Original article

Artificial intelligence decision support in automated breast ultrasound: improving diagnostic accuracy and reducing unnecessary biopsies

Chayaporn Hasdiseve*, Jenjeera Prueksadee

Departments of Radiology, King Chulalongkorn Memorial Hospital, Thai Redcross Society, Bangkok, Thailand

Abstract

Background: The emerging roles of artificial intelligence (AI) support in the imaging of the breast have led to improved radiologist performance.
Objective: This study assessed the diagnostic performance of the AI decision support in the evaluation of breast masses using automated breast ultrasound (ABUS).
Methods: One hundred eighty-two patients (415 breast masses) who received ABUS were included. Two readers, including an experienced breast radiologist (reader 1) and the breast imaging fellow (reader 2), separately reviewed the ABUS images and the AI decision support according to the American College of Radiology BI-RADS 5th edition guidelines.
Results: In the 415 masses that were evaluated, 395 (95.2%) were benign, and 20 (4.8%) were malignant. The area under the receiver operating curve (AUC) of the AI decision support was 0.74 (95% confidence interval (CI) 0.72–0.77) with a sensitivity and specificity of 100.0% and 48.6%, respectively. The integration of AI decision support significantly increased the AUC for both readers, from 0.82 (95% CI 0.74–0.91) to 0.85 (95% CI 0.76–0.93) for reader 1 ($P < 0.001$) and from 0.79 (95% CI 0.71–0.88) to 0.81 (95% CI 0.73–0.89) for reader 2 ($P < 0.001$). Furthermore, the AI decision support led to a 14.2% and 16.9% alteration in BI-RADS, with a 22.2% and 10.7% reduction in biopsies of benign masses for reader 1 and reader 2, respectively.
Conclusion: AI decision support demonstrates diagnostic performance comparable to that of radiologists, exhibiting high sensitivity and a high negative predictive value. Integrating AI into the diagnostic workflow may potentially enhance the diagnostic performance of radiologists across various experience levels and thereby contribute to a reduction in unnecessary biopsies of benign masses.

Keywords: Artificial intelligence, breast neoplasms, ultrasound.

Breast cancer is the most frequently diagnosed cancer among women and the leading cause of cancer-related mortality worldwide. In Southeast Asia, the incidence was 41.7 per 100,000 individuals, while the mortality rate was 14.8 per 100,000 individuals during 2020.[1] According to the breast cancer screening guidelines of the American College of Radiology (ACR), women who have an average risk of developing breast cancer should undergo annual mammography screening at 40 years of age, while those at higher-than-average risk are advised to commence annual mammography with supplemental screening between the ages of 25 and 40 years.[2] Additional handheld ultrasound (HHUS) or automated breast ultrasound (ABUS) has enhanced the rate of breast cancer detection, particularly in women with dense breast tissue, alongside a corresponding increase in false-positive examinations.[2-4] Moreover, less operator dependence, less operator variability, and less interpretation time for ABUS have been observed.[4]

Recent research has validated the improved performance of radiologists with less interobserver variability after the integration of artificial intelligence (AI) decision support.[5-7] Furthermore, a reduced number of biopsies for benign breast masses has been demonstrated.[5, 8]

***Correspondence to: Chayaporn Hasdiseve**, Departments of Radiology, King Chulalongkorn Memorial Hospital, Thai Redcross Society, Bangkok 10330, Thailand.
E- mail: c.hasdiseve@hotmail.com
Received: March 21, 2025
Revised: June 19, 2025
Accepted: August 28, 2025

With the emerging roles of AI decision support in breast imaging, this study aims to assess the diagnostic performance of the AI decision support in the evaluation of breast masses using ABUS.

Material and methods

Study population

This study received expedited review from the Institutional Review Board (IRB no. 0808/66, year issued 2023), in which informed consent was waived. Patients undergoing ABUS examinations at a single center from January 1, 2018, to September 30, 2021, were included in this study. The inclusion criteria included patients with mass lesions that were either a pathological result or follow-up imaging that indicated stability for more than two years. Furthermore, patients with non-mass lesions, such as ductal lesions, calcifications, or architectural distortion, simple cysts, abnormalities of axillary nodes, and negative findings (BI-RADS 1) were excluded from the study **(Figure 1).**

Out of the 1,203 patients who underwent ABUS at our center from January 1, 2018, to September 30, 2021, 182 patients were selected, which included a total of 415 masses. All participants were female, with a mean age of 46.4 ± 13.4 years. Approximately 12.6% of patients had a history of breast cancer. The examination was indicated for screening in 17.0% of cases and for diagnosis in the remaining 83.0%. The average number of masses per patient was 2.3 ± 2.2 masses.

Reader workflow

Two readers, including a breast radiologist with more than 10 years of experience (reader 1) and a breast imaging fellow (reader 2), independently examined the ABUS images in accordance with the BI-RADS 5th edition guidelines. [9] The evaluation was performed in two sessions, with a 4-week washout period between the sessions. To minimize result variation from differences in the region of interest (ROI) boundary, both readers were provided with AI decision support results featuring a similarly drawn ROI.

ABUS and AI decision support

ABUS was conducted using the GE Healthcare Invenia ABUS 2.0 with a reverse curve transducer (C15-6XW) and a frequency range of 6–15 MHz that was managed by technologists. Each breast underwent anterior-posterior (AP), medial, and lateral views, with additional superior and/or inferior views in patients with larger breasts. The AI decision support evaluated in this study was the Koios DS Breast software version 3.3 (Koios Medical), which could be used with the handheld ultrasound by the GE E10 machine and the Invenia ABUS 2.0. The AI decision support was integrated into the ABUS workstation. Following the placement of the marker at the mass (yellow circle in **Figures 2, 3, 4, and 5A**), readers could access the Koios DS Breast software. A tightly fitted rectangular ROI was drawn over the mass in two orthogonal views (yellow rectangle in **Figure 2, 3, 4, and 5A**). There would be a pop-up result window, and the software provided results that were categorized as benign (BI-RADS 2), probably benign (BI-RADS 3), suspicious (combining BI-RADS 4a and 4b), and probably malignant (combining BI-RADS 4c and 5). A confidence level bar displayed a continuum of the BI-RADS assessment categories.

Statistical analysis

The sample size was calculated as per reliability studies for dichotomous responses, where no distinction between the two types of variables was made, which yielded a requirement of 401 samples. [10] Samples were randomly selected from the target population using a simple random method, and data analysis was performed using STATA 18 software. Data were expressed as numbers, percentages, the mean, and standard deviation. Furthermore, the area under the receiver operating characteristics (AUC), 95% confidence interval (CI), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), as well as agreement between the two radiologists with Kappa statistics, were calculated, and statistical significance was established at $P < 0.05$.

Results

Characteristics of the masses

Out of the 415 included masses, 95.2% were categorized as benign, while the remaining 4.8% were categorized as malignant. Among these masses, 17.4% had biopsy-proven results indicating either benign or malignant status. The remaining benign masses exhibited stability in the follow-up imaging for more than two years. The average size of the masses was 10.3 ± 6.2 mm in the transverse dimension and 5.6 ± 3.7 mm in the AP dimension.

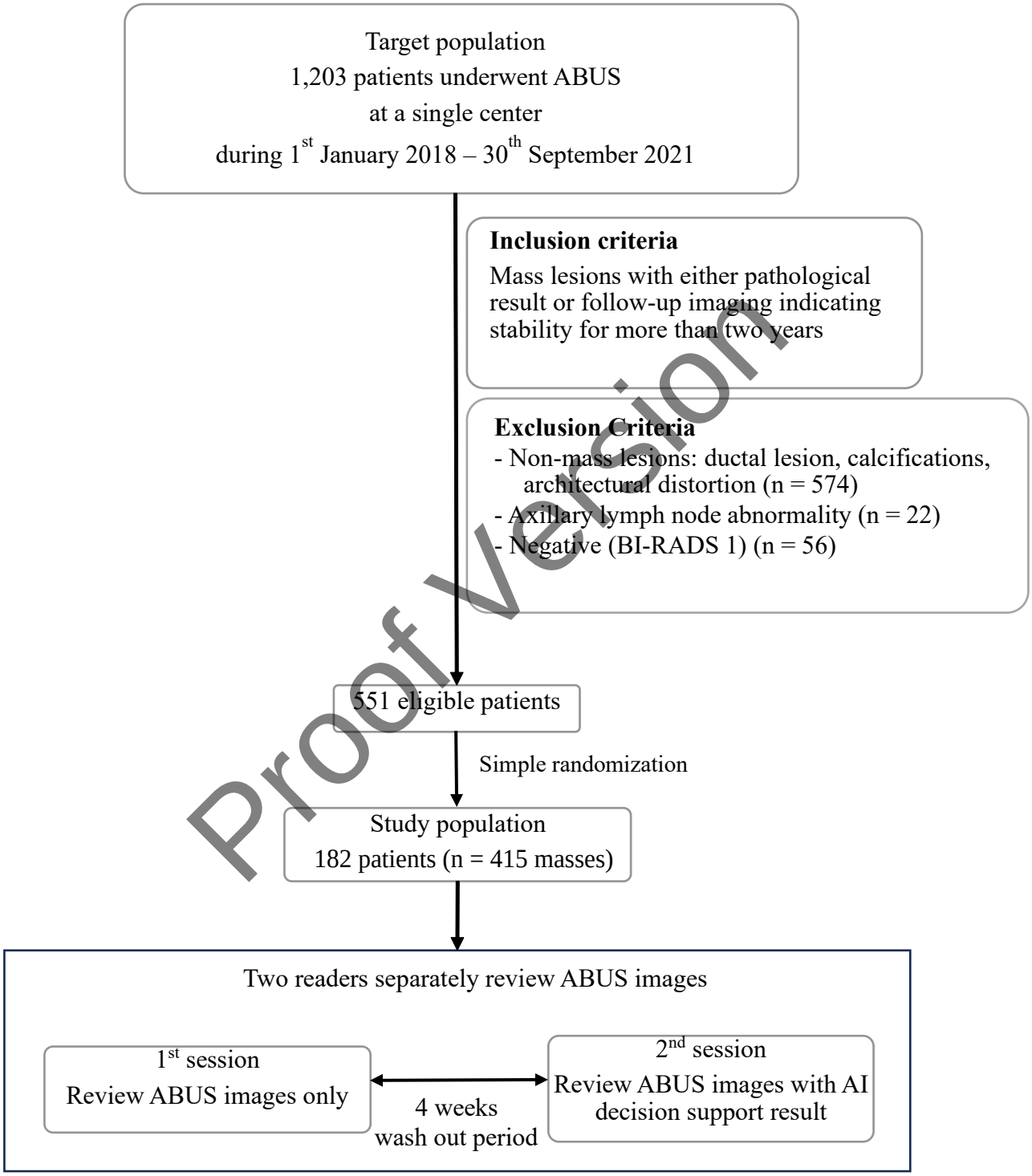

Figure 1. Schematic of the workflow. ABUS, automated breast ultrasound; AI, artificial intelligence; BI-RADS, breast imaging-reporting and data system.

Figure 2. A 64-year-old patient with invasive ductal carcinoma. ABUS reveals an irregular, heterogeneous hypoechoic mass with angular margin and internal calcification in the upper inner quadrant of the right breast, classified as BI-RADS 5 by both readers and probably malignant by the artificial intelligence decision support. ABUS, automated breast ultrasound; BI-RADS, breast imaging-reporting and data system.

Figure 3. A 24-year-old female presented with a palpable mass in her right breast which surgical excision revealed fibrocystic change with fibroadenomatous change. ABUS depicts an irregular, indistinct hypoechoic mass with non-parallel orientation in the upper outer quadrant of the left breast. The mass was classified as BI-RADS 4c by both readers and probably malignant by the artificial intelligence decision support. ABUS, automated breast ultrasound; BI-RADS, breast imaging-reporting and data system.

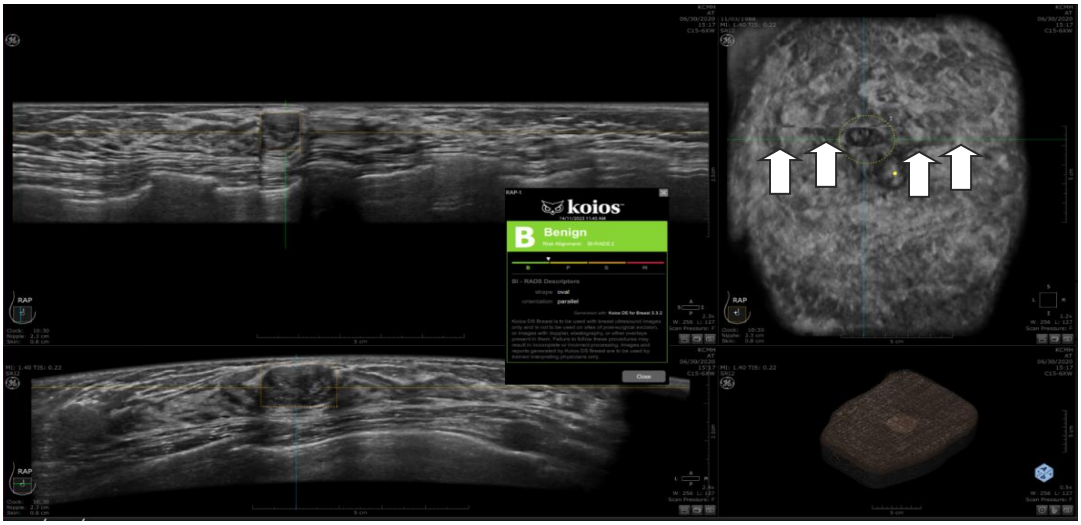

Figure 4. A 31-year-old female with fibroadenoma. ABUS reveals a circumscribed oval heterogeneous hypoechoic mass with skip artifact (arrow) considered benign by the AI decision support. The mass was initially classified as BI-RADS 4a by both readers. After reading with the AI decision support, reader 1 re-classified the mass as BI-RADS 2 and reader 2 as BI-RADS 3 in which biopsy would be omitted. ABUS, automated breast ultrasound; AI, artificial intelligence; BI-RADS, breast imaging-reporting and data system.

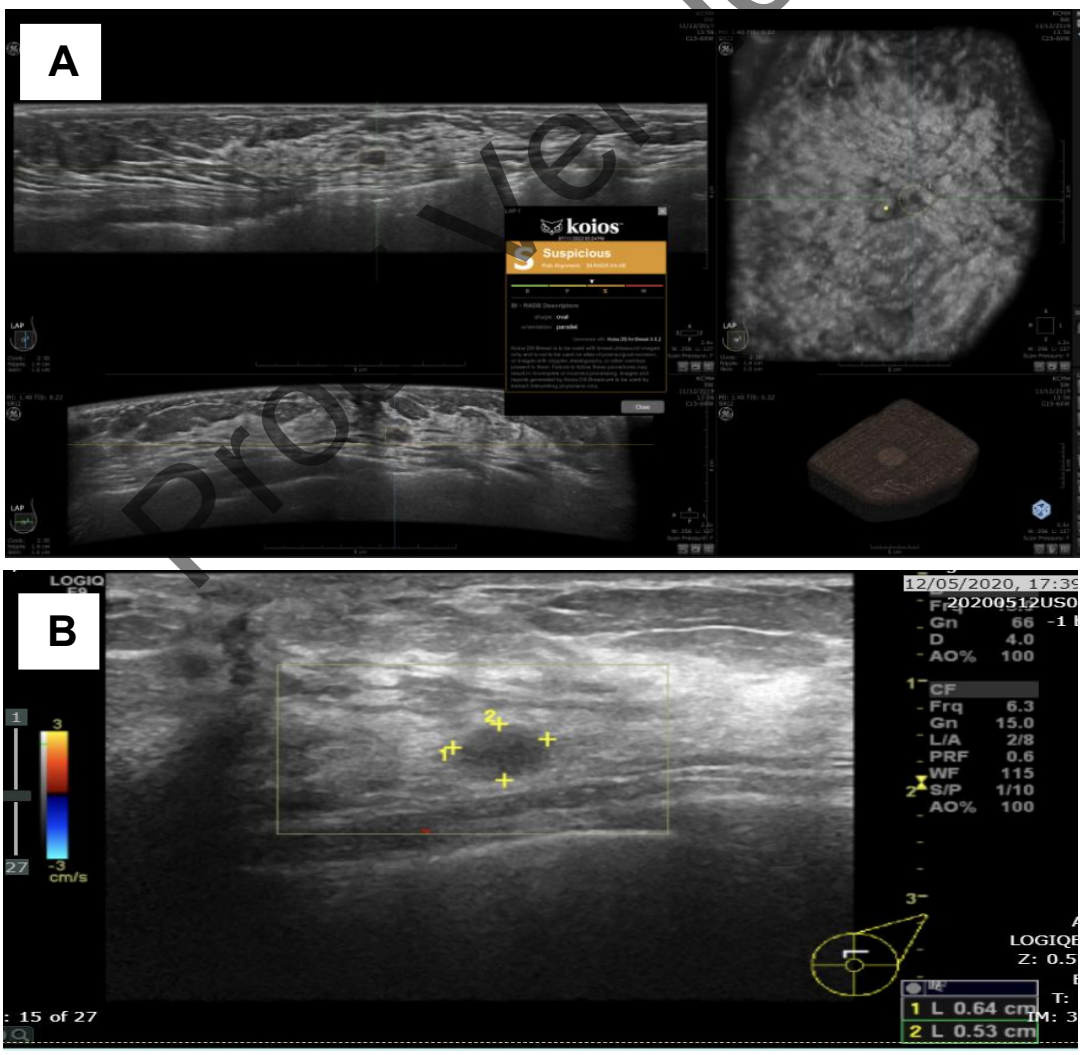

Figure 5. A 60-year-old female with invasive ductal carcinoma. She had a history of contralateral breast cancer. ABUS **(A)** demonstrates a circumscribed hypoechoic mass in the lower outer quadrant of left breast; BI-RADS 3. However, the AI decision support characterized this mass as suspicious. Subsequent 6-month follow-up with HHUS **(B)**; the mass developed partial indistinct border, BI-RADS 4a. ABUS, automated breast ultrasound; AI, artificial intelligence; BI-RADS, breast imaging-reporting and data system.

Table 1. Number of masses and percentage of malignancy in each category by the AI decision support.

Artificial intelligence decision support	Mass	Malignant	Malignant (%)
Benign (BI-RADS 2)	71	0	0.0
Probably benign (BI-RADS 3)	121	0	0.0
Suspicious (BI-RADS 4a and 4b)	214	12	5.6
Probably malignant (BI-RADS 4c and 5)	9	8	88.9

BI-RADS, breast imaging reporting and data systems.

Table 2. Number of masses and percentage of malignancy in each BI-RADS category by reader 1 and reader 2.

BI-RADS	Mass	Malignant	Malignant (%)	Mass	Malignant	Malignant (%)
	Reader 1			**Reader 1 with AI**		
2	92	1	1.1	139	1	0.7
3	225	2	0.9	196	2	1.0
4a	76	5	6.6	57	4	7.0
4b	18	8	44.4	16	7	43.8
4c	1	1	100.0	4	3	75.0
5	3	3	100.0	3	3	100.0
	Reader 2			**Reader 2 with AI**		
2	21	0	0.0	68	0	0.0
3	274	3	1.1	238	3	1.3
4a	97	4	4.1	88	4	4.6
4b	18	10	55.6	10	7	70.0
4c	3	1	33.3	9	4	44.4
5	2	2	100.0	2	2	100.0

AI, artificial intelligence; BI-RADS, breast imaging reporting and data systems.

Table 3. AUC, sensitivity, specificity, PPV and NPV of the AI decision support and two readers.

	AI	Reader 1	Reader 1 with AI	Reader 2	Reader 2 with AI
AUC	0.74	0.82	0.85	0.79	0.81
(95% CI)	(0.72–0.77)	(0.74–0.91)	(0.76–0.93)	(0.71–0.88)	(0.73–0.89)
Sensitivity (%)	100.0	85.0	85.0	85.0	85.0
Specificity (%)	48.6	79.5	84.1	73.9	76.7
PPV (%)	9.0	17.4	21.3	14.2	15.6
NPV	100.0%	99.1%	99.1%	99.0%	99.0%

AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

AI decision support findings

The AI decision support results are displayed in **Table 1**. None of the 71 benign masses (BI-RADS 2) and 121 probably benign masses (BI-RADS 3) were identified as malignant. Out of the 214 suspicious masses (BI-RADS 4a and 4b), 12 masses (5.6%) were confirmed as malignant. Only nine masses were categorized as probably malignant, with eight of them (88.9%) proven to be malignant via biopsy **(Figure 2).** Pathological results of these malignant masses revealed five invasive ductal carcinomas, one solid papillary carcinoma, one mucinous carcinoma, and one ductal carcinoma in situ, whereas surgical excision of the remaining mass revealed fibrocystic change with fibroadenomatous change **(Figure 3).**

The AI decision support results aligned with the malignant rates of BI-RADS 4a for the suspicious masses and BI-RADS 4c for the probably malignant masses. The AUC for the AI decision support system was 0.74 (95% CI 0.72–0.77).

Readers findings

The masses were classified into the BI-RADS categories by the two readers. Reader 1 exhibited a percentage of malignancy in BI-RADS 4a, 4b, 4c, and 5 that corresponded to the expected range for each of the respective BI-RADS categories. However, for reader 2, the malignancy rate was higher than the expected range for BI-RADS 4b and lower than the expected range for BI-RADS 4c **(Table 2).**

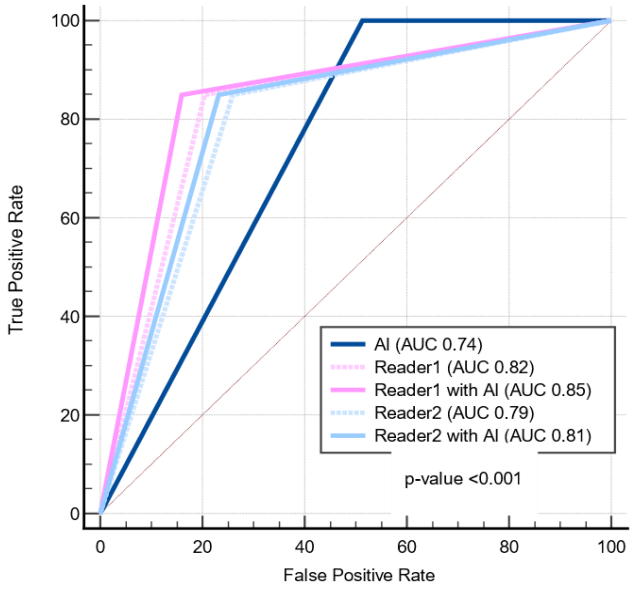

Figure 6. Area under the receiver operating curve of the AI and two readers. AI, artificial intelligence.

Table 4. Agreement between AI decision support and two readers.

Agreement (95% CI)	AI	Reader1	Reader1 with AI	Reader 2
AI	–	–	–	–
Reader1	0.12 (0.09–0.17)	–	–	–
Reader1 with AI	0.34 (0.29–0.37)	0.78 (0.74–0.82)	–	–
Reader2	0.18 (0.12–0.22)	0.41 (0.39–0.41)	0.32 (0.27–0.37)	–
Reader 2 with AI	0.42 (0.38–0.47)	0.31 (0.30–0.32)	0.50 (0.48–0.52)	0.70 (0.68–0.73)

AI, artificial intelligence; CI, confidence interval.

The AUC, sensitivity, specificity, PPV, and NPV of the two readers are presented in **Table 3** and **Figure 6**. The AUC for the AI decision support was calculated as 0.74 (95% CI 0.72–0.77). With the integration of AI decision support, the AUC for reader 1 significantly increased from 0.82 (95% CI 0.74–0.91) to 0.85 (95% CI 0.76–0.93) ($P < 0.001$). Similarly, the AUC for reader 2 significantly increased from 0.79 (95% CI 0.71–0.88) to 0.81 (95% CI 0.73–0.89) ($P < 0.001$). Moreover, no significant differences were observed in the AUC between AI and reader 1 ($P = 0.068$), AI and reader 2 ($P = 0.236$), and AI and reader 2 with AI decision support ($P = 0.129$). The AUC of reader 1 with AI decision support was significantly higher than AI decision support alone ($P < 0.05$). In addition, there was no significant difference in the AUC between reader 1 and reader 2 ($P = 0.46$), as well as between reader 1 with AI decision support and reader 2 with AI decision support ($P = 0.33$). The incorporation of AI decision support resulted in increased specificity, PPV, and NPV for both readers.

The incorporation of AI decision support led to a 14.2% and 16.9% alteration in BI-RADS for reader 1 and reader 2, respectively. For reader 1, 93.2% of these changes were attributed to a decreased BI-RADS classification, which consequently resulted in a 22.2% reduction in biopsies of benign masses (18 of 81 benign masses). In the case of reader 2, 88.6% of the changes were also because of decreased BI-RADS classification, resulting in a 10.7% reduction in biopsies of benign masses (11 of 103 masses) **(Figure 4).**

Agreement

The agreement between the AI decision support and two readers is shown in **Table 4**. Substantial agreement was observed when comparing the readings with and without AI decision support for both readers, with 0.78 (95% CI 0.74–0.82) for reader 1 and 0.70 (95% CI 0.68–0.73) for reader 2. There was also a moderate agreement between reader 1 and reader 2 at 0.41 (95% CI 0.39–0.41), as well as between reader 1 with AI decision support and reader 2 with AI decision support at 0.50 (95% CI 0.48–0.52).

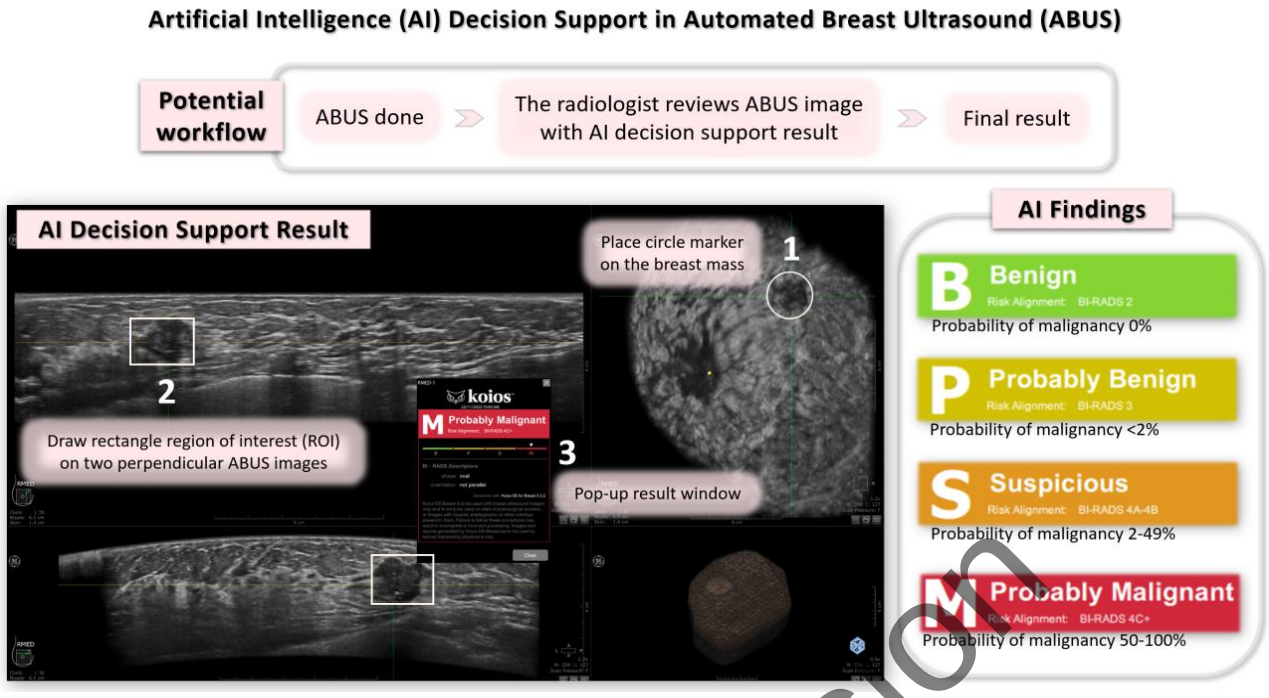

Figure 7. Workflow of AI-assisted decision support for lesion assessment in ABUS. ABUS, automated breast ultrasound; AI, artificial intelligence.

Discussion

The AI decision support demonstrated good diagnostic performance with an AUC comparable to that of the two readers without any significant differences. Notably, the AI decision support exhibited high sensitivity and a high NPV with a malignant percentage of the masses categorized as suspicious (BI-RADS 4a and 4b) and probably malignant (BI-RADS 4c and 5), which fit within the expected ACR BI-RADS range.

A total of four malignancies were missed by both readers, which were categorized as suspicious by the AI decision support. Both readers misinterpreted three malignancies as benign or probably benign. Furthermore, two of these four cases exhibited mammographic abnormalities, such as suspicious calcifications or architectural distortion, which could have assisted the readers in categorizing these lesions as BI-RADS 4–5 in routine practice. Another case was a small multifocal tumor that was considered suspicious by the AI decision support, and the last case was characterized by two readers as BI-RADS 3, but the AI decision support suggested a suspicious classification **(Figure 5).**

The incorporation of AI decision support in reading ABUS significantly increased the AUC for both readers, which aligns with a previous study that demonstrated a pooled AUC increase for readers with varied experience in breast imaging, increasing from 0.83 to 0.87.[5] Another study reported a significant increase in the AUC range of three radiologists from 0.73–0.76 to 0.79–0.83.[6] In addition, the agreement between the two readers increased from 0.406 to 0.497 with the aid of AI decision support, which is consistent with findings from a previous study.[5]

The percentage of change in BI-RADS category for both readers was similar, with rates of 14.2% and 16.9%. Furthermore, no significant difference was observed in the rate of BI-RADS category change with regard to radiologists' experience in a previous study.[11]

When comparing this study's diagnostic performance with that of other studies using Koios software, our study demonstrated a lower AUC at 0.74 compared to 0.77 (malignant rate 27.6%)[11] and 0.88 (malignant rate 47.8%).[5] This difference may be attributed to the lower malignancy rate in our population (4.8%), as it included screening and diagnostic studies. The AI decision support offered by other vendors, such as S-Detect or AI-Sonic, also demonstrated good diagnostic performance and helped decrease the biopsy rate of the benign mass **(Figure 7).**[12, 13]

Limitations of this study include that it was a retrospective study and was performed in a single center. Furthermore, there was a relatively small number of malignant masses. Future studies with an increased number of cases, including a malignant subgroup, will promote enhanced data analysis. This

study evaluated a single commercially available AI decision support system (Koios), which may limit the generalizability of the findings to other AI platforms with different training data, algorithms, and decision thresholds. Although previous studies of AI systems from other vendors have reported similar directional effects, differences in the study design, populations, and evaluation metrics prevent a direct numerical comparison.

Conclusion

The AI decision support demonstrates diagnostic performance comparable to that of radiologists, exhibiting high sensitivity and a high NPV. Integrating AI into the diagnostic workflow has the potential to enhance diagnostic performance for radiologists across various experience levels and contribute to the reduction in unnecessary biopsies of benign masses.

Author contributions

Both authors contributed substantially to the concept and design of this study, acquiring the data, reviewing the literature, and data analysis and interpretation. All authors approved the final version submitted for publication and take responsibility for statements made in the published article.

Acknowledgements

The authors would like to express their gratitude to Ms. Sirilak Uppathamphanon, Technician, Department of Radiology, King Chulalongkorn Memorial Hospital for imaging data organization.

Conflict of interest statement

All authors have completed and submitted the International Committee of Medical Journal Editors (ICMJE) Uniform Disclosure Form for Potential Conflicts of Interest. All authors declare that they have no conflicts of interest.

Data sharing statement

All data generated or analyzed in the present study are included in the published article. Further details are available for non-commercial purposes from the corresponding author upon reasonable request.

ORCID

Chayaporn Hasdiseve https://orcid.org/0009-0003-7953-6691
Jenjeera Prueksadee https://orcid.org/0000-0003-2288-3484

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74:229-63.
2. Mainiero MB, Lourenco A, Mahoney MC, Newell MS, Bailey L, Barke LD, et al. ACR Appropriateness Criteria® breast cancer screening. J Am Coll Radiol 2016;13:R45-R49.
3. Berg WA, Bandos AI, Mendelson EB, Lehrer D, Jong RA, Pisano ED. Ultrasound as the primary screening test for breast cancer: analysis from ACRIN 6666. J Natl Cancer Inst 2015;108:djv367.
4. Brem RF, Tabár L, Duffy SW, Inciardi MF, Guingrich JA, Hashimoto BE, et al. Assessing improvement in detection of breast cancer with three-dimensional automated breast ultrasound in women with dense breast tissue: the SomoInsight Study. Radiology 2015;274:663-73.
5. Mango VL, Sun M, Wynn RT, Ha R. Should we ignore, follow, or biopsy? Impact of artificial intelligence decision support on breast ultrasound lesion assessment. AJR Am J Roentgenol 2020;214:1445-52.
6. Barinov L, Jairaj A, Becker M, Seymour S, Lee E, Schram A, et al. Impact of data presentation on physician performance utilizing artificial intelligence–based computer-aided diagnosis and decision support systems. J Digit Imaging 2019;32:408-16.
7. Berg WA, López Aldrete AL, Jairaj A, Ledesma Parea JC, García CY, McClennan RC, et al. Toward AI-supported US triage of women with palpable breast lumps in a low-resource setting. Radiology 2023;307: e223351.
8. Browne JL, Pascual MÁ, Perez J, Salazar S, Valero B, Rodriguez I, et al. AI: can it make a difference to the predictive value of ultrasound breast biopsy? Diagnostics (Basel) 2023;13:811.
9. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. ACR BI-RADS® atlas : breast imaging reporting and data system. Reston, VA: American College of Radiology; 2013.
10. Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability studies: review and new results. Stat Methods Med Res 2004;13: 251-71.
11. Berg WA, Gur D, Bandos AI, Nair B, Gizienski TA, Tyma CS, et al. Impact of original and artificially improved artificial intelligence-based computer-aided diagnosis on breast US interpretation. J Breast Imaging 2021;3:301-11.
12. Lyu SY, Zhang Y, Zhang MW, Zhang BS, Gao LB, Bai LT, Wang J. Diagnostic value of artificial intelligence automatic detection systems for breast BI-RADS 4 nodules. World J Clin Cases 2022;10:518-27.
13. Wang X, Meng S. Diagnostic accuracy of S-Detect for breast cancer on ultrasonography: a meta-analysis (PRISMA). Medicine (Baltimore) 2022;101:e30359.