligence in breast ultrasonography. *Ultrasonography* 2021;40:183-190.
6. Villa-Camacho JC, Baikpour M, Chou SS. Artificial Intelligence for Breast US. *J Breast Imaging* 2023;5:11-20.
7. Lee SE, Han K, Youk JH, et al. Differing benefits of artificial intelligence-based computer-aided diagnosis for breast US according to workflow and experience level. *Ultrasonography* 2022;41:718-727.
8. Mango VL, Sun M, Wynn RT, Ha R. Should We Ignore, Follow, or Biopsy? Impact of Artificial Intelligence Decision Support on Breast Ultrasound Lesion Assessment. *AJR Am J Roentgenol* 2020;214:1445-1452.
9. Bahl M, Chang JM, Mullen LA, Berg WA. Artificial Intelligence for Breast Ultrasound: AJR Expert Panel Narrative Review. *AJR Am J Roentgenol* 2024;223:e2330645.
10. Lai YC, Chen HH, Hsu JF, Hong YJ, Chiu TT, Chiou HJ. Evaluation of physician performance using a concurrent-read artificial intelligence system to support breast ultrasound interpretation. *Breast* 2022;65:124-135.
11. 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.
12. 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.

13. Park HJ, Kim SM, La Yun B, 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.
14. Wei Q, Zeng SE, Wang LP, et al. The Added Value of a Computer-Aided Diagnosis System in Differential Diagnosis of Breast Lesions by Radiologists With Different Experience. *J Ultrasound Med* 2022;41:1355-1363.
15. Yoon JH, Han K, Kim HW, Kim WH, Kim J. Evaluating the benefit of using AI system for breast US: Results of a multicenter, reader study Poster presentation at the Radiological Society of North America 2024. McCormick Place, Chicago, IL, USA, 2024.
16. Berg WA, López Aldrete AL, Jairaj A, et al. Toward AI-supported US Triage of Women with Palpable Breast Lumps in a Low-Resource Setting. *Radiology* 2023;307:e223351.
17. Lee S, Lee HS, Lee E, Kim WH, Kim J, Yoon JH. Improving breast ultrasonography education: the impact of AI-based decision support on the performance of non-specialist medical professionals. *Ultrasonography* 2025;44:124-133.
18. Berg WA, Gur D, Bandos AI, et al. Impact of Original and Artificially Improved Artificial Intelligence-based Computer-aided Diagnosis on Breast US Interpretation. *J Breast Imaging* 2021;3:301-311.
19. He P, Chen W, Bai MY, et al. Deep Learning-Based Computer-Aided Diagnosis for Breast Lesion Classification on Ultrasound: A Prospective Multicenter Study of Radiologists Without Breast Ultrasound Expertise. *AJR Am J Roentgenol* 2023;221:450-459.

유방 초음파 진단 보조 인공지능의 상용화 시대의 앞으로의 방향

윤정현¹ · 정진² · 김원화^{3,4} · 김재일^{4,5}

¹연세대학교 의과대학 세브란스병원 영상의학과, ²이화여자대학교 의과대학 이대목동병원 영상의학과,

³경북대학교 의과대학 칠곡경북대학교 병원 영상의학과, ⁴(주) 빔웍스, ⁵경북대학교 컴퓨터공학부

유방초음파 진단 보조 인공지능 시스템은 현재 상용화되어 일상에서 사용되고 있다. 이러한 인공지능 시스템은 높은 진단 성적을 보이고 있지만 유방 초음파의 경우 다양한 임상 환경에서 다양한 사용자에 의해 이루어지기 때문에 인공지능 시스템을 사용함에 따라 결과가 영향을 받을 수 있다. 인공지능을 사용하기에 앞서 사용자는 인공지능 시스템의 개발 목적, 결과 산출 방법과 결과 해석에 대해 숙지하고 판독 보조로 사용해야 한다. 또한, 인공지능 진단 보조를 이용할 경우 유방 초음파 검사자 혹은 진단 환경에 따라 인공지능 시스템을 적용한 결과가 달라질 수 있어 차후 전향 연구를 통해 검증이 필요하다. 본 중설에서는 현재 초음파용 인공지능 시스템의 현황을 정리하고 향후 유방영상 전문의들이 최상의 진단 효과를 얻기 위해 고려해야 할 점들에 대해 정리해보고자 한다.

Index words: Breast; Ultrasonography; Artificial intelligence

Corresponding author: Jung Hyun Yoon, M.D., Ph.D.