

Table S1 Diagnostic performance of radiologists with and without assistance of AI system using BI-RADS category 3 as threshold

Radiologist	Sensitivity (%)			Specificity (%)			AUC (95% CI)		
	BI-RADS _{initial}	BI-RADS _{adjusted}	BI-RADS _{second reading}	BI-RADS _{initial}	BI-RADS _{adjusted}	BI-RADS _{second reading}	BI-RADS _{initial}	BI-RADS _{adjusted}	BI-RADS _{second reading}
Total									
Junior	97.4	99.1	99.1	70.3	69.0	75.2	0.879 (0.852, 0.902)	0.901 (0.876, 0.922)	0.921 (0.898, 0.940)
P	0.125	>0.99	0.125	0.18	<0.001	<0.001	<0.001	<0.001	<0.001
Senior	99.1	99.1	97.4	68.5	67.2	69.4	0.950 (0.931, 0.965)	0.946 (0.926, 0.962)	0.951 (0.932, 0.966)
P	>0.99	0.125	0.125	0.031	0.002	0.125	0.059	0.399	0.907
D ≤2 cm									
Junior	94.1	98.0	98.0	75.3	74.3	80.7	0.864 (0.831, 0.893)	0.895 (0.865, 0.920)	0.924 (0.897, 0.945)
P	0.125	>0.99	0.125	0.388	<0.001	<0.001	<0.001	<0.001	<0.001
Senior	98.0	98.0	94.1	73.3	72.3	74.3	0.935 (0.909, 0.954)	0.935 (0.910, 0.955)	0.936 (0.911, 0.956)
P	>0.99	0.125	0.125	0.125	0.008	0.125	0.898	0.921	0.902
D >2 cm									
Junior	100.0	100.0	100.0	26.1	21.7	26.1	0.744 (0.672, 0.808)	0.774 (0.704, 0.834)	0.794 (0.726, 0.852)
P	–	–	–	0.50	0.50	>0.99	0.036	0.143	0.003
Senior	100.0	100.0	100.0	26.1	21.7	26.1	0.930 (0.881, 0.963)	0.910 (0.857, 0.948)	0.906 (0.852, 0.945)
P	–	–	–	0.50	0.50	>0.99	0.067	0.832	0.136

BI-RADS assessment category 3, 4B, 4C, and 5 were considered positive for malignant for calculation of sensitivity and specificity. P values in the BI-RADS_{initial} column were calculated by comparing with BI-RADS_{adjusted}; P values in the BI-RADS_{adjusted} column were calculated by comparing with BI-RADS_{second reading}; P values in the BI-RADS_{second reading} column were calculated by comparing with BI-RADS_{initial}. AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System; AUC, area under the curve; CI, confidence interval; BI-RADS_{initial}, radiologists analyze the images without the assistance of AI; BI-RADS_{adjusted}, incorporate BI-RADS provided by AI into BI-RADS_{initial}; BI-RADS_{second reading}, combining ultrasound features extracted by AI to conduct a second reading; D, largest diameter of breast lesions.

Table S2 Management decision changes of radiologists with assistance of AI system using BI-RADS category 3 as threshold

Parameters	BI-RADS _{adjusted}			BI-RADS _{second reading}		
	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 3, 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_{adjusted}, incorporate BI-RADS provided by AI into BI-RADS_{initial}; BI-RADS_{initial}, radiologists analyze the images without the assistance of AI; BI-RADS_{second reading}, combining ultrasound features extracted by AI to conduct a second reading; D, largest diameter of breast lesions.