Egyptian Journal of Radiology and Nuclear Medicine

RESEARCH Open Access

Potential role of AI decision support in reducing unnecessary ultrasound-guided breast biopsies

Meliha Akin[1*], Servet Erdemli[1], Onur Dulgeroglu[2], Fatma Tokat[3] and Fusun Taskin[1]

Abstract

Background Ultrasound (US)-guided breast biopsy is widely used for evaluating suspicious breast lesions; yet, many biopsied lesions prove to be benign, resulting in potentially unnecessary invasive procedures. Improving diagnostic specificity without compromising sensitivity remains a key challenge in breast imaging, and artificial intelligence-based decision support systems have emerged as promising tools to assist radiologists in biopsy decision-making. This study aimed to compare the diagnostic effectiveness of an AI-DS system with that of experienced breast radiologists in evaluating lesions that underwent US-guided biopsy, and to assess the potential of AI-DS in reducing unnecessary biopsies.

Methods BI-RADS scores assigned by radiologists and the AI-DS were retrospectively analyzed in 453 breast lesions biopsied between October 2021 and March 2024. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and area under the curve (AUC) were calculated for reader A, reader B, and the AI-DS. Diagnostic performance was compared using ROC analysis, McNemar tests, and generalized estimating equations.

Results The AI-DS demonstrated significantly higher specificity, PPV, and accuracy than both radiologists (all $p < 0.001$). The AUC was 0.681 (95% CI = 0.633–0.729) for the AI-DS, 0.556 (95% CI = 0.502–0.610) for reader A, and 0.574 (95% CI = 0.521–0.627) for reader B. Subgroup analyses showed that the AI-DS yielded more accurate results than radiologists across most subcategories. In a hypothetical scenario, if 84 BI-RADS 4A lesions downgraded by AI-DS had not been biopsied, the biopsy rate could have been reduced by 18.6% (84/453) while missing only one malignant lesion.

Conclusion The AI-DS demonstrated sensitivity comparable to radiologists while providing higher specificity and accuracy. In cases where radiologists classify lesions as BI-RADS 4A but the AI-DS predicts BI-RADS 2 or 3, the AI output could prompt reconsideration of biopsy necessity and support short-term imaging follow-up instead.

Keywords Artificial intelligence, Breast neoplasms, Image-guided biopsy, Ultrasonography

Background

Mammography screening reduces breast cancer-related mortality [1, 2]. Increased breast density is associated with both a reduction in mammographic sensitivity and a higher risk of breast cancer and mortality [3, 4]. Therefore, complementary imaging modalities are recommended for women with dense breast tissue.

Ultrasonography (US) is a crucial complementary imaging technique for breast cancer diagnosis. When

*Correspondence:
Meliha Akin
meloakin@gmail.com
[1] Department of Radiology, Acibadem Mehmet Ali Aydinlar University Research Institute of Senology, Istanbul, Turkey
[2] Department of General Surgery, Acibadem Mehmet Ali Aydinlar University Research Institute of Senology, Istanbul, Turkey
[3] Department of Pathology, Acibadem Mehmet Ali Aydinlar University Research Institute of Senology, Istnabul, Turkey

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

used in addition to mammographic screening, US improves the detection rate of early-stage cancers [5, 6]. It is a real-time and cost-effective imaging method. However, its major limitations include operator and equipment dependency [7] as well as low specificity [5, 6]. Moreover, the positive predictive value (PPV) of US is low and ranges from 6.5 to 25.4% [8, 9]. Consequently, US may lead to unnecessary follow-ups, additional imaging, and biopsies [7, 10, 11].

Artificial intelligence decision support (AI-DS) algorithms have been developed to address these limitations. Studies indicate that AI-DS systems can enhance radiologists' diagnostic performance in classifying breast lesions on US [12, 13]. Additionally, AI-DS systems have been shown to increase specificity and reduce unnecessary biopsies [12–15].

However, most AI-DS systems rely on "black box" models that generate predictions without revealing the underlying reasoning or the features influencing those decisions [13, 16–18]. This lack of transparency limits clinical interpretability, as understanding how AI systems generate their outputs is important for radiologist confidence, appropriate validation, and safe clinical implementation. Therefore, the validation and auditing of a model by experienced radiologists are expected to enhance the reliability of AI-DS systems.

The present study aimed to compare the diagnostic performance of AI-DS and dedicated breast radiologists in evaluating lesions that underwent US-guided biopsy. We also sought to assess the potential of AI-DS to reduce unnecessary biopsies. Furthermore, to explore the comparative advantages of AI-DS and radiologists, additional subgroup analyses were performed based on age, breast composition, lesion depth, size, and imaging characteristics.

Methods

Study design and participants

This single-center retrospective study was approved by our institutional review board, which waived the requirement for informed consent. We included patients who underwent US-guided biopsy between October 2021 and March 2024 and had lesions clearly visible on US. Lesions that lacked clear visualization in two orthogonal planes or showed radiologic–pathologic discordance without subsequent surgical excision were excluded. A flowchart illustrating the inclusion and exclusion process is presented in Fig. 1.

The study was designed to compare three diagnostic approaches:

(1) Reader A (the initial clinical radiologists): one of two breast radiologists with 5 and 25 years of experience who assigned BI-RADS during US-guided biopsy with full access to clinical history, prior imaging, and additional ultrasound techniques.
(2) AI-DS (Koios DS™ Breast): evaluated lesions using two orthogonal pre-biopsy static images, without any clinical or imaging context.
(3) Reader B: a radiologist with 5 years of experience who retrospectively interpreted only the pre-biopsy US images in a laboratory setting, without access to clinical, radiologic, or histopathologic information—mimicking the AI-DS evaluation environment.

Table 1 outlines the differences in information access and working conditions between reader A and reader B.

Subgroup analyses were performed according to breast composition, lesion depth, size, and imaging characteristics based on reader A's biopsy reports and lesion images. Histopathology from core needle biopsy or surgical excision served as the reference standard.

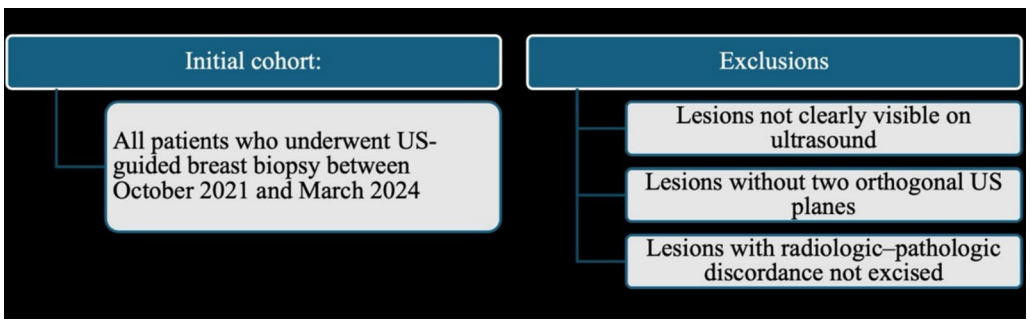


Fig. 1 Flowchart illustrating the lesion selection process, including the initial cohort and applied exclusion criteria

Table 1 Summary of reader A and reader B conditions and data access

Parameter	Reader A	Reader B
Setting	Clinical environment (during US-guided biopsy)	Laboratory (retrospective) environment
Experience	One of two radiologists (5 and 25 years of experience in breast imaging)	Radiologist with 5 years of experience in breast imaging
Access to information	Full access to patient history, prior imaging, and additional US techniques	No access to clinical information, prior imaging, or additional modalities
Data used	Real-time and stored ultrasound images obtained during the biopsy procedure	Two orthogonal static ultrasound images used to simulate AI-DS evaluation
Purpose	Assigned BI-RADS categories during the biopsy procedure under real-world clinical conditions	Simulated AI-DS evaluation without contextual or clinical data

US imaging and biopsy procedure

US examinations were performed using a GE LOGIQ E10S Plus US system (GE Healthcare) equipped with a high-frequency linear probe (4–20 MHz). Biopsies were conducted with an automatic biopsy gun (Becton Dickinson) using a 14-gauge needle.

BI-RADS 3 lesions were biopsied in high-risk women, those newly diagnosed with breast cancer, patients with clinical symptoms, or when physical examination suggested suspicious findings.

US image analyses by radiologists and AI-DS

Breast composition, lesion size, and depth were measured on stored PACS images.

Reader A's BI-RADS scores were extracted from biopsy reports. Reader B independently assigned BI-RADS scores using only pre-biopsy US images. Both radiologists classified lesions as BI-RADS 3, 4A, 4B, 4C, or 5, without knowledge of AI-DS results. For AI-DS evaluation, a radiologist manually defined ROIs on two orthogonal static images. The AI-DS automatically categorized lesions into BI-RADS 2 (benign), BI-RADS 3 (probably benign), BI-RADS 4A–4B (suspicious), or BI-RADS 4C+ (probably malignant). Due to software-related export limitations, Koios-annotated images were not systematically archived.

Statistical analysis

Because the AI-DS categories differed from the radiologist-assigned BI-RADS classifications, all final assessments were dichotomized as negative (BI-RADS 2–3) and positive (BI-RADS 4–5). According to the ACR BI-RADS Atlas, lesions with an estimated malignancy likelihood greater than 2% fall into BI-RADS 4 or higher and warrant tissue diagnosis [19]; therefore, BI-RADS 4A lesions—despite their relatively low malignancy probability (2–10%)—were included in the positive group.

For subgroup analyses based on ultrasonographic features, diagnostic performance was compared between negative and positive categories, whereas BI-RADS-based subgroup analyses were conducted by dividing lesions into three subgroups: BI-RADS 2–3, BI-RADS 4A–4B, and BI-RADS 4C–5.

Diagnostic performance was assessed using sensitivity, specificity, PPV, NPV, accuracy, and ROC AUC. Comparisons between the radiologists and the AI-DS were performed using McNemar's test and generalized estimating equations. Statistical significance was defined as $p < 0.05$ (two-tailed) with 95% confidence intervals. Agreement between the radiologists and the AI-DS was evaluated using Cohen's kappa (κ), interpreted as slight (0–0.2), fair (0.2–0.4), moderate (0.4–0.6), substantial (0.6–0.8), or almost perfect (> 0.8). All analyses were carried out using SPSS version 26 (IBM Corp).

Results

Patient population and lesion characteristics

A total of 453 lesions (395 masses and 58 non-mass lesions) were evaluated. Among these lesions, 297/453 (65.5%) were benign, and 156/453 (34.4%) were malignant. The mean patient age was 45.0 years ± 12.0 (range: 15–87 years). The mean lesion size was 19.5 ± 14.8 mm (range: 3–110 mm). Benign lesions had a mean size of 17.4 ± 14.4 mm (range: 3–110 mm), while malignant lesions had a mean size of 23.4 ± 14.7 mm (range: 4–80 mm). Thirty-two out of the 453 lesions were excised due to either high-risk pathology or radiologic–pathologic discordance. The mean follow-up duration after the final diagnosis was 6 months (range: 0–36 months).

The most common benign diagnoses were fibroadenoma (44.1%, 131/297) and adenosis/stromal fibrosis (33.7%, 100/297). The most common malignant diagnoses were invasive breast carcinoma of no special type (IBC-NST) (86.5%, 135/156), invasive lobular carcinoma (5.8%, 9/156), and ductal carcinoma in situ (5.1%, 8/156). Detailed histopathological results are provided in Table 2.

Table 2 Histopathological diagnosis of breast lesions

Histopathological diagnosis	n (%)
Benign lesions (n = 297)	
Fibroadenoma	44.1%[131/297]
Adenosis and stromal fibrosis	33.7%[100/297]
Papilloma	7.1%[21/297]
Proliferative lesions without atypia	4.4%[13/297]
Mastitis	3.0%[9/297]
Phyllodes tumor	1.7%[5/297]
Complex / Cellular fibroadenoma	1.3%[4/297]
Radial scar	1.0%[3/297]
Flat epithelial atypia	1.0%[3/297]
Lobular carcinoma in situ	0.7%[2/297]
Lactation adenoma	0.7%[2/297]
Atypical ductal hyperplasia	0.3%[1/297]
Atypical lobular hyperplasia	0.3%[1/297]
Fat necrosis	0.3%[1/297]
Intramammary lymph node	0.3%[1/297]
Malign lesions (n = 156)	
Invasive breast cancer of no special type	86.5%[135/156]
Invasive lobular carcinoma	5.8%[9/156]
Ductal carcinoma in situ	5.1%[8/156]
Mucinous carcinoma	0.6%[1/156]
B-cell lymphoblastic lymphoma	1.3%[2/156]
Malignant melanoma	0.6%[1/156]

Inter-reader agreement in lesion differential diagnosis

There was substantial agreement between the readers regarding dichotomized BI-RADS classifications (κ = 0.615, 95% CI = 0.570–0.659). However, there was only fair agreement between reader A and the AI-DS (κ = 0.214, 95% CI = 0.176–0.252) and likewise between reader B and the AI-DS (κ = 0.274, 95% CI = 0.233–0.315).

Diagnostic performances of the AI-DS and radiologists

The AI-DS had significantly higher specificity, PPV, and accuracy than the radiologists ($p < 0.001$ for all results). However, no significant differences were observed in sensitivity or NPV between the AI-DS and the radiologists (Table 3).

The AI-DS showed a higher AUC (AUC = 0.681, 95% CI = 0.633–0.729) than the radiologists (reader A: AUC = 0.556, 95% CI = 0.502–0.610; reader B: AUC = 0.574, 95% CI = 0.521–0.627) (Fig. 2).

Subgroup analysis

In the subgroup analysis, the AI-DS demonstrated significantly higher accuracy than both reader A and reader B for lesions with a depth of less than 10 mm ($p < 0.001$ for all results). For lesions measuring 10 mm or more, its performance was comparable to that of reader A ($p > 0.001$) and remained superior to that of reader B ($p < 0.001$).

When evaluating lesions based on surrounding breast tissue composition, the AI-DS outperformed both readers for lesions surrounded by heterogeneous or homogeneous fibroglandular tissue ($p < 0.001$ for all results), while achieving similar accuracy to both for lesions surrounded by homogeneous fat tissue ($p > 0.001$ for all results).

Regarding lesion type, the AI-DS showed higher accuracy than both reader A and reader B for mass lesions ($p < 0.001$ for all results), whereas its accuracy for non-mass lesions was comparable to that of both readers ($p > 0.001$ for all results).

For mass lesions, the AI-DS achieved higher accuracy for lesions with parallel orientation, oval shape, circumscribed margins, and absence of posterior features ($p < 0.001$ for all results). Additionally, it outperformed reader B for lesions with a round shape ($p < 0.001$). However, for lesions with non-parallel orientation, irregular shape, non-circumscribed margin, or posterior enhancement, the AI-DS performed comparably to both readers ($p > 0.001$ for all results).

The diagnostic performance of the AI-DS system in comparison with reader A and reader B, based on lesion characteristics, is presented in detail in Tables 4 and 5.

In subgroup analyses based on BI-RADS classification, the AI-DS demonstrated higher accuracy than both readers for BI-RADS 4A–4B categories but

Table 3 Comparison of diagnostic performance between the readers and the AI-DS System

Variable	AI-DS system	Reader A	*p*-value	AI-DS system	Reader B	*p*-value
Sensitivity	96.2%[150/156]	98.7%[154/156]	0.289[a]	96.2%[150/156]	99.4%[155/156]	0.063[a]
Specificity	40.1%[119/297]	12.5%[37/260]	**< 0.001[a]***	40.1%[119/297]	15.5%[46/297]	**< 0.001[a]***
PPV	45.7%[150/328]	37.2%[154/414]	**0.019[b]***	45.7%[150/328]	38.2%[155/406]	**0.039[b]***
NPV	95.2%[119/125]	94.9%[37/39]	0.934[b]	95.2%[119/125]	97.9%[46/47]	0.442[b]
Accuracy	59.4%[269/453]	42.2%[191/453]	**< 0.001[a]***	59.4%[269/453]	44.4%[201/453]	**< 0.001[a]***

AI-DS = artificial intelligence decision support, reader A = initial radiologist, reader B = second radiologist, PPV = positive predictive value, NPV = negative predictive value. *$p < 0.05$, a = McNemar test, b = generalized estimating equation

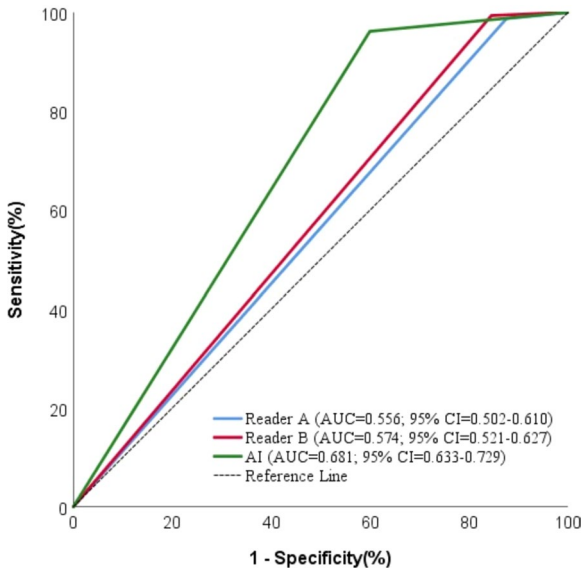


Fig. 2 Receiver operating characteristic curve for the readers and the AI-DS system

showed similar accuracy for BI-RADS 2–3 and BI-RADS 4C–5 (all $p < 0.001$). The diagnostic accuracy rates of the AI-DS and the radiologists according to BI-RADS classification are provided in Table 6.

Hypothetical clinical implementation

The AI-DS upgraded 13 lesions (12 masses and one non-mass lesion) that were initially classified as BI-RADS 3 by reader A to BI-RADS 4A–4B and 4C+. Among the 13 lesions, six were diagnosed as fibroadenomas, three as adenosis with stromal fibrosis, two as mastitis, and two as malignant lesions (IBC-NST). The AI-DS assigned both malignant cases to BI-RADS 4A–4B.

The AI-DS downgraded 99 lesions that reader A had classified as BI-RADS 4A, 4B, or 4C to BI-RADS 2 or 3. Among these, 93 were confirmed benign and six malignant. Among the malignant cases, five were classified as BI-RADS 4C and one as BI-RADS 4A by reader A. If biopsy had been omitted for the 84 lesions initially classified as BI-RADS 4A by reader A, the overall biopsy rate would have decreased by 18.6% (84/453), while only one malignancy would have been missed. Importantly, this missed malignant lesion was a non-mass lesion with microcalcifications, for which biopsy would have been performed regardless. Fig. 3 (mass lesion) and Fig. 4 (non-mass lesion) are shown as examples of lesions that were incorrectly downgraded by the AI-DS system.

Discussion

Accurate decision-making in ultrasound-guided breast biopsy is critical to balance early cancer detection with the avoidance of unnecessary invasive procedures. Despite the widespread use of BI-RADS-based assessment, variability among radiologists and the high prevalence of benign outcomes in biopsied lesions continue to pose challenges in clinical practice [5–9]. AI-DS systems have therefore been increasingly investigated as tools to enhance diagnostic consistency and improve specificity in breast imaging [12–14].

In this study, we assessed the diagnostic efficacy of the AI-DS in lesions undergoing US-guided biopsy and evaluated its potential to reduce biopsy rates performed by experienced breast radiologists. The AI-DS demonstrated sensitivity and NPV comparable to radiologists while showing significantly higher specificity, PPV, and overall accuracy. The system also exhibited a higher AUC. Kim et al. [14] found higher specificity, PPV, and accuracy for an AI system, whereas Wanderley et al. [15] reported that the AI system and radiologists demonstrated similar diagnostic performance. Studies involving multiple radiologists with varying experience levels also noted higher AUC values for AI systems [12, 13].

In other studies, AI systems showed sensitivity, specificity, and AUC values ranging from 79.2 to 100.0% [14, 15, 20–23], 39.0 to 82.0% [14, 15, 20–23], and 0.725 to 0.962, respectively [12–14, 23]. Readers demonstrated sensitivity, specificity, and AUC values ranging from 70 to 100% [14, 15, 20–23], 24.0 to 75% [14, 15, 20–23], and 0.653 to 0.924, respectively [12–14, 23]. Compared with these reports, both the AI-DS and radiologists in our study displayed lower specificity and AUC. Furthermore, we found a lower inter-reader agreement ($\kappa = 0.21–0.27$) between the AI-DS and radiologists than previously reported ($\kappa = 0.56–0.62$) [14, 20, 21]. This difference may primarily be attributed to the inclusion of only biopsy-proven lesions, which enriched the study population with diagnostically challenging cases—particularly BI-RADS 4A lesions—resulting in a narrower diagnostic spectrum. The inclusion of non-mass lesions, which are often more difficult to classify on ultrasound, may have further contributed to this difference. Additionally, our tertiary referral setting, which serves a high-risk population, limits comparability with screening-based cohorts. Moreover, in one previous study, the benign–malignant threshold was based on BI-RADS 4B rather than 4A, which would be expected to yield higher specificity and AUC [22].

The AI-DS upgraded 13 lesions, two of which were malignant. In other studies, malignancy rates were found to be 1.6 to 8.3% in lesions upgraded by the AI-DS from BI-RADS 2–3 to BI-RADS 4A–4B or 4C+ [15,

Table 4 Diagnostic performance of the AI-DS system vs. reader A by lesion characteristics

		Sensitivity			Specificity			Accuracy		
Variables	**Category**	**AI-DS**	**Reader A**	***p*-value**	**AI-DS**	**Reader A**	***p*-value**	**AI-DS**	**Reader A**	***p*-value**
Age (year)	≤ 44 year	97.9%[47/48]	95.8%[46/48]	> 0.999	45.4%[88/194]	14.9%[29/194]	**< 0.001***	55.8%[135/242]	31%[75/242]	**< 0.001***
	> 44 year	95.4%[103/108]	100%[108/108]	N/A	30.1%[31/103]	7.8%[8/103]	**< 0.001***	63.5%[134/211]	55%[116/211]	**0.001***
Size	≤ 10 mm	100%[30/30]	96.7%[29/30]	N/A	36.1%[35/97]	7.2%[7/97]	**< 0.001***	51.2%[65/127]	28.3%[36/127]	**< 0.001***
	> 10 mm	95.2%[120/126]	99.2%[125/126]	0.125	42% [84/200]	15%[30/200]	**< 0.001***	62.6%[204/326]	47.5%[155/326]	**< 0.001***
Breast density	Homogeneous-fat	95.2%[20/21]	100%[21/21]	N/A	23.1%[6/26]	7.7%[2/26]	0.219	55.3%[26/47]	48.9%[23/47]	0.453
	Heterogeneous	94.8%[73/77]	100%[77/77]	N/A	43.6%[51/117]	6%[7/117]	**< 0.001***	63.9%[124/194]	43.3%[84/194]	**< 0.001***
	Homogeneous-fibroglanduler	98.3%[57/58]	96.6%[56/58]	> 0.999	40.3%[62/154]	18.2%[28/154]	**< 0.001***	56.1%[119/212]	39.6%[84/212]	**< 0.001***
Skin distance	0–4	97.9%[47/48]	97.9%[47/48]	1	49.1%[52/106]	11.3%[12/106]	**< 0.001***	64.3%[99/154]	38.3%[59/154]	**< 0.001***
	5–9	97%[65/67]	98.5%[66/67]	> 0.999	36.3%[41/113]	10.6%[12/113]	**0.004***	58.9%[106/180]	43.3%[78/180]	**< 0.001***
	≥ 10	92.7%[38/41]	100%[41/41]	N/A	33.3%[26/78]	16.7%[13/78]	**< 0.001***	53.8%[64/119]	45.4%[54/119]	0.052
Mass	Mass	96.5%[139/144]	98.6%[142/144]	0.453	44.6%[112/251]	14.3%[36/251]	**< 0.001***	63.5%[251/395]	45.1%[178/395]	**< 0.001***
	Non-mass	91.7%[11/12]	100%[12/12]	N/A	15.2%[7/46]	2.2%[1/46]	0.070	31%[18/58]	22.4%[13/58]	0.180
Orientation	Parallel	93.8%[60/64]	96.9%[62/64]	0.687	48.9%[108/221]	15.8%[35/221]	**< 0.001***	58.9%[168/285]	34%[97/285]	**< 0.001***
	Not Parallel	98.8%[79/80]	100%[80/80]	N/A	13.3%[4/30]	3.3%[1/30]	0.375	75.5%[83/110]	73.6%[81/110]	0.687
Echo pattern	Complex cystic/solid	100%[2/2]	50%[1/2]	N/A	34.6%[9/26]	3.8%[1/26]	**0.008***	39.3%[11/28]	7.1%[2/28]	**0.004***
	Hypoechoic	96.1%[73/76]	98.7%[75/76]	0.625	46.6%[41/88]	23.9%[21/88]	**0.001***	69.5%[114/164]	58.5%[96/164]	**0.004***
	Isoechoic	100%[2/2]	100%[2/2]	N/A	52.5%[21/40]	15%[6/40]	**< 0.001***	54.8%[23/42]	19%[8/42]	**< 0.001***
	Heterogeneous	96.9%[62/64]	100%[64/64]	N/A	42.3%[41/97]	8.2%[8/97]	**< 0.001***	64%[103/161]	44.7%[72/161]	**< 0.001***
Shape	Oval	94.7%[18/19]	89.5%[17/19]	> 0.999	53%[87/164]	21.3%[35/164]	**< 0.001***	57.4%[105/183]	28.4%[52/183]	**< 0.001***
	Round	80%[8/10]	N/A	N/A	56.1%[23/41]	N/A	N/A	60.8%[31/51]	N/A	N/A
	Irregular	98.3%[113/115]	100%[115/115]	N/A	4.3%[2/46]	2.2%[1/46]	> 0.999	71.4%[115/161]	72%[116/161]	> 0.999
Margin	Circumscribed	81.8%[9/11]	81.8%[9/11]	> 0.999	59.3%[99/167]	17.4%[29/167]	**< 0.001***	60.7%[108/178]	21.3%[38/178]	**< 0.001***
	Angulated	100%[11/11]	100%[11/11]	N/A	25%[5/20]	15%[3/20]	0.625	51.6%[16/31]	45.2%[14/31]	0.625
	Microlobulated	100%[4/4]	N/A	N/A	25%[2/8]	N/A	N/A	50%[6/12]	N/A	N/A
	Indistinct	95.2%[60/63]	100%[63/63]	N/A	13.3%[6/45]	6.7%[3/45]	0.453	61.1%[66/108]	61.1%[66/108]	1
	Spiculated	N/A	100%[55/55]	N/A	N/A	9.1%[1/11]	N/A	N/A	84.8%[56/66]	N/A
Posterior features	No features	95.6%[109/114]	99.1%[113/114]	0.219	41%[112/273]	12.1%[33/273]	**< 0.001***	57.1%[221/387]	37.7%[146/387]	**< 0.001***
	Enhancement	75%[3/4]	75%[3/4]	1	58.3%[7/12]	25%[3/12]	0.219	62.5%[10/16]	37.5%[6/16]	0.289
	Shadowing	N/A	100%[38/38]	N/A	N/A	8.3%[1/12]	N/A	N/A	78%[39/50]	N/A

AI-DS = artificial intelligence decision support, reader A = initial radiologists. *$p < 0.05$, a: McNemar test, *NA*, not available

Table 5 Diagnostic performance of the AI-DS system vs. reader B by lesion characteristics

Variables	Category	Sensitivity			Specificity			Accuracy		
		AI-DS	Reader B	*p*-value	AI-DS	Reader B	*p*-value	AI-DS	Reader B	*p*-value
Age (year)	≤ 44 year	97.9%[47/48]	100%[48/48]	N/A	45.4%[88/194]	15.5%[30/194]	**< 0.001***	55.8%[135/242]	32.2%[78/242]	**< 0.001***
	> 44 year	95.4%[103/108]	99.1%[107/108]	0.125	30.1%[31/103]	15.5%[16/103]	**0.003***	63.5%[134/211]	58.3%[123/211]	0.052
Size	≤ 10 mm	100%[30/30]	100%[30/30]	N/A	36.1%[35/97]	11.3%[11/97]	**< 0.001***	51.2%[65/127]	32.3%[41/127]	**< 0.001***
	> 10 mm	95.2%[120/126]	99.2%[125/126]	0.063	42% [84/200]	17.5%[35/200]	**< 0.001***	62.6%[204/326]	49.1%[160/326]	**< 0.001***
Breast density	Homogeneous-fat	95.2%[20/21]	95.2%[20/21]	> 0.999	23.1%[6/26]	7.7%[2/26]	0.219	55.3%[26/47]	46.8%[22/47]	0.219
	Heterogeneous	94.8%[73/77]	100%[77/77]	N/A	43.6%[51/117]	7.7%[9/117]	**< 0.001***	63.9%[124/194]	44.3%[86/194]	**< 0.001***
	Homogeneous-fibroglanduler	98.3%[57/58]	100%[58/58]	N/A	40.3%[62/154]	22.7%[35/154]	**< 0.001***	56.1%[119/212]	43.9%[93/212]	**< 0.001***
Skin distance	0–4	97.9%[47/48]	100%[48/48]	N/A	49.1%[52/106]	16%[17/106]	**< 0.001***	64.3%[99/154]	42.2%[65/154]	**< 0.001***
	5–9	97%[65/67]	100%[67/67]	N/A	36.3%[41/113]	13.3%[15/113]	**< 0.001***	58.9%[106/180]	45.6%[82/180]	**< 0.001***
	≥ 10	92.7%[38/41]	97.6%[40/41]	0.500	33.3%[26/78]	17.9%[14/78]	**0.004***	53.8%[64/119]	45.4%[54/119]	**0.031***
Mass	Mass	96.5%[139/144]	99.3%[143/144]	0.125	44.6%[112/251]	18.3%[46/251]	**< 0.001***	63.5%[251/395]	47.8%[189/395]	**< 0.001***
	Non-mass	91.7%[11/12]	100%[12/12]	N/A	15.2%[7/46]	0%[0/46]	N/A	31%[18/58]	20.7%[12/58]	0.070
Orientation	Parallel	93.8%[60/64]	98.4%[63/64]	0.250	48.9%[108/221]	20.4%[45/221]	**< 0.001***	58.9%[168/285]	37.9%[108/285]	**< 0.001***
	Not Parallel	98.8%[79/80]	100%[80/80]	N/A	13.3%[4/30]	3.3%[1/30]	0.375	75.5%[83/110]	73.6%[81/110]	0.687
Echo pattern	Complex cystic/solid	100%[2/2]	100%[2/2]	N/A	34.6%[9/26]	3.8%[1/26]	**0.008***	39.3%[11/28]	10.7%[3/28]	**0.008***
	Hypoechoic	96.1%[73/76]	98.7%[75/76]	0.500	46.6%[41/88]	29.5%[26/88]	**0.011***	69.5%[114/164]	61.6%[101/164]	**0.035***
	Isoechoic	100%[2/2]	100%[2/2]	N/A	52.5%[21/40]	17.5%[7/40]	**0.001***	54.8%[23/42]	21.4%[9/42]	**0.001***
	Heterogeneous	96.9%[62/64]	100%[64/64]	N/A	42.3%[41/97]	12.4%[12/97]	**< 0.001***	64%[103/161]	47.2%[76/161]	**< 0.001***
Shape	Oval	94.7%[18/19]	100%[19/19]	N/A	53%[87/164]	26.2%[43/164]	**< 0.001***	57.4%[105/183]	33.9%[62/183]	**< 0.001***
	Round	80%[8/10]	90%[1/10]	> 0.999	56.1%[23/41]	2.4%[1/41]	**< 0.001***	60.8%[31/51]	19.6%[10/51]	**< 0.001***
	Irregular	98.3%[113/115]	100%[115/115]	N/A	4.3%[2/46]	4.3%[2/46]	1	71.4%[115/161]	72.7%[117/161]	0.687
Margin	Circumscribed	81.8%[9/11]	90.9%[10/11]	> 0.999	59.3%[99/167]	24%[40/167]	**< 0.001***	60.7%[108/178]	28.1%[50/178]	**< 0.001***
	Angulated	100%[11/11]	100%[11/11]	N/A	25%[5/20]	5%[1/20]	0.125	51.6%[16/31]	38.7%[12/31]	0.125
	Microlobulated	100%[4/4]	100%[4/4]	N/A	25%[2/8]	N/A	N/A	50%[6/12]	33.3%[4/12]	0.500
	Indistinct	95.2%[60/63]	100%[63/63]	N/A	13.3%[6/45]	8.9%[4/45]	0.727	61.1%[66/108]	62%[67/108]	> 0.999
	Spiculated	N/A	100%[55/55]	N/A	N/A	9.1%[1/11]	N/A	N/A	84.8%[56/66]	N/A
Posterior features	No features	95.6%[109/114]	99.1%[113/114]	0.125	41%[112/273]	15.4%[42/273]	**< 0.001***	57.1%[221/387]	40.1%[155/387]	**< 0.001***
	Enhancement	75%[3/4]	100%[4/4]	N/A	58.3%[7/12]	25%[3/12]	0.289	62.5%[10/16]	43.8%[7/16]	0.508
	Shadowing	N/A	100%[38/38]	N/A	N/A	8.3%[1/12]	N/A	N/A	78%[39/50]	N/A

AI-DS = artificial intelligence decision support, reader B = second radiologist. *$p < 0.05$, a: McNemar test, *NA,* not available

Table 6 Diagnostic accuracy of the AI-DS system vs. readers by BI-RADS category

BI-RADS classification	Reader(s)	Accuracy		*p*-value
		Reader	AI-DS	
BI-RADS 2–3	A	94.9%[37/39]	95.2%[119/125]	0.935[a]
	B	97.9%[46/47]	95.2%[119/125]	0.347[a]
BI-RADS 4A-4B	A	6%[15/251]	25.2%[52/206]	**< 0.001[a]***
	B	8.4%[21/251]	25.2%[52/206]	**< 0.001[a]***
BI-RADS 4C-5	A	85.3%[139/163]	80.3%[98/122]	0.730[a]
	B	86.5%[134/155]	80.3%[98/122]	0.539[a]

Reader A = initial radiologist, reader B = second radiologist.*$p < 0.001$, a: generalized estimating equation

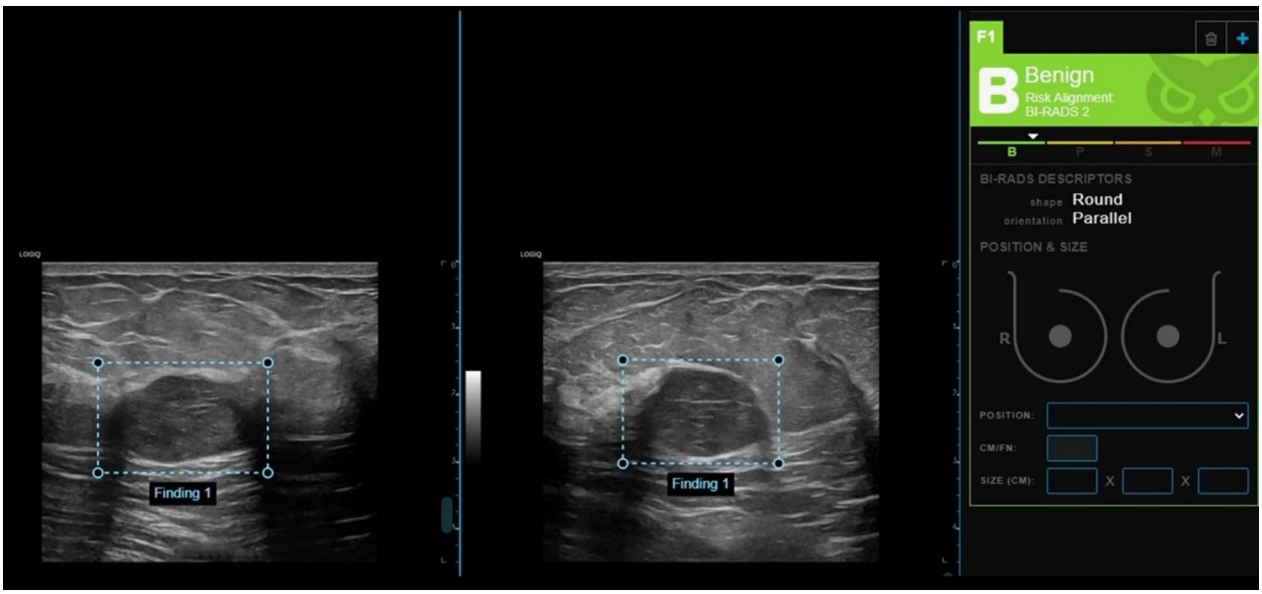


Fig. 3 A 74-year-old female patient presented with a palpable mass. Breast ultrasound revealed an oval-shaped, circumscribed hypoechoic lesion in the right breast measuring 24 × 17 × 14 mm. The lesion was categorized as BI-RADS 4C by reader A, BI-RADS 4B by reader B, and BI-RADS 2 by the AI-DS system. Histopathological examination confirmed a diagnosis of solid papillary carcinoma. AI-DS = artificial intelligence decision support, reader A = initial radiologist, reader B = second radiologist

20, 21, 24]. Given the small sample sizes of the studies conducted so far, larger cohorts are needed to make informed recommendations regarding which lesions may be appropriately upgraded by AI-DS systems. Fig. 5,6,7, and 8 illustrate representative cases from different categories.

Our study showed that, in a hypothetical scenario, if biopsies were not performed on lesions categorized as BI-RADS 4A by radiologists but classified as BI-RADS 2 or 3 by the AI-DS, the biopsy rate could be reduced by 18.6% while maintaining the same sensitivity. Wanderley et al. [15] reported that if biopsies were avoided for lesions classified as BI-RADS 2 or 3 by both the AI system and radiologists, as well as those downgraded from BI-RADS 4A to BI-RADS 2 or 3 by the AI system, biopsy rates could be reduced by 21.1%. Shen et al. [13] found that AI assistance reduced the number of biopsies recommended by radiologists by 27.8% while preserving sensitivity. They also noted that 53.3% of false positives fell within the BI-RADS 4A category and that 68.2% of these false positives could be avoided using an AI system. Wang et al. [24] similarly reported that if lesions classified as BI-RADS 4A by radiologists and downgraded by the AI were not biopsied, biopsy rates for BI-RADS 4A lesions could be reduced by 32.6%. Brown et al. [25] observed that AI-DS increased PPV and reduced benign biopsies but emphasized that several malignancies were downgraded by AI, underscoring the need to balance biopsy reduction with the risk of missed cancers. Collectively, these data indicate that AI-DS systems may help

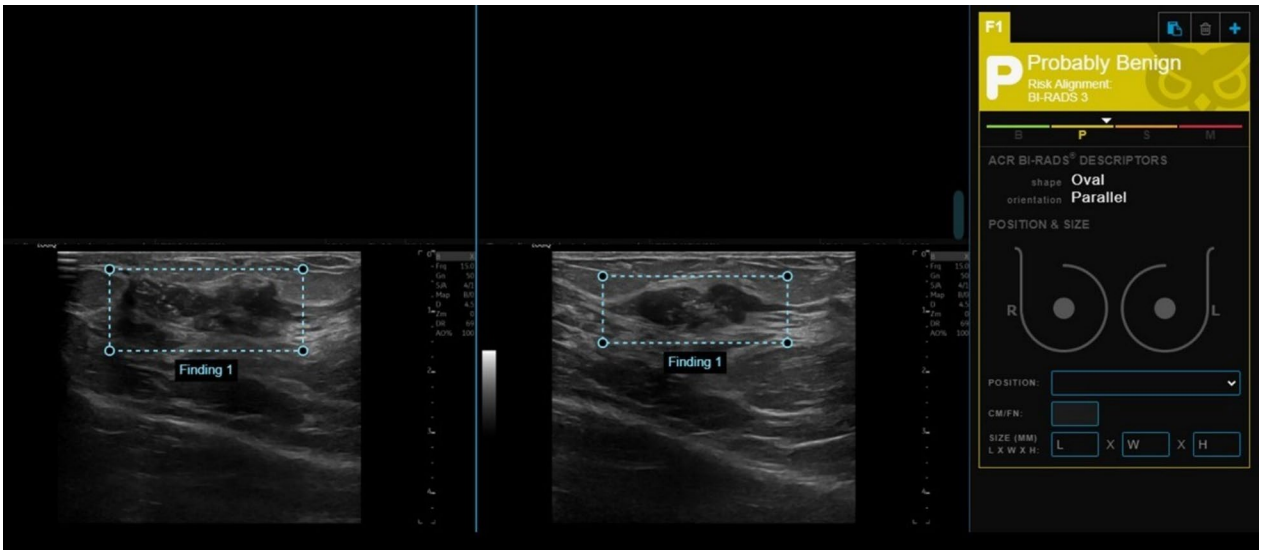


Fig. 4 A 61-year-old female patient presented with skin redness on her left breast. Breast ultrasound revealed a heterogeneous hypoechoic non-mass lesion in the left breast measuring 31 × 12 × 10 mm and containing microcalcifications. The lesion was classified as BI-RADS 4C by reader A, BI-RADS 4B by reader B, and BI-RADS 3 by the AI-DS system. Histopathological examination confirmed the diagnosis of HER2-enriched IBC-NST. AI-DS = artificial intelligence decision support, reader A = initial radiologist, reader B = second radiologist

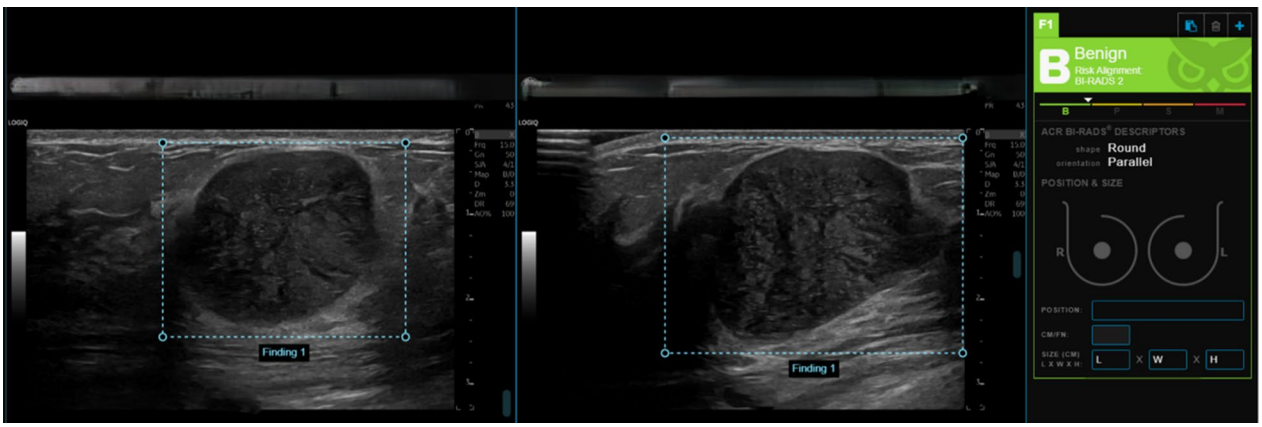


Fig. 5 A 15-year-old female patient presented with a rapidly growing breast mass. Breast ultrasound demonstrated a round, well-circumscribed, heterogeneous mass measuring 35 × 24 × 22 mm. The lesion was classified as BI-RADS 4A by both reader A and reader B, whereas it was classified as BI-RADS 2 by the AI-DS system. Histopathological examination confirmed the diagnosis of fibroadenoma. AI-DS = artificial intelligence decision support, reader A = initial radiologist, reader B = second radiologist

reduce unnecessary biopsies for BI-RADS 4A lesions; however, the possibility of false negatives remains a critical concern.

In a potential real-time clinical workflow, the AI-DS may be incorporated as a second-reader or consensus-support tool rather than a primary decision-maker. After the radiologist's initial BI-RADS classification, the AI-DS output could be reviewed in parallel. If the AI-DS suggests a lower-risk category for a lesion initially classified as BI-RADS 4A, this discrepancy could prompt a structured re-evaluation of the case. During re-evaluation, radiologists would integrate clinical history, interval imaging changes, Doppler and elastography findings, and the presence of microcalcifications—factors that AI systems are currently unable to assess. As reported in previous studies, these additional clinical and imaging factors are often considered in biopsy decision-making [15, 24]. In this model, the AI-DS functions as a triage or consensus aid that supports decision-making while the radiologist retains full diagnostic authority.

This approach underscores that the role of AI-DS is complementary rather than autonomous. However,

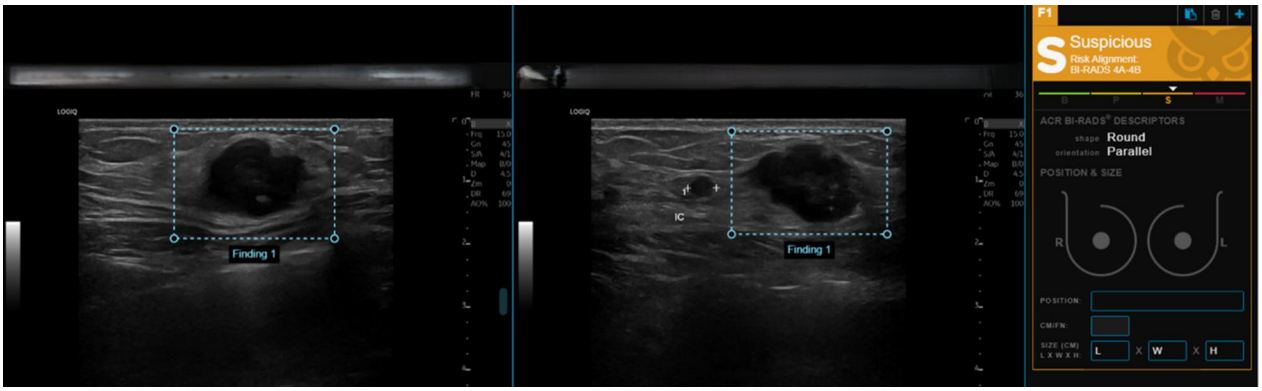


Fig. 6 A 32-year-old female patient presented with a palpable breast mass. Breast ultrasound demonstrated an oval mass with angulated margins, hypoechoic echotexture, and parallel orientation, measuring 20 × 14 × 12 mm. The lesion was classified as BI-RADS 4A by reader A and BI-RADS 4B by reader B, whereas it was classified as BI-RADS 4A–4B by the AI-DS system. Histopathological examination confirmed the diagnosis of invasive ductal carcinoma, no special type (NST). AI-DS = artificial intelligence decision support, reader A = initial radiologist, reader B = second radiologist

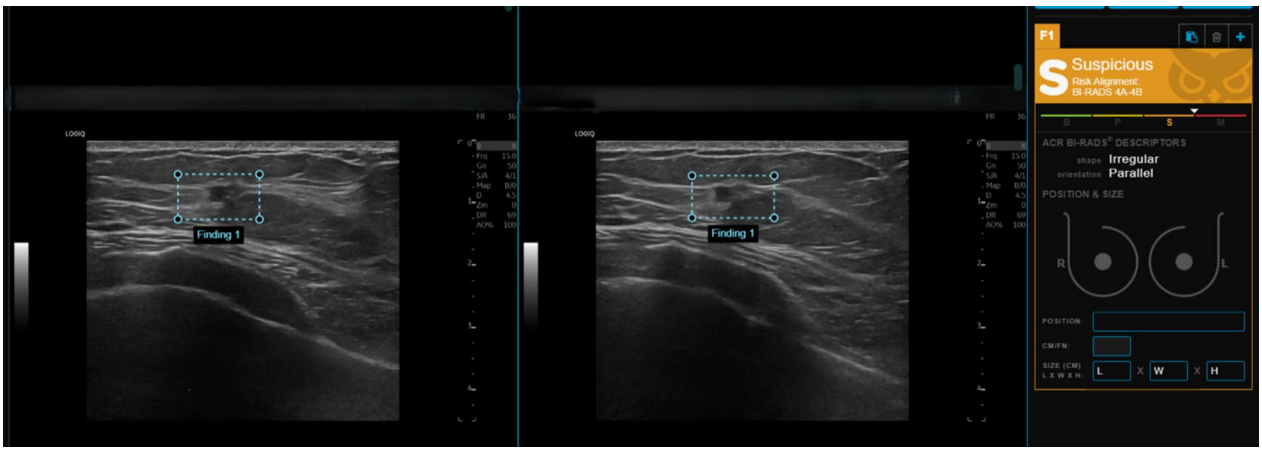


Fig. 7 A 61-year-old female patient presented for screening. Breast ultrasound demonstrated an irregular mass with spiculated margins, heterogeneous echotexture, and non-parallel orientation, measuring 7 × 6 × 5 mm. The lesion was classified as BI-RADS 4B by both reader A and reader B, whereas it was classified as BI-RADS 4C+ by the AI-DS system. Histopathological examination confirmed the diagnosis of invasive ductal carcinoma, no special type (NST). AI-DS = artificial intelligence decision support, reader A = initial radiologist, reader B = second radiologist

further investigation is needed to determine whether such a workflow is feasible and safe in real-world clinical practice. Prospective, multi-center studies with substantial sample sizes are required to evaluate the performance of AI-DS systems in real-time clinical settings, establish criteria for when re-evaluation should be triggered, and ensure that biopsy reduction does not compromise cancer detection.

In our study, the AI-DS demonstrated higher accuracy than radiologists for the BI-RADS 4A–4B category but comparable accuracy for BI-RADS 2–3 and BI-RADS 4C–5 categories. Wanderley et al. [15] similarly reported that benign and malignant rates were comparable between the AI-DS and radiologists for BI-RADS 2–3 (AI-DS: 97.0% vs. 3.0%; radiologist: 98.3% vs. 1.7%) and BI-RADS 4C–5 (AI-DS: 6.9% vs. 93.1%; radiologist: 7.6% vs. 92.4%) categories. However, in the BI-RADS 4A–4B category, AI-DS yielded a higher malignancy rate (AI-DS: 69.7% vs. 30.3%; radiologist: 80.4% vs. 19.6%), suggesting that AI-DS may more accurately downgrade lesions that radiologists categorize as BI-RADS 4A.

Subgroup analyses indicated that the diagnostic performance of the AI-DS varied according to lesion characteristics and imaging context. The AI-DS tended to show higher accuracy for lesions located superficially (depth < 10 mm) and for those surrounded by fibroglandular tissue, where ultrasound interpretation can be limited by acoustic heterogeneity. For these subgroups, the AI-DS outperformed both readers, whereas for deeper lesions (≥ 10 mm) or those surrounded by homogeneous

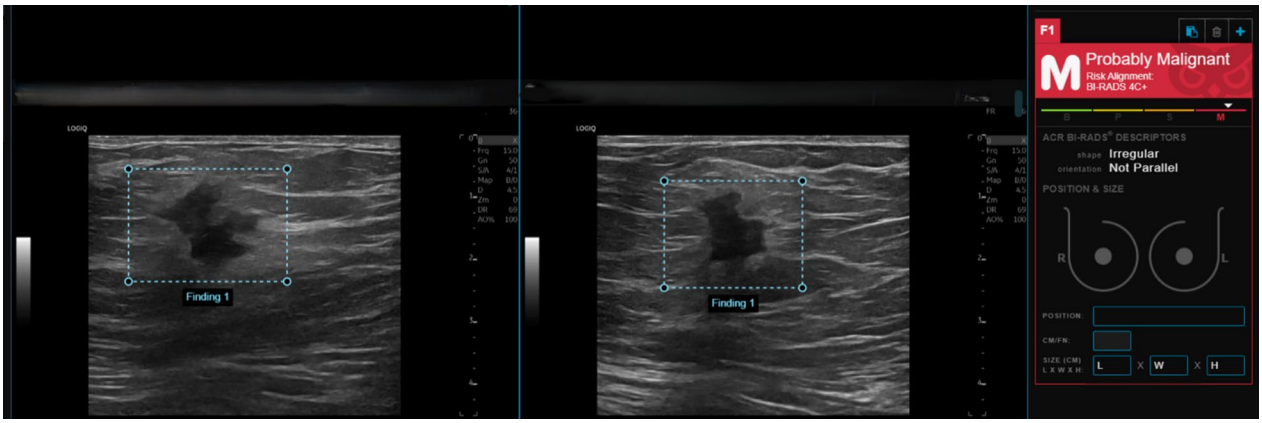


Fig. 8 A 58-year-old female patient presented with a palpable breast mass. Breast ultrasound demonstrated an irregular mass with spiculated margins, heterogeneous echotexture, and non-parallel orientation, measuring 15 × 12 × 8 mm. The lesion was classified as BI-RADS 5 by both reader A and reader B, whereas it was classified as BI-RADS 4C+ by the AI-DS system. Histopathological examination confirmed the diagnosis of invasive ductal carcinoma, no special type (NST). AI-DS = artificial intelligence decision support, reader A = initial radiologist, reader B = second radiologist

fat tissue, its accuracy was comparable to that of reader A.

Regarding lesion type, the AI-DS achieved higher accuracy for mass lesions—particularly those with parallel orientation, oval shape, circumscribed margins, and no posterior features—while performing similarly to the radiologists for non-mass lesions and for masses with irregular shape or posterior enhancement. These findings suggest that the AI-DS may provide incremental benefit in well-defined or superficially located lesions and in cases where background tissue composition complicates interpretation. However, given the relatively small size of some subgroups, these observations should be interpreted with caution and validated in larger, more diverse cohorts.

Lai et al. [23] reported comparable trends, showing that AI assistance improved diagnostic performance across most subgroups, except for lesions deeper than 2 cm. Heller et al. [20], however, found higher accuracy for irregular and hypoechoic masses but lower accuracy for round or anechoic lesions. Such discrepancies across studies likely reflect differences in subgroup definitions, lesion distribution, and the degree of subjectivity in morphological classification.

Only a few studies have included non-mass lesions [21, 23, 26], and the diagnostic performance of AI systems in this subgroup is not yet well established. In our study, the AI-DS performed similarly to radiologists for non-mass lesions; however, given the small number of non-mass lesions in our cohort, larger studies are needed to validate these preliminary findings.

This study has several limitations. First, it was conducted retrospectively at a single tertiary referral center, which may limit the generalizability of the findings. The study population predominantly consisted of women with personal or family histories of breast cancer, and the results may therefore differ in screening settings or general populations. Second, all examinations were performed by experienced breast radiologists using a high-end ultrasound system; thus, diagnostic performance may differ in centers with less experienced operators or different equipment. Third, only biopsied lesions were included, introducing potential selection bias and producing a narrower diagnostic spectrum (e.g., a higher proportion of diagnostically challenging lesions such as BI-RADS 4A and select non-mass lesions). Fourth, subgroup analyses were based on subjective stratification, and the relatively small number of cases within each subgroup may have reduced statistical power. Finally, the real-world role of AI-DS was not evaluated, and prospective studies are needed to determine its practical utility and integration into clinical workflows.

Conclusions

Our study found that the AI-DS demonstrated sensitivity comparable to that of experienced breast radiologists in lesions undergoing US-guided biopsy while also exhibiting higher specificity and accuracy. Additionally, the AI-DS showed higher specificity and accuracy across stratified subgroups based on age, lesion size, depth, breast composition, and morphological features. Moreover, AI-DS systems have the potential to reduce unnecessary biopsies for BI-RADS 4A lesions while maintaining comparable sensitivity.

Conflict of interest

The authors declare no competing interests.

Abbreviations

AI-DS	Artificial intelligence decision support
US	Ultrasonography
PPV	Positive predictive value
AUC	Area under the curve
NPV	Negative predictive value
CI	Confidence interval
IBC-NST	Invasive breast carcinoma of no special type

Author contribution

All authors contributed to the study conception and design.Study design, Data acquisition: Meliha akin, Fusun Taskin. Data enrollment and confirmation: Meliha akin, Fusun Taskin, Servet Erdemli, Onur Dulgeroglu, Fatma Tokat. Data analysis: Meliha akin, Fusun Taskin, Servet Erdemli, Onur Dulgeroglu, Fatma Tokat. Draft of the manuscript: Meliha akin, Fusun Taskin. All authors read and approved the manuscript.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to institutional policy but are available from the corresponding author on reasonable request.

Received: 1 October 2025 Accepted: 6 January 2026
Published online: 28 January 2026

References

1. Coldman A, Phillips N, Warren L, Kan L (2007) Breast cancer mortality after screening mammography in British Columbia women. Int J Cancer 120(5):1076–1080
2. Broeders M, Moss S, Nyström L, Njor S, Jonsson H, Paap E, Massat N, Duffy S, Lynge E, Paci E (2012) The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen 19(suppl 1):14–25
3. Chiu SYH, Duffy S, Yen AMF, Tabár L, Smith RA, Chen HH (2010) Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. Cancer Epidemiol Biomarkers Prev 19(5):1219–1228
4. Von Euler-Chelpin M, Lillholm M, Vejborg I, Nielsen M, Lynge E (2019) Sensitivity of screening mammography by density and texture: a cohort study from a population-based screening program in Denmark. Breast Cancer Res 21(1):111
5. Ohuchi N, Suzuki A, Sobue T, Kawai M, Yamamoto S, Zheng YF, Shiono YN, Saito H, Kuriyama S, Tohno E, Endo T, Fukao A, Tsuji I, Yamaguchi T, Ohashi Y, Fukuda M, Ishida T (2016) Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the J-START trial. Lancet 387(10016):341–348
6. Berg WA, Bandos AI, Mendelson EB, Lehrer D, Jong RA, Pisano ED (2016) Ultrasound as the primary screening test for breast cancer: analysis from ACRIN 6666. J Natl Cancer Inst 108(4):djv367
7. Berg WA, Blume JD, Cormack JB, Mendelson EB (2006) Operator dependence of physician-performed whole-breast US: lesion detection and characterization. Radiology 241(2):355–365
8. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE (2012) Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. Radiology 265(1):59–69
9. Schaefer FK, Waldmann A, Katalinic A, Wefelnberg C, Heller M, Jonat W, Schreer I (2010) Influence of additional breast ultrasound on cancer detection in a cohort study for quality assurance in breast diagnosis. Eur Radiol 20(5):1085–1092
10. Chang JM, Koo HR, Moon WK (2015) Radiologist-performed hand-held ultrasound screening at average risk of breast cancer: results from a single health screening center. Acta Radiol 56(6):652–658
11. Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH, Shah K (2016) Supplemental screening for breast cancer in women with dense breasts: a systematic review. Ann Intern Med 164(4):268–278
12. Mango VL, Sun M, Wynn RT, Ha R (2020) Should we ignore, follow, or biopsy? Impact of artificial intelligence decision support on breast ultrasound lesion assessment. AJR Am J Roentgenol 214(6):1445–1452
13. Shen Y, Shamout FE, Oliver JR, Witowski J, Kannan K, Park J, Wu N, Huddleston C, Wolfson S, Millet A, Ehrenpreis R, Awal D, Tyma C, Samreen N, Gao Y, Chhor C, Gandhi S, Lee C, Kumari-Subaiya S, Leonard C, Mohammed R, Moczulski C, Altabet J, Babb J, Lewin A, Reig B, Moy L, Heacock L, Geras KJ (2021) Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. Nat Commun 12(1):5645
14. Kim K, Song MK, Kim EK, Yoon JH (2017) Clinical application of S-Detect to breast masses on ultrasonography: diagnostic performance and agreement with a dedicated breast radiologist. Ultrasonography 36(1):3–9
15. Wanderley MC, Soares CMA, Morais MMM, Cruz RM, Lima IRM, Chojniak R, Bitencourt AGV (2023) Application of artificial intelligence in predicting malignancy risk in breast masses on ultrasound. Radiol Bras 56(5):229–234
16. Becker AS, Mueller M, Stoffel E, Marcon M, Ghafoor S, Boss A (2018) Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study. Br J Radiol 91(1083):20170576
17. Tanaka H, Chiu SW, Watanabe T, Kaoku S, Yamaguchi T (2019) Computer-aided diagnosis system for breast ultrasound images using deep learning. Phys Med Biol 64(23):235013
18. Cao Z, Duan L, Yang G, Yue T, Chen Q (2019) An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. BMC Med Imaging 19(1):51
19. Mendelson EB, Böhm-Vélez M, Berg WA, Whitman GJ, Feldman MI, Madjar H, Rizzatto G, Baker JA, Zuley M, Stavros AT, Comstock C, Van Duyn Wear V. ACR BI-RADS® Ultrasound. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology; 2013.
20. Heller SL, Wegener M, Babb JS, Gao Y (2021) Can an artificial intelligence decision aid decrease false-positive breast biopsies? Ultrasound Q 37(1):10–15
21. Guldogan N, Taskin F, Icten GE, Yilmaz E, Turk EB, Erdemli S, Parlakkilic UT, Turkoglu O, Aribal E (2024) Artificial intelligence in BI-RADS categorization of breast lesions on ultrasound: can we omit excessive follow-ups and biopsies? Acad Radiol 31(6):2194–2202
22. O'Connell AM, Bartolotta TV, Orlando A, Jung S, Baek J, Parker KJ (2022) Diagnostic performance of an artificial intelligence system in breast ultrasound. J Ultrasound Med 41(1):97–105
23. Lai YC, Chen HH, Hsu JF, Hong YJ, Chiu TT, Chiou HJ (2022) Evaluation of physician performance using a concurrent-read artificial intelligence system to support breast ultrasound interpretation. Breast 65:124–135
24. Wang XY, Cui LG, Feng J, Chen W (2021) Artificial intelligence for breast ultrasound: an adjunct tool to reduce excessive lesion biopsy. Eur J Radiol 138:109624
25. Browne JL, Pascual MÁ, Perez J, Salazar S, Valero B, Rodriguez I, Cassina D, Alcázar JL, Guerriero S, Graupera B (2023) AI: can it make a difference to the predictive value of ultrasound breast biopsy? Diagnostics 13(4):811
26. Xiang H, Wang X, Xu M, Zhang Y, Zeng S, Li C, Liu L, Deng T, Tang G, Yan C, Ou J, Lin Q, He J, Sun P, Li A, Chen H, Heng PA, Lin X (2023) Deep learning-assisted diagnosis of breast lesions on US images: a multivendor, multicenter study. Radiol Artif Intell 5(5):e220185

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.