Editorial

Artificial Intelligence for Breast Ultrasound: Will It Impact Radiologists' Accuracy?

Manisha Bahl, MD, MPH[1],*

[1]Massachusetts General Hospital, Department of Radiology, Boston, MA, USA

*Address correspondence to M.B. (e-mail: mbahl1@mgh.harvard.edu)

In the setting of breast density notification legislation and the attendant interest in supplemental screening for women with dense breast tissue, screening US has become widely integrated into breast imaging practices (1,2). Screening US detects additional invasive cancers in women with dense breast tissue, with an incremental cancer detection rate averaging 2.1–2.7 per 1000 examinations (3). However, this modality is currently limited by its high recall rate and low positive predictive value of biopsy (1–3). In fact, a recent review article reported that the positive predictive value of biopsies prompted by screening US examinations averages only 9%–11% (3). These limitations, in addition to the wide variability in performance metrics across studies, have fueled interest in the application of artificial intelligence (AI) to screening breast US examinations in order to avoid unnecessary recalls and biopsies (4–10).

AI offers the potential of improved accuracy, speed, and quality of breast imaging interpretation. While traditional computer-aided detection (CADe) and computer-aided diagnosis (CADx) are programmed based on human-engineered features, such as shape and margins, AI algorithms can learn the necessary features to categorize a lesion as benign or malignant, discover features that are not perceptible by humans, and continually improve with exposure to more images (11,12). AI algorithms for image interpretation use deep learning, which is based on neural networks with multiple layers that first learn to recognize pixels, then edges and shapes, and then more complex shapes and higher-level features (12,13). Deep learning-based clinical decision support tools are now commercially available for breast imaging applications.

In this issue of the *Journal of Breast Imaging*, Berg and colleagues evaluate the standalone performance of one of these commercially available AI-based CAD systems for breast US interpretation (Koios Medical, Chicago, IL) and its impact on radiologist performance (14). For a user-selected region-of-interest containing the breast lesion, the AI-based CAD system generates a probability of cancer that is translated into a categorical output (such as "probably benign"). One unique aspect of the reader study by Berg and colleagues was that the authors had access to a research version of the AI-based CAD system and thus were able to set it to one of three modes: the *original mode* with outputs of benign, probably benign, suspicious, or malignant; a *high-sensitivity* mode with outputs of benign or malignant; and a *high-specificity* mode also with outputs of benign or malignant. Nine breast imaging radiologists interpreted US images, mostly from whole-breast screening US examinations, of 319 lesions (enriched with 88 cancers), with and without AI support in each of the three modes.

Although use of the original mode did not impact radiologists' accuracy, as measured by the mean area under the receiver operating characteristic curve (AUC), both the high-sensitivity and high-specificity modes led to improvements (14). For the original mode, the standalone AUC of the AI system was 0.77, and the radiologists' mean AUC was 0.82 with and without AI support ($P = 0.92$). For the high-sensitivity mode, the standalone AUC of the AI system was 0.86, and the radiologists' mean AUC was higher with AI support (0.88 versus 0.83, $P < 0.001$). For the high-specificity mode, the standalone AUC of the AI system was 0.88, and the radiologists' mean AUC was also higher with AI support (0.89 versus 0.82, $P < 0.001$). With each of the three modes, radiologists changed their interpretations in approximately one-quarter of the cases (23% with the original mode, 24% with the high-sensitivity mode, and 26% with the

high-specificity mode). The authors' main conclusion is that radiologists improved their performance and were more responsive to the AI-based CAD system in the high-sensitivity and high-specificity modes, particularly in the high-specificity mode, which had fewer false-positive cues.

Evidence about the utility of AI systems is based largely on reader studies, such as this one, but the behavior of radiologists may differ in real-world clinical practice. The actual impact of an AI system in clinical practice may be influenced by several factors, which include the radiologist's confidence in the AI system, the radiologist's confidence in his or her own independent interpretation, the accessibility of the rationale being used by the AI system, and the radiologist's interactions with the AI system (eg, the number of clicks needed to access the AI output) (15). With regard to confidence in the AI system, Berg and colleagues suggest that radiologists are likely to trust the recommendations of more specific AI-based CAD, as demonstrated by their acceptance of a higher proportion of "malignant" CAD cues in the high-specificity mode of the CAD system (which produces the fewest number of malignant cues) (14). The authors suggest that their observation of higher radiologist acceptance and responsiveness in the setting of fewer false-positive cues should be taken into consideration when developing and implementing AI algorithms.

One other unanswered question about AI systems is their degree of impact on radiologists with differing levels of experience and expertise. In the study by Berg and colleagues, all nine radiologist readers were specialists in breast imaging or currently in breast imaging fellowship, with experience ranging from 0.5 to 29 years (14). Neither the group of radiologists with less than 10 years of experience nor the group with more than 10 years of experience showed improvement with the AI-based CAD system in the original mode, and both groups had similar improvements with the high-sensitivity and high-specificity modes. However, in a reader study by Mango et al with the same AI-based CAD system for breast US interpretation (Koios Medical, Chicago, IL), the degree of improvement varied with the radiologist reader's initial operating point (8). Experienced subspecialized breast imaging radiologists may not benefit from AI-based CAD systems if the system does not perform better than or at least at the same level as experts, or if the system's output is less likely to be trusted and accepted by experienced readers. It is possible, however, that AI systems could help radiologists with less experience or those who are not fellowship-trained in breast imaging achieve better performance, which could ultimately improve the quality of breast imaging across the world (16).

The study by Berg and colleagues shows that AI-based CAD can improve radiologists' accuracy in classifying breast lesions on US as benign or malignant, which has become an increasingly important application of AI in the setting of widespread use of supplemental screening US for women with dense breast tissue (14). The authors also found that breast imaging specialists are more likely to act appropriately on the output generated by the AI system if fewer false-positive cues are provided, which should be taken into consideration as AI system development continues to progress. Further improvements in AI-based CAD for breast US interpretation are expected, and thorough validation of these systems with large studies, diverse populations, and prospective study design in real-world clinical environments are necessary before widespread deployment.

Funding

Conflict of Interest Statement

Dr Bahl is a consultant for Lunit (medical AI software company) and an expert panelist for 2nd.MD (digital health company). There are no other conflicts of interest.

References

1. Brem RF, Lenihan MJ, Lieberman J, Torrente J. Screening breast ultrasound: past, present, and future. *AJR Am J Roentgenol* 2015;204(2):234–240.

2. Butler RS, Hooley RJ. Screening breast ultrasound: update after 10 years of breast density notification laws. *AJR Am J Roentgenol* 2020;214(6):1424–1435.

3. Berg WA, Vourtsis A. Screening breast ultrasound using handheld or automated technique in women with dense breasts. *J Breast Imaging* 2019;1(4):283–296.

4. Barinov L, Jairaj A, Becker M, et al. Impact of data presentation on physician performance utilizing artificial intelligence-based computer-aided diagnosis and decision support systems. *J Digit Imaging* 2019;32(3):408–416.

5. Choi JS, Han BK, Ko ES, et al. Effect of a deep learning framework-based computer-aided diagnosis system on the diagnostic performance of radiologists in differentiating between malignant and benign masses on breast ultrasonography. *Korean J Radiol* 2019;20(5):749–758.

6. 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(3):e14146.

7. Heller SL, Wegener M, Babb JS, Gao Y. Can an artificial intelligence decision aid decrease false-positive breast biopsies? *Ultrasound Q* 2020;37(1):10–15.

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(6):1445–1452.

9. Dong F, She R, Cui C, et al. One step further into the blackbox: a pilot study of how to build more confidence around an AI-based decision system of breast nodule assessment in 2D ultrasound. *Eur Radiol* 2021 [Online ahead of print].

10. Kim S, Choi Y, Kim E, et al. Deep learning-based computer-aided diagnosis in screening breast ultrasound to reduce false-positive diagnoses. *Sci Rep* 2021;11(1):395.

11. Geras KJ, Mann RM, Moy L. Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. *Radiology* 2019;293(2):246–259.

12. Bahl M. Artificial intelligence: a primer for breast imaging radiologists. *J Breast Imaging* 2020;2(4):304–314.

13. Tang A, Tam R, Cadrin-Chênevert A, et al; Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian Association of Radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J* 2018;69(2):120–135.

14. Berg WA, Gur D, Bandos AI, et al. Impact of original and artificially improved AI-based CADx on breast US interpretation. *J Breast Imaging* 2021;3(3):301–311.

15. Hsu W, Hoyt AC. Using time as a measure of impact for AI systems: implications in breast screening. *Radiol Artif Intell* 2019;1(4):e190107.

16. Bahl M. Detecting breast cancers with mammography: will AI succeed where traditional CAD failed? *Radiology* 2019;290(2):315–316.