



# Application of an artificial intelligence-assisted diagnostic system for breast ultrasound: a prospective study

Zhi-Ying Jin, Jun-Kang Li, Rui-Lan Niu, Nai-Qin Fu, Ying Jiang, Shi-Yu Li, Zhi-Li Wang

Department of Ultrasound, The First Medical Center, Chinese PLA General Hospital, Beijing, China

**Contributions:** (I) Conception and design: ZY Jin, JK Li, ZL Wang; (II) Administrative support: RL Niu, ZL Wang; (III) Provision of study materials or patients: NQ Fu, Y Jiang, SY Li; (IV) Collection and assembly of data: ZY Jin, JK Li, NQ Fu; (V) Data analysis and interpretation: ZY Jin, RL Niu, Y Jiang, JK Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

**Correspondence to:** Zhi-Li Wang, MD. Department of Ultrasound, The First Medical Center, Chinese PLA General Hospital, 28 Fuxing Road, Beijing 100853, China. Email: wzllg@sina.com.

**Background:** Accurate diagnosis of breast cancer is of great importance to improve the prognosis of patients. Artificial intelligence (AI)-assisted diagnostic system for breast ultrasound is gradually being applied in the identification of benign and malignant breast lesions. This study aimed to evaluate the diagnostic performance and optimal application of AI-assisted ultrasonography for breast lesions in clinical setting.

**Methods:** A total of 501 consecutive patients with 679 breast lesions were prospectively included in the study. Junior and senior radiologists were asked to interpret images of lesions with and without AI assistance, respectively. Three application modes of AI were employed: AI alone, adjusted Breast Imaging Reporting and Data System (BI-RADS; incorporating BI-RADS obtained by AI into BI-RADS obtained by radiologists), and second reading mode (combining characteristic information extracted by AI to conduct a second reading so as to obtain a new BI-RADS). The diagnostic performances of these application modes were analyzed and compared.

**Results:** The area under the curve (AUC) of junior radiologists increased from 0.879 to 0.921 in BI-RADS<sub>second reading</sub>, which was higher than that in BI-RADS<sub>adjusted</sub> (0.901), similar to that in AI alone (0.924), and lower than that obtained by senior radiologists (0.950). Using BI-RADS category 4A as the threshold, the sensitivity of junior radiologists was found to increase from 0.83 to 0.92 ( $P<0.001$ ). Furthermore, the specificity increased from 0.79 to 0.85, which was higher than those of AI alone and BI-RADS<sub>adjusted</sub> ( $P<0.001$ ). The unnecessary biopsy rate decreased by 14.70% ( $P=0.01$ ). For senior radiologists, the sensitivity increased from 0.91 to 0.96 ( $P=0.01$ ). Similar results were observed in the subgroup analysis of lesions  $\leq 2$  cm. For lesions  $>2$  cm, only the specificity of junior radiologists increased from 0.39 to 0.52 ( $P=0.03$ ).

**Conclusions:** AI-assisted ultrasound is useful for the diagnosis of breast lesions, particularly for junior radiologists and lesions  $\leq 2$  cm. The use of the second reading mode can achieve excellent diagnostic performance.

**Keywords:** Breast cancer; ultrasonography; artificial intelligence (AI); diagnosis

Submitted May 31, 2024. Accepted for publication Nov 05, 2024. Published online Dec 10, 2024.

doi: 10.21037/gs-24-213

**View this article at:** <https://dx.doi.org/10.21037/gs-24-213>

## Introduction

Breast cancer is the most commonly diagnosed malignant disease in women and the leading cause of cancer death worldwide; it has an incidence of 24.5% and a mortality of 15.5% (1). An estimated 29 million new cases of breast

cancer in women and four million deaths were reported in 2022 (2). According to a previous study, breast cancer diagnosed at an early stage has a 5-year survival rate of 90%, with a better prognosis than middle or advanced breast cancer (3). Hence, early identification and treatment

of breast cancer are the most effective strategies to reduce the rate of breast cancer mortality.

Several tools have been used to screen breast cancer, mainly ultrasound, mammography, and magnetic resonance imaging (MRI). Mammography is widely employed for breast cancer screening to reduce breast cancer-related mortality; however, dense mammary gland may decrease the sensitivity of the mammography and lead to a false diagnosis (4). MRI has a higher diagnostic efficiency than mammography; however, its price and time cost limit its application in breast cancer screening. Meanwhile, ultrasound is a supplementary tool widely used in breast cancer screening, particularly for intensive mammary glands. However, ultrasound image characteristics of benign and malignant breast lesions can overlap with conventional ultrasound examination (5,6). The diagnostic accuracy of ultrasound imaging is largely dependent on the operator's personal experience and technology. The interpretation of ultrasound findings of each radiologist varies between observers. In recent years, the combination of artificial intelligence (AI) and medical imaging technologies has become possible. Previous studies have reported that the emergence of an AI-assisted ultrasound diagnostic system helps overcome the limitations of ultrasound imaging (7-9). To date, computer-aided decision tools for breast ultrasound, such as S-Detect, have been mainly applied in

clinical setting (10-12). Systematic evaluation of the value of AI in real-world clinical setting has not yet been conducted. Furthermore, it remains unclear how AI can maximize its effectiveness in clinical setting. The diagnostic accuracy of ultrasound imaging is related to the size of breast lesions. The role of AI in the diagnosis of breast lesions with different sizes is unknown. Moreover, whether AI can improve the diagnostic efficiency of small breast lesions remains unclear. Therefore, this study aimed to investigate the diagnostic performance and optimal use of AI-assisted system in diagnosing breast lesions in clinical setting. We present this article in accordance with the STARD reporting checklist (available at <https://gs.amegroups.com/article/view/10.21037/gs-24-213/rc>).

## Methods

### Patients

This prospective study was approved by the Ethics Committee of the Chinese PLA General Hospital (No. S2021-683-01), and informed consent was obtained from all patients. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). From January 2022 to May 2022, a total of 538 consecutive patients with 716 lesions who underwent conventional ultrasound examination in our hospital and were assessed using an AI-assisted diagnostic system were included in this study. The exclusion criteria were (I) poor-quality ultrasound image data or incomplete AI information (n=12), (II) loss to follow-up or less than 1 year of follow-up (n=22); and (III) neoadjuvant chemotherapy treatment before ultrasound examination (n=3). Finally, 501 criteria-matched patients with 679 lesions were included in final analysis. Of these lesions, 346 were subjected to percutaneous biopsy or surgical biopsy and 333 were followed up for more than 1 year. The lesions that remained stable in size and morphology over 1 year of ultrasound follow-up were defined as benign (*Figure 1*).

### Equipment

All ultrasound examinations were conducted using the Mindray Resona 7s ultrasound system (Mindray Medical International Co., Shenzhen, China) with a L11-3 linear array probe at 5.6–10.0 MHz. The AI-assisted ultrasound diagnostic system used in this study was Yizhun AI (Beijing Yizhun Intelligent Technology Co., Ltd., Beijing, China),

### Highlight box

#### Key findings

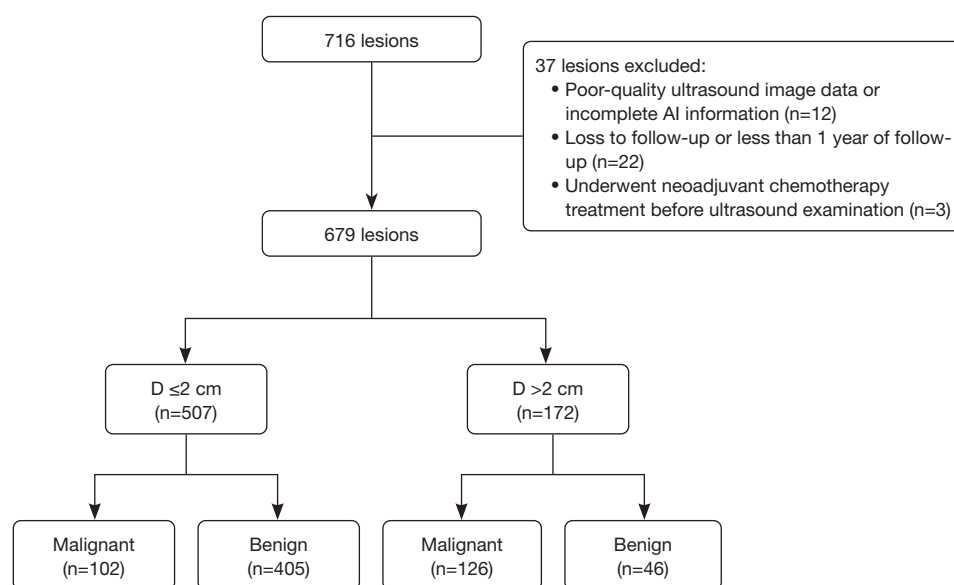
- Artificial intelligence (AI) is an effective auxiliary diagnostic tool for the differentiation of benign and malignant breast lesions.
- Junior radiologists can significantly improve their diagnostic performance with the help of AI, especially for breast lesions  $\leq 2$  cm.
- Utilizing of the second reading mode can achieve excellent diagnostic performance.

#### What is known and what is new?

- In recent years, the combination of AI and medical imaging technologies has become possible. However, systematic evaluation of the value of AI in real-world clinical setting has not yet been conducted.
- AI-assisted ultrasound is useful for diagnosing breast lesions, particularly for junior radiologists and lesions  $\leq 2$  cm.

#### What is the implication, and what should change now?

- This study investigated the diagnostic performance and optimal use method of AI-assisted system for breast ultrasound in clinical setting.



**Figure 1** Flowchart of the process of patient enrollment. AI, artificial intelligence; D, largest diameter of breast lesions.

which was trained using deep learning techniques. The AI system mainly consists of a host and a display screen that can be connected to any desktop ultrasound machine. While the radiologist is performing a dynamic breast scan, the screen of the AI system displays the ultrasound image in real-time. In this process, the AI system can automatically detect lesions, even those that flash in milliseconds. Moreover, it can provide ultrasound features according to the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS), such as characteristic information about breast lesions, including shape (oval, round, or irregular), orientation (parallel or not parallel), margin (circumscribed or not circumscribed), echo pattern (hypoechoic or complex cystic and solid), posterior features (no posterior features, shadowing, enhancement, or combined pattern), and calcifications (no calcifications, microcalcification, coarse calcifications, or combined calcification). By analyzing these features, breast lesions can be classified according to the BI-RADS.

### Ultrasound examination procedures

All ultrasound examinations were conducted by one radiologist (Y.J.; with 8 years of experience in breast imaging). The participants were placed in the supine or lateral position with both the upper limbs raised. The entire breast was carefully scanned, without missing any part, and when a lesion was found, the dynamic imaging video

was recorded for several minutes and stored for offline analysis. In addition, the maximum diameter of each lesion was measured three times, and then the mean value was recorded.

### Image analysis

All images were analyzed according to the BI-RADS and categorized into 3, 4A, 4B, 4C, and 5. The analyses were independently conducted by two junior radiologists (Z.Y.J. and N.Q.F.; with 2 and 3 years of experience in breast imaging, respectively) and two senior radiologists (J.K.L. and Z.L.W.; with 10 and 15 years of experience in breast imaging, respectively) who were blinded to the pathological information of the lesion. The final decisions of radiologists with equivalent seniority were reached by consensus. The image analysis mainly involved three stages: first, the radiologists analyzed the images without the assistance of AI to obtain the BI-RADS<sub>initial</sub>. Second, the AI system was applied to analyze images, automatically extract ultrasonic features, and obtain BI-RADS<sub>AI</sub>. The first and second stages could be carried out simultaneously. Third, in order to explore the most suitable application of AI, two modes were employed to combine radiologists and AI. The radiologists combined the characteristic information extracted by AI, evaluated the ultrasound images via second reading according to personal experience, and then obtained the final BI-RADS<sub>second reading</sub>. In addition, the radiologists

**Table 1** Characteristics of patients and breast lesions

Characteristic	Total	D ≤2 cm	D >2 cm
Patients	501	335	166
Age (years)	46±11	44±10	49±13
Lesions	679 (100.00)	507 (74.67)	172 (25.33)
Largest diameter (cm)	1.50±1.39	0.88±0.43	3.33±1.61
Malignant	228 (33.58)	102 (20.12)	126 (73.26)
Benign	451 (66.42)	405 (79.88)	46 (26.74)

Data are presented as n, mean ± standard deviation, and n (%). D, largest diameter of breast lesions.

combined the BI-RADS<sub>AI</sub> to obtain BI-RADS<sub>adjusted</sub> according to the following protocol: when BI-RADS<sub>AI</sub> is category 3, the BI-RADS<sub>initial</sub> is decreased by one level (BI-RADS<sub>initial</sub> category 3 remains unchanged); when BI-RADS<sub>AI</sub> is category 4A, the BI-RADS<sub>initial</sub> remains unchanged; and when BI-RADS<sub>AI</sub> is category 4B or above, the BI-RADS<sub>initial</sub> is increased by one level (BI-RADS<sub>initial</sub> category 5 remains unchanged). In general, we allowed BI-RADS category 4 (suspicious) assessment by radiologists could be upgraded or be downgraded by addition of BI-RADS<sub>AI</sub>.

### Statistical analysis

SPSS 25.0 (SPSS Inc., Chicago, IL, USA) and MedCalc 20.0 (MedCalc Software Ltd., Ostend, Belgium) were used for the statistical analysis. Numerical data with a normal distribution were expressed as mean ± standard deviation and compared using Student *t*-test. Otherwise, they were expressed as medians and interquartile ranges and then compared using the Mann-Whitney *U* test is used. Qualitative data were expressed as numbers (percentages) and compared using the Pearson Chi-squared test or Fisher exact test. To satisfy the clinical application environment, the lesions were divided into the high- and low-risk groups with threshold values of BI-RADS category 3 (BI-RADS category 3 was classified as negative for cancer, whereas 4A, 4B, and 4C were classified as positive for cancer) and 4A (BI-RADS 3 and 4A were classified as negative for cancer, whereas 4B and 4C were classified as positive for cancer), respectively. High-risk lesions are recommended for biopsy, whereas low-risk one can be followed up regularly. The sensitivity and specificity of BI-RADS<sub>initial</sub>, BI-RADS<sub>AI</sub>, BI-RADS<sub>second reading</sub>, and BI-RADS<sub>adjusted</sub> were calculated, with BI-RADS categories of 3 and 4A as two different thresholds and using pathological and follow-up results as a reference

standard. Sensitivity and specificity were compared using the McNemar test. Furthermore, the receiver operating characteristic curves were drawn and compared using the DeLong test. Unnecessary biopsy rate was defined as the number of benign lesions among the biopsy-recommended lesions. *P*<0.05 was considered statistically significant.

## Results

### Patient characteristics

All participants were female with a mean age of 46±11 years (range, 22–88 years). The median size of all lesions was 1.0 cm (interquartile range, 0.6–2.0 cm; range, 0.3–10.7 cm). The basic characteristics of all the patients are presented in *Table 1*. Using the histopathological results of biopsy and follow-up outcomes, 228 malignant and 451 benign lesions were found. No adverse events were reported during ultrasound examinations and AI application (*Table 1*).

### Diagnostic performance without assistance of the AI system

The area under the curve (AUC) of all lesions, lesions >2 cm, and lesions ≤2 cm were significantly higher in senior than in junior radiologists (0.950 *vs.* 0.879, *P*<0.001; 0.930 *vs.* 0.744, *P*<0.001; and 0.935 *vs.* 0.864, *P*<0.001, respectively) (*Table 2*).

When BI-RADS category 3 was used as the threshold, the sensitivity and specificity of senior and junior radiologists for all breast lesions were 99.1% *vs.* 97.4% and 68.5% *vs.* 70.3%, respectively. For lesions size ≤2 cm, the sensitivity and specificity were 98.0% *vs.* 94.1% and 73.3% *vs.* 75.3%, respectively (*Table S1*), whereas for breast lesions >2 cm, they were 100.0% *vs.* 100.0% and 26.1% *vs.* 26.1%, respectively. When BI-RADS category 4A was used as the

**Table 2** Diagnostic performance of radiologists with and without assistance of AI system using BI-RADS category 4A as threshold

Radiologist	Sensitivity (%)			Specificity (%)			AUC (95% CI)		
	BI-RADS <sub>initial</sub>	BI-RADS <sub>adjusted</sub>	BI-RADS <sub>second reading</sub>	BI-RADS <sub>initial</sub>	BI-RADS <sub>adjusted</sub>	BI-RADS <sub>second reading</sub>	BI-RADS <sub>initial</sub>	BI-RADS <sub>adjusted</sub>	BI-RADS <sub>second reading</sub>
<b>Total</b>									
Junior	82.5	94.7	92.1	79.2	79.6	85.1	0.879 (0.852, 0.902)	0.901 (0.876, 0.922)	0.921 (0.898, 0.940)
P	<0.001	0.03	<0.001	0.002	<0.001	<0.001	<0.001	<0.001	<0.001
Senior	91.2	96.5	95.6	88.9	80.5	89.8	0.950 (0.931, 0.965)	0.946 (0.926, 0.962)	0.951 (0.932, 0.966)
P	<0.001	0.50	0.01	<0.001	<0.001	0.39	0.059	0.40	0.91
<b>D ≤2 cm</b>									
Junior	68.6	88.2	86.3	83.7	81.2	88.9	0.864 (0.831, 0.893)	0.895 (0.865, 0.920)	0.924 (0.897, 0.945)
P	<0.001	0.50	<0.001	0.002	<0.001	<0.001	<0.001	<0.001	<0.001
Senior	82.4	92.2	92.2	92.1	84.7	92.1	0.935 (0.909, 0.954)	0.935 (0.910, 0.955)	0.936 (0.911, 0.956)
P	0.002	>0.99	0.002	<0.001	<0.001	>0.99	0.90	0.92	0.90
<b>D &gt;2 cm</b>									
Junior	93.7	100.0	96.8	39.1	39.1	52.2	0.744 (0.672, 0.808)	0.774 (0.704, 0.834)	0.794 (0.726, 0.852)
P	–	–	0.13	>0.99	0.03	0.03	0.04	0.14	0.003
Senior	98.4	100.0	98.4	60.9	43.5	69.6	0.930 (0.881, 0.963)	0.910 (0.857, 0.948)	0.906 (0.852, 0.945)
P	–	–	>0.99	0.008	<0.001	0.13	0.07	0.83	0.14

BI-RADS assessment category 4B, 4C, and 5 were considered positive for malignant for calculation of sensitivity and specificity. P values in the BI-RADS<sub>initial</sub> column were calculated by comparing with BI-RADS<sub>adjusted</sub>; P values in the BI-RADS<sub>adjusted</sub> column were calculated by comparing with BI-RADS<sub>second reading</sub>; P values in the BI-RADS<sub>second reading</sub> column were calculated by comparing with BI-RADS<sub>initial</sub>. AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System; AUC, area under the curve; CI, confidence interval; BI-RADS<sub>initial</sub>, radiologists analyze the images without the assistance of AI; BI-RADS<sub>adjusted</sub>, incorporate BI-RADS provided by AI into BI-RADS<sub>initial</sub>; BI-RADS<sub>second reading</sub>, combining ultrasound features extracted by AI to conduct a second reading; D, largest diameter of breast lesions.

threshold, the diagnostic sensitivity and specificity of senior and junior radiologists for all breast lesions were 91.2% *vs.* 82.5% and 88.9% *vs.* 79.2%, respectively. For lesions ≤2 cm, the sensitivity and specificity were 82.4% *vs.* 68.6% and 92.1% *vs.* 83.7%, respectively, whereas for lesions >2 cm, they were 98.4% *vs.* 93.7% and 60.9% *vs.* 39.1%, respectively (Table 2).

### Diagnostic performance of the AI system alone

Using the AI system alone, the AUC was 0.924 for all lesions, 0.917 for lesions ≤2 cm, and 0.821 for lesions size >2 cm (Table 3).

When BI-RADS category 3 was used as the threshold, the sensitivities of the AI system for all lesions, lesions

≤2 cm, and lesions >2 cm were all 100.0%, whereas the specificities were 44.6%, 48.6%, and 8.7%, respectively. When BI-RADS category 4A was used as the threshold, the sensitivities were 92.1%, 86.3%, and 96.8%, whereas the specificities were 79.6%, 83.7%, and 43.5%, respectively (Table 3).

### Diagnostic performance with assistance of AI system

For all breast lesions, BI-RADS<sub>adjusted</sub> improved the AUC of junior radiologists to 0.901 [95% confidence interval (CI): 0.876, 0.922] and BI-RADS<sub>second reading</sub> improved it to 0.921 (95% CI: 0.898, 0.940). For breast lesions ≤2 cm, BI-RADS<sub>adjusted</sub> improved the AUC to 0.895 (95% CI: 0.865, 0.920) and BI-RADS<sub>second reading</sub> improved it to 0.924 (95%



**Table 3** Diagnostic performance of AI system alone

Threshold	Sensitivity (%)	Specificity (%)	+LR	–LR	Accuracy	AUC (95% CI)
Total						0.924 (0.901, 0.943)
3	100.0	44.6	1.80	0	0.632	
4A	92.1	79.6	4.52	0.099	0.838	
D ≤2 cm						0.917 (0.890, 0.940)
3	100.0	48.6	1.95	0	0.590	
4A	86.3	83.7	5.29	0.16	0.842	
D >2 cm						0.821 (0.755, 0.875)
3	100.0	8.7	1.10	0	0.756	
4A	96.8	43.5	1.71	0.073	0.826	

AI, artificial intelligence; +LR, positive likelihood ratio; –LR, negative likelihood ratio; AUC, area under the curve; CI, confidence interval; D, largest diameter of breast lesions.

CI: 0.897, 0.945). For breast lesions >2 cm, BI-RADS<sub>adjusted</sub> improved the AUC to 0.774 (95% CI: 0.704, 0.834) and BI-RADS<sub>second reading</sub> improved it to 0.794 (95% CI: 0.726, 0.852). And they were all statistically different compared with the junior radiologists alone (all  $P < 0.05$ ). Furthermore, the diagnostic performance of BI-RADS<sub>second reading</sub> is always better than that of BI-RADS<sub>adjusted</sub>. For the senior radiologists, neither BI-RADS<sub>adjusted</sub> nor BI-RADS<sub>second reading</sub> improved AUC ( $P > 0.05$ ) (Figure 2).

When BI-RADS category 3 was used as the threshold, the AI system alone was also not effective in improving the sensitivity and specificity of the radiologists (Table 3 and Table S1). When BI-RADS category 4A was used as the threshold, for the junior radiologists, the specificity of BI-RADS<sub>second reading</sub> was significantly higher than that of BI-RADS<sub>adjusted</sub> for all breast lesions, lesions ≤2 cm, and lesions >2 cm (85.1% *vs.* 79.6%,  $P < 0.001$ ; 88.9% *vs.* 81.2%,  $P < 0.001$ ; and 52.2% *vs.* 39.1%,  $P = 0.03$ , respectively). For the junior radiologists, BI-RADS<sub>second reading</sub> improved its sensitivity for all lesions and lesions ≤2 cm (92.1% *vs.* 82.5%,  $P < 0.001$ , and 86.3% *vs.* 68.6%,  $P < 0.001$ , respectively). While the specificity was all improved for all lesions, lesions ≤2 cm, and lesions >2 cm (85.1% *vs.* 79.2%,  $P < 0.001$ ; 88.9% *vs.* 83.7%,  $P < 0.001$ ; and 52.2% *vs.* 39.1%,  $P = 0.03$ , respectively). For the senior radiologists, BI-RADS<sub>second reading</sub> improved their sensitivity for all lesions and lesions ≤2 cm (95.6% *vs.* 91.2%,  $P = 0.01$ , and 92.2% *vs.* 82.4%,  $P = 0.002$ , respectively) (Table 2).

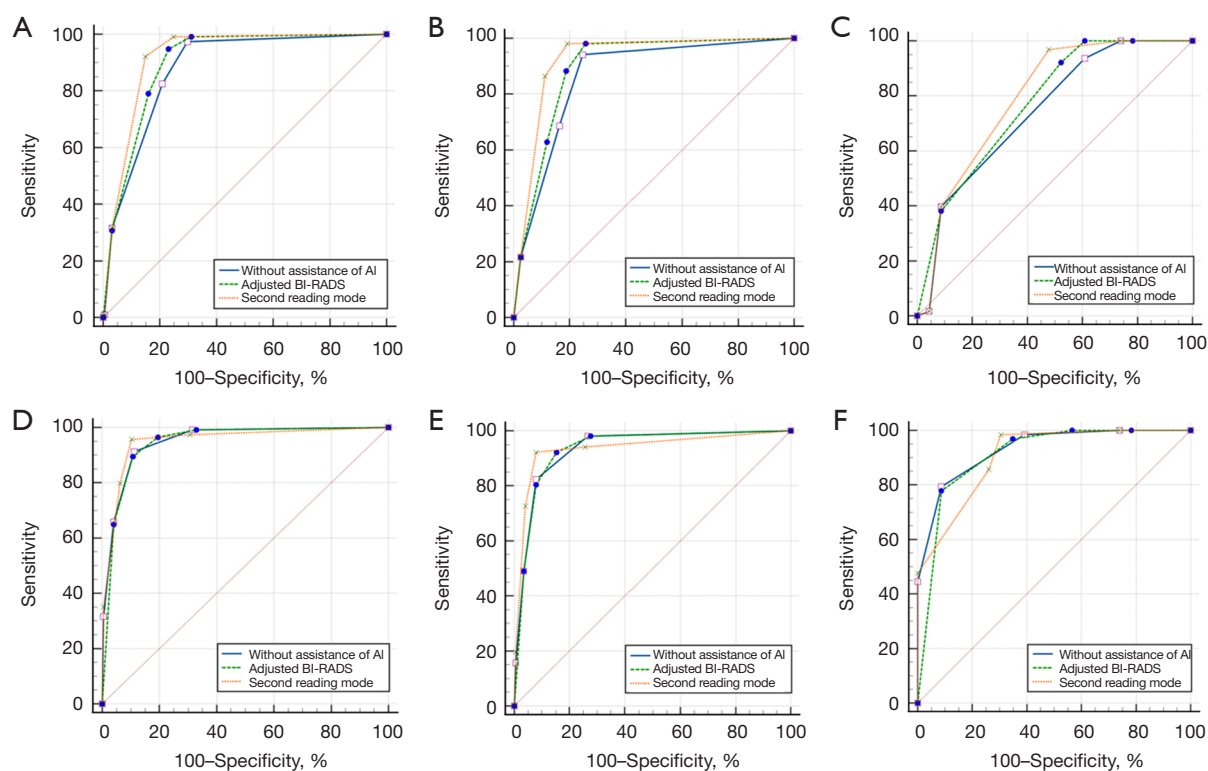
For junior radiologists, the AUC of AI system alone was significantly higher than that of BI-RADS<sub>adjusted</sub> (0.924 *vs.* 0.901,  $P = 0.005$ ), but there was no statistical difference

compared with that of BI-RADS<sub>second reading</sub> (0.924 *vs.* 0.921,  $P = 0.72$ ). In addition, the specificity of BI-RADS<sub>second reading</sub> was higher than that of AI alone for all lesions and lesions ≤2 cm (85.1% *vs.* 79.6%,  $P < 0.001$ , and 88.9% *vs.* 83.7%,  $P < 0.001$ , respectively). For the senior radiologists, the AUC of AI alone was significantly lower than those of BI-RADS<sub>adjusted</sub> and BI-RADS<sub>second reading</sub> (0.924 *vs.* 0.946,  $P = 0.003$ , and 0.924 *vs.* 0.951,  $P < 0.001$ , respectively) (Tables 2, 3).

### Change of management decision

When BI-RADS category 3 was used as the threshold, the junior radiologists correctly changed the management decision from follow-up to biopsy for four malignant lesions and from biopsy to follow-up for four benign lesions using BI-RADS<sub>adjusted</sub>. However, incorrect change from follow-up to biopsy was observed for 10 benign lesions. Using BI-RADS<sub>second reading</sub>, four malignant lesions were accurately changed from follow-up to biopsy and 12 benign lesions were changed from biopsy to follow-up. Compared with junior radiologists alone, BI-RADS<sub>second reading</sub> successfully reduced the rate of unnecessary biopsy by 4.50% in all breast lesions and 7.20% in lesions ≤2 cm, no statistical difference was observed ( $P = 0.39$  and  $P = 0.41$ ) (Table S2).

When BI-RADS category 4A was used as the threshold, the junior radiologists correctly changed the management decision from follow-up to biopsy for 28 malignant lesions using BI-RADS<sub>adjusted</sub>. However, using BI-RADS<sub>second reading</sub>, 22 malignant lesions were correctly changed from follow-up to biopsy and 12 benign lesions from biopsy



**Figure 2** ROC curves of three reading modes. (A) ROC curves of the junior radiologists for all breast lesions. (B) ROC curves of the junior radiologists for breast lesions size  $\leq 2$  cm. (C) ROC curves of the junior radiologists for breast lesions size  $> 2$  cm. (D) ROC curves of the senior radiologist for all breast lesions. (E) ROC curves of the senior radiologist for breast lesions size  $\leq 2$  cm. (F) ROC curves of the senior radiologist for breast lesions size  $> 2$  cm. AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System; ROC, receiver operating characteristic.

to follow-up. Compared with junior radiologists alone, BI-RADS<sub>second reading</sub> reduced the rate of unnecessary biopsies by 9.15% in all lesions ( $P=0.17$ ), by 14.70% in lesions  $\leq 2$  cm ( $P=0.01$ ), and by 3.90% in lesions  $> 2$  cm ( $P=0.38$ ) (Table 4).

AI was not as helpful for senior radiologists as it for junior radiologists in changing management decisions. Furthermore, no statistically significant difference was observed in the change of the rate of unnecessary biopsies, regardless of the threshold used.

## Discussion

The experience of radiologists has a certain impact on diagnostic accuracy. The rapid developments in AI technology are expected to bridge the difference in the experience of radiologists. Therefore, it is necessary to analyze the diagnostic performance of AI in clinical setting. The present study investigated the performance and

optimal use of AI for breast ultrasound to diagnose solid and cystic-solid lesions. Furthermore, it compared the different seniorities and ways of applying AI using histology or 1-year follow-up as the standard of reference. The results indicated that AI is helpful for breast ultrasound, particularly for junior radiologists. The diagnostic efficiency of AI for lesions  $\leq 2$  cm was higher than that of lesions  $> 2$  cm. When the threshold was set to BI-RADS category 4A and BIRADS<sub>second reading</sub> was applied, AI can exhibit excellent performance.

This study demonstrates that AI-assisted ultrasound diagnostic systems are extremely helpful for radiologists, particularly junior radiologists, in differentiating the benign and malignant breast lesions. A previous study has reported that AI systems can improve the diagnostic performance of radiologists by increasing their specificity, accuracy, and positive predictive value (13). In this study, two combined diagnosis modes were used individually to improve the diagnostic performance of radiologists.

**Table 4** Management decision changes of radiologists with assistance of AI system using BI-RADS category 4A as threshold

Parameters	BI-RADS <sub>adjusted</sub>			BI-RADS <sub>second reading</sub>		
	No change	Downgraded	Upgraded	No change	Downgraded	Upgraded
Malignant (n=228)						
Total						
Junior	200	0	28	206	0	22
Senior	216	0	12	214	2	12
D ≤2 cm						
Junior	82	0	20	84	0	18
Senior	92	0	10	92	0	10
D >2 cm						
Junior	118	0	8	122	0	4
Senior	124	0	2	122	2	2
Benign (n=451)						
Total						
Junior	441	0	10	439	12	0
Senior	413	0	38	439	8	4
D ≤2 cm						
Junior	395	0	10	393	12	0
Senior	375	0	30	397	4	4
D >2 cm						
Junior	46	0	0	46	0	0
Senior	38	0	8	42	4	0

BI-RADS assessment category 4B, 4C, and 5 were considered positive for malignant for calculation of management decision changes. Downgrade means moving the recommendation from biopsy to follow-up and upgrade means moving from follow-up to biopsy. AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System; BI-RADS<sub>adjusted</sub>, incorporate BI-RADS provided by AI into BI-RADS<sub>initial</sub>; BI-RADS<sub>initial</sub>, radiologists analyze the images without the assistance of AI; BI-RADS<sub>second reading</sub>, combining ultrasound features extracted by AI to conduct a second reading; D, largest diameter of breast lesions.

The results indicated that compared with BI-RADS<sub>adjusted</sub>, BI-RADS<sub>second reading</sub> was more effective in improving the AUC of junior radiologists while increasing their sensitivity and specificity. In a previous study, junior radiologists subjectively made diagnostic decisions based on the region of interest provided by the AI system, and the AUC improved from 0.73 to 0.80 (14). This indicated that the application of BI-RADS<sub>second reading</sub> was more suitable.

The worsening effect of breast biopsy on psychological distress is well known. As percutaneous core needle biopsies are expensive, reducing the rate of nonessential puncture biopsies can reduce the economic burden of breast cancer diagnosis on the healthcare system (15). The present study

found that when changing management decisions using the AI system, the rate of unnecessary biopsy of breast lesions could be significantly reduced. In the second reading mode of the junior radiologists, 12 benign lesions were correctly changed from biopsy to follow-up in all breast lesions, and the rate of unnecessary biopsy was reduced by 9.15%. Previous research has also proved that if residents were aware of the diagnosis results of S-Detect and made the diagnosis of breast lesions according to their own clinical experience, the rate of unnecessary biopsy could be reduced from 44.4% to 32.7% (P=0.03) (16).

This study aimed to determine whether implementing an AI-assisted ultrasound diagnostic system enhances the



diagnostic performance of radiologists and to investigate how the full potential of AI can be realized to maximize its value. Therefore, we also analyzed the difference between the application of the AI system to breast lesions  $>2$  cm and lesions  $\leq 2$  cm. Compared with lesions  $>2$  cm, BI-RADS<sub>second reading</sub> exhibited lower sensitivity and higher specificity for lesions  $\leq 2$  cm, but its overall diagnostic performance was better (AUC: 0.924 *vs.* 0.794). Furthermore, although BI-RADS<sub>second reading</sub> successfully reduced the rate of unnecessary biopsies in both the aforementioned lesions, statistical difference was observed only for lesions  $\leq 2$  cm. This could be attributed to the fact that when examining larger breast lesions, the image of the lesion displayed in real-time in the AI-assisted ultrasound diagnostic system exceeds the range of the display and interferes with the AI's analysis of the lesion. Similarly, some studies have reported that AI-assisted ultrasound diagnosis has a more important application value in breast lesions  $\leq 2$  cm (17,18). The second reading mode has limited value for the additional diagnosis of lesions  $>2$  cm, where delineating morphological features of lesions becomes significantly difficult for AI. In the future, for lesions  $>2$  cm, larger cohort data will be needed to optimize the AI diagnostic model so as to meet the diagnostic needs of larger lesions. The present study found that AI-assisted ultrasound diagnostic system can play a better role with BI-RADS category 4A as the threshold. A previous study has demonstrated that when the threshold is set at 4A, S-Detect significantly outperforms the radiologists in terms of specificity, positive predictive value, and accuracy, further corroborating our conclusions (19). In addition, some studies used 4A as the threshold when comparing the diagnostic effectiveness between conventional ultrasound and AI systems (20). The AI system was found to provide a better value for lesions  $\leq 2$  cm and was most appropriate when BI-RADS 4A was used as the threshold.

In the present study, the effect of AI system on senior radiologists was not obvious, and when junior radiologists were using AI system, the diagnostic efficiency remainder was lower than that of senior radiologists. Previous research on AI-assisted diagnostic ultrasound systems has reported that combined diagnosis with AI improves the diagnostic efficiency of radiologists regardless of their experience (21). Other studies have found that AI technology can improve the diagnostic efficiency of doctors with less experience in ultrasound, reaching a level similar to that of doctors with more experience (22,23). Some studies have also reported that diagnosis with AI assistance is not useful for senior

radiologists (24). In this study, the senior radiologists were breast specialists with experience of 15 years; thus, a ceiling effect may exist. Nevertheless, the AI should not displace the well-informed clinical judgment of physicians but should instead be used as a complementary tool for treatment plan decision-making, particularly for junior radiologists.

There are limitations in this study. First, the sample size of breast lesions  $>2$  cm was relatively small. Second, to avoid additional healthcare costs and patient anxiety, the study defined BI-RADS 3 and most BI-RADS 4A lesions that remained stable in size and morphology over 1 year of ultrasound follow-up as benign, which may introduce some bias. Third, the reliance of the study on a single AI system limits the generalizability of the results to other AI platforms. Further research is warranted to determine whether the results are consistent across different platforms. Finally, this study did not include lesions of BI-RADS category 2, indicating certain selection bias. However, according to ACR BI-RADS-US 2013, BI-RADS 2 lesions are considered to be benign with essentially 0% likelihood of malignancy and are generally not misdiagnosed. The application of AI to such lesions cannot improve the accuracy of diagnosis, but may increase the time cost; thus, BI-RADS 2 lesions are not the best indication for AI.

## Conclusions

In the setting of this study, AI-assisted ultrasound is a valuable tool for diagnosing solid and cystic-solid breast lesions, particularly for junior radiologists and lesions  $\leq 2$  cm. Moreover, the utilization of the second reading mode (combining characteristic information of lesions extracted by AI to conduct a second reading so as to obtain a new BI-RADS) can achieve excellent diagnostic performance.

## Acknowledgments

*Funding:* This work was supported by the National Natural Science Foundation of China (No. 82071925), the Military Health Project (No. 22BJZ23), and the Equipment Comprehensive Research Project (No. LB20211A010011).

## Footnote

*Reporting Checklist:* The authors have completed the STARD reporting checklist. Available at <https://gs.amegroups.com/>

[article/view/10.21037/ggs-24-213/rc](https://www.amegroups.com/article/view/10.21037/ggs-24-213/rc)

*Data Sharing Statement:* Available at <https://gs.amegroups.com/article/view/10.21037/ggs-24-213/dss>

*Peer Review File:* Available at <https://gs.amegroups.com/article/view/10.21037/ggs-24-213/prf>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://gs.amegroups.com/article/view/10.21037/ggs-24-213/coif>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This prospective study was approved by the Ethics Committee of the Chinese PLA General Hospital (No. S2021-683-01), and informed consent was obtained from all patients. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
2. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. *CA Cancer J Clin* 2022;72:7-33.
3. Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. *CA Cancer J Clin* 2023;73:17-48.
4. Pirikahu S, Lund H, Cadby G, et al. The impact of breast density notification on rescreening rates within a population-based mammographic screening program. *Breast Cancer Res* 2022;24:5.
5. Gu Y, Tian JW, Ran HT, et al. The Utility of the Fifth Edition of the BI-RADS Ultrasound Lexicon in Category 4 Breast Lesions: A Prospective Multicenter Study in China. *Acad Radiol* 2022;29 Suppl 1:S26-34.
6. Alshoabi SA, Alareqi AA, Alhazmi FH, et al. Utility of Ultrasound Imaging Features in Diagnosis of Breast Cancer. *Cureus* 2023;15:e37691.
7. Shen YT, Chen L, Yue WW, et al. Artificial intelligence in ultrasound. *Eur J Radiol* 2021;139:109717.
8. Wang J, Jiang J, Zhang D, et al. An integrated AI model to improve diagnostic accuracy of ultrasound and output known risk features in suspicious thyroid nodules. *Eur Radiol* 2022;32:2120-9.
9. Xue P, Si M, Qin D, et al. Unassisted Clinicians Versus Deep Learning-Assisted Clinicians in Image-Based Cancer Diagnostics: Systematic Review With Meta-analysis. *J Med Internet Res* 2023;25:e43832.
10. Zhao C, Xiao M, Ma L, et al. Enhancing Performance of Breast Ultrasound in Opportunistic Screening Women by a Deep Learning-Based System: A Multicenter Prospective Study. *Front Oncol* 2022;12:804632.
11. Xia Q, Cheng Y, Hu J, et al. Differential diagnosis of breast cancer assisted by S-Detect artificial intelligence system. *Math Biosci Eng* 2021;18:3680-9.
12. Wu JY, Zhao ZZ, Zhang WY, et al. Computer-Aided Diagnosis of Solid Breast Lesions With Ultrasound: Factors Associated With False-negative and False-positive Results. *J Ultrasound Med* 2019;38:3193-202.
13. 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;20:749-58.
14. Choi JH, Kang BJ, Baek JE, et al. Application of computer-aided diagnosis in breast ultrasound interpretation: improvements in diagnostic performance according to reader experience. *Ultrasonography* 2018;37:217-25.
15. Pfof A, Sidey-Gibbons C, Barr RG, et al. Intelligent multi-modal shear wave elastography to reduce unnecessary biopsies in breast cancer diagnosis (INSPIRED 002): a retrospective, international, multicentre analysis. *Eur J Cancer* 2022;177:1-14.
16. Wang XY, Cui LG, Feng J, et al. Artificial intelligence for breast ultrasound: An adjunct tool to reduce excessive lesion biopsy. *Eur J Radiol* 2021;138:109624.
17. Yongping L, Zhou P, Juan Z, et al. Performance of Computer-Aided Diagnosis in Ultrasonography for Detection of Breast Lesions Less and More Than 2 cm: Prospective Comparative Study. *JMIR Med Inform*

- 2020;8:e16334.
18. Zhang D, Jiang F, Yin R, et al. A Review of the Role of the S-Detect Computer-Aided Diagnostic Ultrasound System in the Evaluation of Benign and Malignant Breast and Thyroid Masses. *Med Sci Monit* 2021;27:e931957.
  19. Kim K, Song MK, Kim EK, et al. 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.
  20. Xing B, Chen X, Wang Y, et al. Evaluating breast ultrasound S-detect image analysis for small focal breast lesions. *Front Oncol* 2022;12:1030624.
  21. Li J, Sang T, Yu WH, et al. The value of S-Detect for the differential diagnosis of breast masses on ultrasound: a systematic review and pooled meta-analysis. *Med Ultrason* 2020;22:211-9.
  22. Zhao C, Xiao M, Jiang Y, et al. Feasibility of computer-assisted diagnosis for breast ultrasound: the results of the diagnostic performance of S-detect from a single center in China. *Cancer Manag Res* 2019;11:921-30.
  23. 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;41:718-27.
  24. 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;41:1355-63.

**Cite this article as:** Jin ZY, Li JK, Niu RL, Fu NQ, Jiang Y, Li SY, Wang ZL. Application of an artificial intelligence-assisted diagnostic system for breast ultrasound: a prospective study. *Gland Surg* 2024;13(12):2221-2231. doi: 10.21037/gs-24-213