Original Article | Breast Imaging

eISSN 2005-8330
https://doi.org/10.3348/kjr.2023.0907
Korean J Radiol 2024;25(4):343-350

Korean Journal of Radiology
KJR

Positive Predictive Values of Abnormality Scores From a Commercial Artificial Intelligence-Based Computer-Aided Diagnosis for Mammography

Si Eun Lee, Hanpyo Hong, Eun-Kyung Kim

Department of Radiology, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Republic of Korea

Objective: Artificial intelligence-based computer-aided diagnosis (AI-CAD) is increasingly used in mammography. While the continuous scores of AI-CAD have been related to malignancy risk, the understanding of how to interpret and apply these scores remains limited. We investigated the positive predictive values (PPVs) of the abnormality scores generated by a deep learning-based commercial AI-CAD system and analyzed them in relation to clinical and radiological findings.

Materials and Methods: From March 2020 to May 2022, 656 breasts from 599 women (mean age 52.6 ± 11.5 years, including 0.6% [4/599] high-risk women) who underwent mammography and received positive AI-CAD results (Lunit Insight MMG, abnormality score ≥ 10) were retrospectively included in this study. Univariable and multivariable analyses were performed to evaluate the associations between the AI-CAD abnormality scores and clinical and radiological factors. The breasts were subdivided according to the abnormality scores into groups 1 (10–49), 2 (50–69), 3 (70–89), and 4 (90–100) using the optimal binning method. The PPVs were calculated for all breasts and subgroups.

Results: Diagnostic indications and positive imaging findings by radiologists were associated with higher abnormality scores in the multivariable regression analysis. The overall PPV of AI-CAD was 32.5% (213/656) for all breasts, including 213 breast cancers, 129 breasts with benign biopsy results, and 314 breasts with benign outcomes in the follow-up or diagnostic studies. In the screening mammography subgroup, the PPVs were 18.6% (58/312) overall and 5.1% (12/235), 29.0% (9/31), 57.9% (11/19), and 96.3% (26/27) for score groups 1, 2, 3, and 4, respectively. The PPVs were significantly higher in women with diagnostic indications (45.1% [155/344]), palpability (51.9% [149/287]), fatty breasts (61.2% [60/98]), and certain imaging findings (masses with or without calcifications and distortion).

Conclusion: PPV increased with increasing AI-CAD abnormality scores. The PPVs of AI-CAD satisfied the acceptable PPV range according to Breast Imaging-Reporting and Data System for screening mammography and were higher for diagnostic mammography.

Keywords: Breast neoplasms; Digital mammography; Computer-aided diagnosis; Artificial intelligence

See the invited Editorial "Caveats in Using Abnormality/Probability Scores from Artificial Intelligence Algorithms: Neither True Probability nor Level of Trustworthiness" at https://doi.org/10.3348/kjr.2024.0144.

INTRODUCTION

Mammography, the standard method for detecting breast cancer, has inherent limitations due to its two-dimensional projectional nature. The sensitivity varies from 60%–90% and is significantly affected by breast density [1,2]. Recently, artificial intelligence-based computer-

Received: July 24, 2023 **Revised:** November 17, 2023 **Accepted:** December 5, 2023
Corresponding author: Eun-Kyung Kim, MD, PhD, Department of Radiology, Yongin Severance Hospital, Yonsei University College of Medicine, 363 Dongbaekjukjeon-daero, Giheung-gu, Yongin 16995, Republic of Korea
• E-mail: ekkim@yuhs.ac

aided diagnosis (AI-CAD) has been increasingly integrated into mammography, displaying a diagnostic accuracy comparable to or even superior to that of radiologists, while significantly enhancing the diagnostic performance of radiologists [3-7].

Contrary to traditional CAD algorithms that rely on radiologist-determined features, AI-CAD algorithms built on deep learning networks do not elucidate how they arrive at their final scores or results using a calculation process expressed in continuous numbers [8-10]. The score generated by AI-CAD is generally accepted as the likelihood of cancer, and most commercially available AI-CAD applications offer a heatmap denoting abnormality scores. Nevertheless, the importance of the score itself remains ambiguous, such as the clinical implications of higher and lower scores, compared to the comparatively intuitive and straightforward Breast Imaging-Reporting and Data System (BI-RADS) by the American College of Radiology (ACR). BI-RADS suggests representative imaging findings for each category based on the corresponding positive predictive values (PPVs); therefore, we hypothesized that a comprehensive analysis of abnormality scores from AI-CAD in relation to PPVs would help radiologists and clinicians better understand the clinical significance of the abnormality scores.

In this study, we evaluated the factors associated with the abnormality scores generated by commercial AI-CAD system and the PPVs of these scores according to the clinical and radiological characteristics of mammograms.

MATERIALS AND METHODS

This retrospective study was approved by the Institutional Review Board of the Yongin Severance Hospital (IRB No. 9-2022-0118), and the requirement for informed consent was waived.

Study Population

Between March 2020 and May 2022, 10900 mammograms were performed at our institution. Among them, 798 breasts from 728 patients with abnormal AI results (abnormality score ≥ 10 generated by an AI-CAD explained below) in screening and diagnostic mammograms were enrolled. We excluded patients with a history of breast cancer.

The AI-CAD scores were provided for the left and right breasts; therefore, we applied inclusion and exclusion criteria for each breast. We included 342 breasts that underwent biopsy or surgery, 194 that were stable for at least 12 months, and 120 with BI-RADS scores of 1 or 2 on diagnostic ultrasound (US) and additional views. We excluded 117 breasts with incomplete assessment due to insufficient follow-up mammography in < 12 months, 22 breasts that underwent neoadjuvant chemotherapy, 2 breasts that developed interstitial mammoplasty, and 1 male breast. Finally, we included 656 breasts from 599 patients (mean age 52.6 ± 11.5 years) (Fig. 1), consisting of one mammogram from 569 patients and two mammograms from 30 patients. Based on the AI results, we included both the unilateral and bilateral breasts in these mammograms.

Out of 599 patients (mean age 52.6 ± 11.5 years), 61

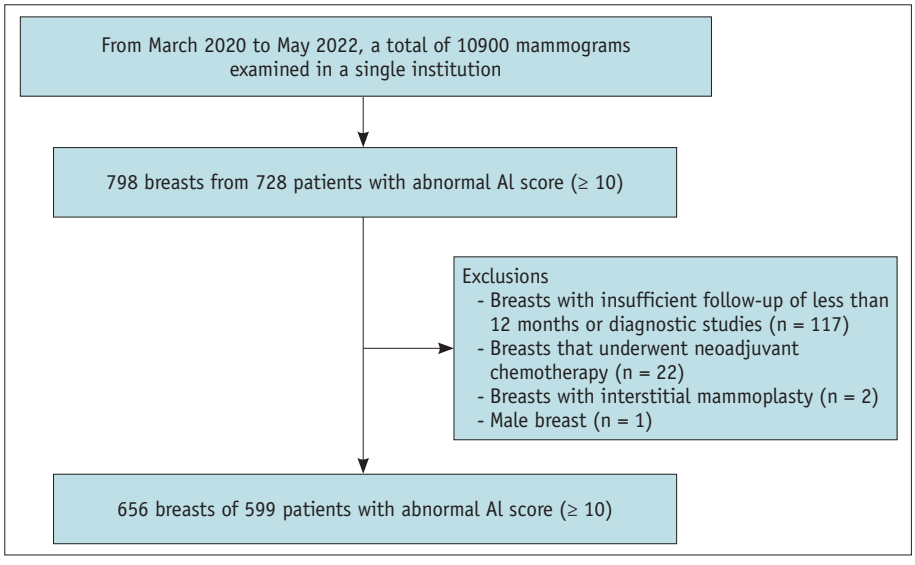

Fig. 1. Study population. AI = artificial intelligence

https://doi.org/10.3348/kjr.2023.0907

Korean Journal of Radiology
KJR

(10.2%) had a family history of breast cancer and 4 met the criteria for high risk for breast cancer, having two or more first-degree relatives diagnosed with breast cancer [11].

Image Analysis and AI-CAD Application

Mammograms were obtained using dedicated equipment (Pristina; GE Healthcare, Milwaukee, WI, USA). A senior radiologist (E-KK, with 24 years of experience in breast imaging) retrospectively reviewed the mammograms with AI-CAD results and recorded the mammographic findings correlating with the AI-detected area. Breast density was visually assessed based on the ACR BI-RADS 5th edition and mammographic findings were recorded in six groups: asymmetry, mass, mass with calcifications, calcifications only, distortions, and negative.

We used a deep learning-based commercial AI-CAD program (Lunit Insight MMG; https://insight.lunit.io, version 1.1.1.0 to 1.1.7.1) which was developed and validated through multinational studies [7,12,13]. In this program, the AI-CAD result is provided as two abnormality scores in percentages of 0%–100% per breast with a heatmap or grayscale map. An abnormality score of < 10 presents as "low" and is regarded as a test-negative result. Abnormality scores from 0 to 100, rounded to two decimal places, were obtained from the raw data.

Statistical Analysis

Data are presented as medians (interquartile ranges) or frequencies with percentages (%), as appropriate. The *P*-value was calculated using the Mann–Whitney U test and the Kruskal–Wallis test to compare the median values. Univariable and multivariable regression analyses were performed to explore the associations between the abnormality scores and clinical factors.

PPV was defined as the number of patients diagnosed as breast cancers per the number of patients whose mammograms got an abnormality score on AI-CAD more than 10. We employed the optimal binning method to categorize the continuous variables of the abnormality score into intervals. This procedure involves discretizing the scale variables by assigning their values to specific bins guided by a categorical variable. The optimal cutoff point was selected to maximize the difference in cancer prevalence among the score groups. The bins were categorized into four groups with optimal cutoff values of 50, 70, and 90. Groups 1, 2, 3, and 4 had the scores of 10–49, 50–69, 70–89, and 90–99, respectively. Chi-square and proportion tests were conducted to calculate *P*-values and determine the statistical significance of differences between the groups.

Statistical analyses were performed using the SPSS software (version 26; IBM Corp., Armonk, NY, USA) and R software (version 3.6.0; http://cran.r-project.org/). Statistical significance was set at a two-sided *P*-value < 0.05 was considered statistically significant.

RESULTS

Among the 656 breasts, 213 had malignant tumors, consisting of 186 invasive breast cancers, 25 ductal carcinomas in situ (DCIS), 1 malignant phyllodes tumor, and 1 mesenchymal tumor. Invasive cancers had higher median abnormality scores than DCIS (95 vs. 77, $P < 0.001$). The remaining 129 breasts had benign biopsy results, and 314 breasts were stable at follow-up or were benign in the diagnostic study. The distribution of AI-CAD scores for the benign and malignant groups is shown in Figure 2.

Nearly half of the mammography studies (312/656, 48%) were screened. Forty-four percent of the breasts (287/656) had subjective palpable symptoms, and 39% (255/656) underwent mammography with a metallic marker placed in the breast. Microcalcifications were the most common imaging findings (159/656, 24%), followed by masses (145/656, 22%), asymmetry (115/656, 18%), microcalcifications (66/656, 10%), and distortion (13/656, 2%). Twenty-four percent of breasts showed no apparent findings.

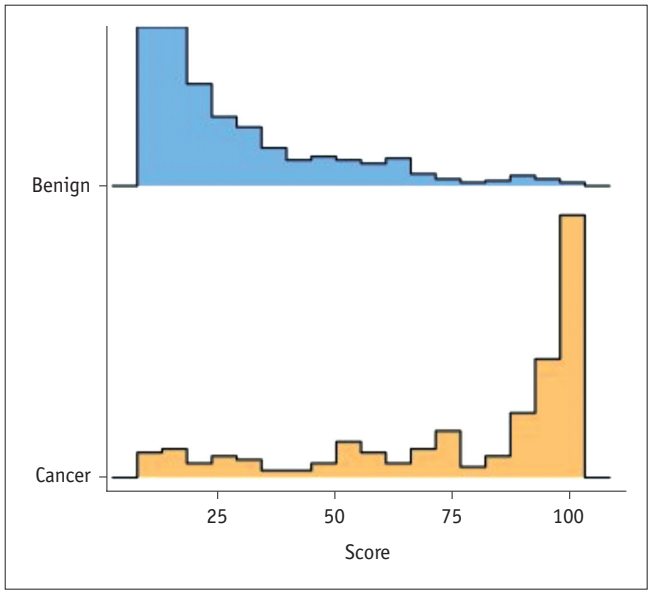

Fig. 2. Distribution of the artificial intelligence-based computer-aided diagnosis score in the benign and malignant groups.

https://doi.org/10.3348/kjr.2023.0907

Table 1. Median abnormality scores assigned by AI-CAD according to clinical and radiological characteristics of the 656 breasts

Variable	n (%)	Median score (IQR)	*P**
Indication			< 0.001
Screening	312 (48)	26 (15, 51)	
Diagnostic	344 (52)	52 (21, 95)	
Palpability			< 0.001
Yes	287 (44)	61 (27, 97)	
No	369 (56)	24 (15, 51)	
Density			< 0.001
Category A	16 (2)	61 (39, 99)	
Category B	82 (13)	71 (26, 98)	
Category C	347 (53)	32 (18, 76)	
Category D	211 (32)	29 (16, 56)	
Density (binary)			< 0.001
Fatty	98 (15)	70 (26, 99)	
Dense	558 (85)	31 (17, 66)	
Finding			< 0.001
Asymmetry	115 (18)	30 (17, 58)	
Distortion	13 (2)	63 (30, 70)	
Mass	145 (22)	74 (37, 97)	
Mass with microcalcifications	66 (10)	98 (61, 100)	
Microcalcifications only	159 (24)	33 (20, 61)	
Negative	158 (24)	18 (13, 27)	

**P*-values are for comparing the median scores and were calculated using the Mann-Whitney U test for indication, palpability, and density and using the Kruskal-Wallis test for density and finding.
AI-CAD = artificial intelligence-based computer-aided diagnosis, IQR = interquartile range

Factors associated with AI-CAD Scores

Diagnostic mammography showed higher median abnormality scores than screening mammography (52 vs. 26, respectively; $P < 0.001$). Palpable symptoms were also associated with higher median abnormality scores ($P < 0.001$). Compared to dense breasts, fatty breasts tended to have higher scores ($P < 0.001$). Among the positive imaging findings, masses with microcalcifications had the highest abnormality scores, followed by masses, distortions, asymmetry, and microcalcifications (Table 1; all $P < 0.001$).

In multivariable linear regression, diagnostic indications and all positive imaging findings deemed by a radiologist were significantly associated with abnormality scores (Table 2; all $P < 0.001$). From the standardized coefficients, masses with microcalcifications, followed by mass, distortion, microcalcifications, and asymmetry were highly associated with the abnormality scores. When we classified breasts as fatty and dense, the abnormality scores tended to increase in the fatty breasts ($P = 0.053$).

PPV of the AI-CAD Score

The overall PPV of AI-CAD was 32.5% (213/656). When we divided the AI-CAD scores into four score groups using the optimal binning method, the PPVs increased significantly as the scores increased ($P < 0.001$).

The overall PPV for screening mammography was 18.6%

Table 2. Association between the abnormality score and clinical and radiological factors on univariable and multivariable linear regression

Variable	Univariable			Multivariable			
	Unstandardized		*P**	Unstandardized		Standardized coefficients	*P**
	Coefficients	SE		Coefficients	SE		
Indication							
Screening	Ref			Ref			
Diagnostic	18.9	2.41	< 0.001	9.6	2.17	0.299	< 0.001
Palpability†							
No	Ref						
Yes	25.4	2.34	< 0.001				
Density							
Fatty	18.8	3.46	< 0.001	5.9	3.01	0.181	0.053
Dense	Ref			Ref			
Finding							
Asymmetry	17.7	3.25	< 0.001	15.1	3.25	0.468	< 0.001
Distortion	31.4	7.65	< 0.001	30.6	7.54	0.948	< 0.001
Mass	43.6	3.05	< 0.001	38.1	3.24	1.182	< 0.001
Mass with microcalcifications	56.6	3.89	< 0.001	52.0	3.95	1.612	< 0.001
Microcalcifications only	20.8	2.98	< 0.001	20.5	2.93	0.637	< 0.001
Negative	Ref			Ref			

**P*-values were calculated using multivariable linear regression, †Palpability was excluded from the multivariable analysis due to its large overlap with diagnostic indications.
SE = standard error, Ref = reference category

https://doi.org/10.3348/kjr.2023.0907

Korean Journal of Radiology
KJR

Table 3. PPVs for the four score groups according to clinical and radiological factors

Variable	Group 1 (10–49)	Group 2 (50–69)	Group 3 (70–89)	Group 4 (90–99)	Overall	*P**
Overall	9.3 (38/409)	33.8 (25/74)	66.0 (33/50)	95.1 (117/123)	32.5 (213/656)	< 0.001
Indication						
Screening	5.1 (12/235)	29.0 (9/31)	57.9 (11/19)	96.3 (26/27)	18.6 (58/312)	< 0.001
Diagnostic	14.9 (26/174)	37.2 (16/43)	71.0 (22/31)	94.8 (91/96)	45.1 (155/344)	< 0.001
P†	0.001	0.457	0.349	0.725	< 0.001	
Palpability						
Yes	16.0 (20/125)	47.1 (16/34)	69.7 (23/33)	94.7 (90/95)	51.9 (149/287)	< 0.001
No	6.3 (18/284)	22.5 (9/40)	58.8 (10/17)	96.4 (27/28)	17.3 (64/369)	< 0.001
P†	0.007	0.023	0.449	0.686	< 0.001	
Density						
Category A	16.7 (1/6)	75.0 (3/4)	100.0 (2/2)	100.0 (4/4)	62.5 (10/16)	0.038
Category B	20.0 (7/35)	66.7 (4/6)	77.8 (7/9)	100.0 (32/32)	61.0 (50/82)	< 0.001
Category C	7.3 (16/218)	27.5 (11/40)	58.6 (17/29)	90.0 (54/60)	28.2 (98/347)	< 0.001
Category D	9.3 (14/150)	29.2 (7/24)	70.0 (7/10)	100.0 (27/27)	26.1 (55/211)	< 0.001
P†	0.105	0.076	0.501	0.085	< 0.001	
Finding						
Asymmetry	13.9 (11/79)	31.6 (6/19)	28.6 (2/7)	80.0 (8/10)	23.5 (27/115)	< 0.001
Distortion	40.0 (2/5)	75.0 (3/4)	100.0 (2/2)	100.0 (2/2)	69.2 (9/13)	0.279
Mass	22.4 (11/49)	64.7 (11/17)	77.8 (21/27)	96.2 (50/52)	64.1 (93/145)	< 0.001
Mass with microcalcifications	35.7 (5/14)	40.0 (2/5)	75.0 (3/4)	97.7 (42/43)	78.8 (52/66)	< 0.001
Microcalcifications only	8.1 (9/111)	13.6 (3/22)	50.0 (5/10)	93.8 (15/16)	20.1 (32/159)	< 0.001
Negative	0.0 (0/151)	0.0 (0/7)	NA (0/0)	NA (0/0)	0.0 (0/158)	NA
P†	0.008	0.009	0.079	0.213	< 0.001	

Data are presented as a percentage of the number of breasts.
**P*-value calculated for the difference in PPV among score groups, †*P*-value calculated for difference in PPV between/among column categories in each variable (e.g., screening vs. diagnostic).
PPV = positive predictive value, NA = not applicable

(58/312), and the PPV for scores of 1, 2, 3, and 4 were 5.1% (12/235), 29.0% (9/31), 57.9% (11/19), and 96.3% (26/27), respectively (Table 3). The overall PPV for diagnostic mammography was 45.1% (155/344), which was much higher than that for screening indications (P < 0.001). To note, breasts with palpable symptoms showed a higher PPV of 51.9% (149/287) compared to those without symptoms that showed a PPV of 17.3% (64/369, P < 0.001). Except for group 4, the abnormal score groups showed higher PPV on diagnostic mammography and palpable breasts.

Fatty breasts showed a higher PPV of 61.2% (60/98) than dense breasts, with a PPV of 27.4% (153/558, P < 0.001). Among the positive imaging findings, masses with microcalcifications had the highest PPV (52/66, 78.8%), followed by distortions (69.2%, 9/13), masses (64.1%, 93/145), asymmetry (23.5%, 27/115), and microcalcifications (20.1%, 32/159) (Table 3). In contrast, 158 mammograms with negative imaging findings obtained by a radiologist had no cancer diagnosis.

Caution in Interpretation of the AI-CAD Score

The overall PPV of group 1 was 9.3% (38/409), which included 38 malignant cases, including 8 DCIS, 29 invasive ductal carcinoma, and 1 malignant phyllodes tumor (Fig. 3). Explainable imaging findings, in the order of PPV, distortion, mass with calcifications, mass, asymmetry, and microcalcification showed a PPV of at least 8.1% (9/111).

In contrast, 6 benign cases were observed in group 4 (Fig. 4). Four patients underwent a biopsy and were diagnosed with granulomatous mastitis, sclerosing adenosis, or acute/chronic inflammation. The other was a typically ruptured epidermal cyst on US, and the last was a grouped microcalcification that had been stable for more than 5 years.

DISCUSSION

We found that the PPV of AI-CAD increased serially, depending on the abnormality score. We present evidence indicating that the continuous nature of abnormality scores generated by AI-CAD for mammography is related

 https://doi.org/10.3348/kjr.2023.0907

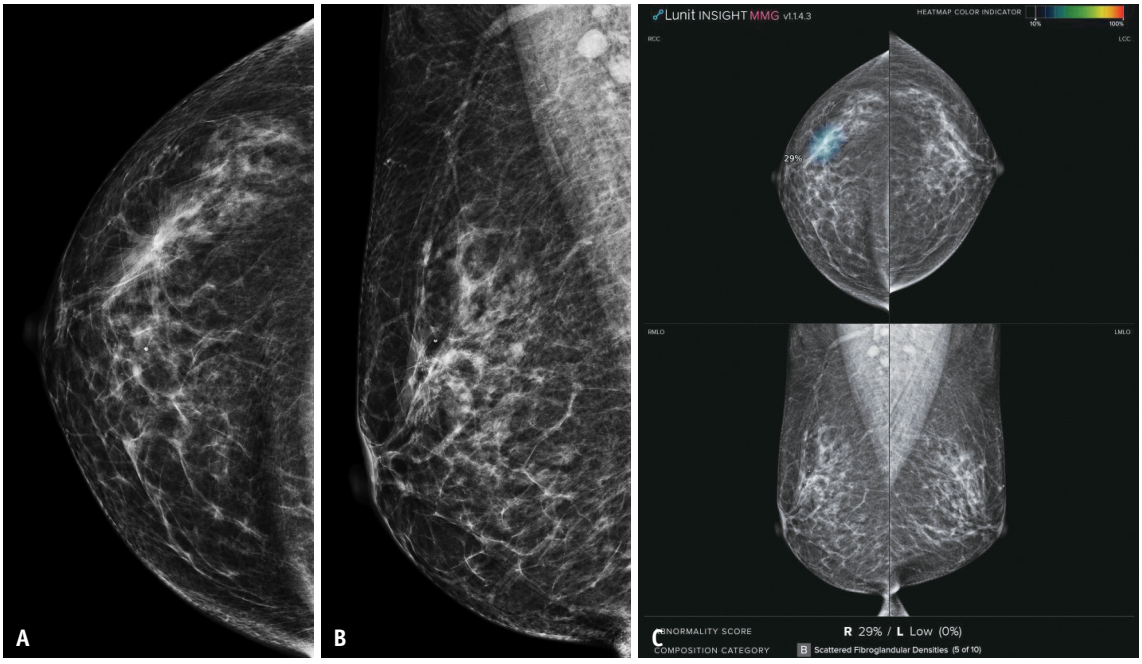

Fig. 3. 63-year-old woman visited for screening. **A, B:** Right mediolateral oblique **(A)** and right craniocaudal mammograms **(B)** show asymmetry with architectural distortion at right upper outer portion. **C:** Screenshot of output of AI tool. AI localized area in the right outer breast, as depicted by region of interest color map. Tool assigned abnormality score of 29% to the right breast and of "low" to the left breast. Ultrasound-guided core biopsy revealed invasive ductal carcinoma. Case represents malignant pathology with relatively low AI score. AI = artificial intelligence

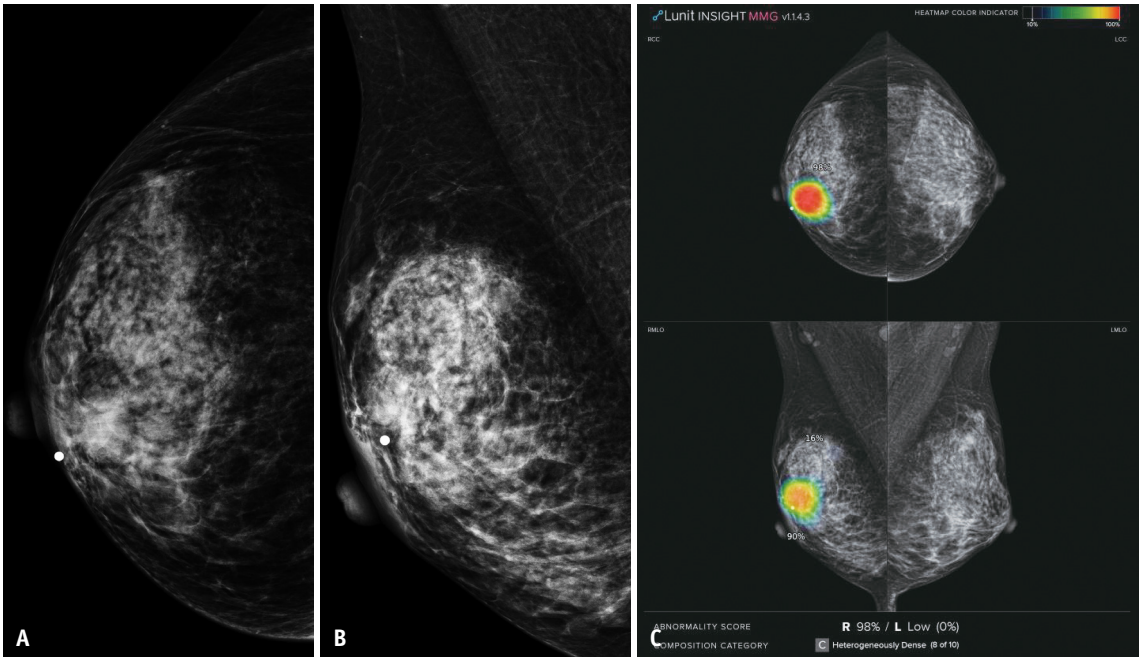

Fig. 4. 40-year-old woman visited due to right breast pain with lump. **A, B:** Right mediolateral oblique **(A)** and right craniocaudal mammograms **(B)** show mass opacity at right upper central portion with bb marker. **C:** Screenshot of output of AI tool. AI localized area in the right upper central breast, as depicted by region of interest color map. Tool assigned abnormality score of 98% to the right breast and of "low" to the left breast. Ultrasound-guided core biopsy revealed granulomatous mastitis. Case represents benign pathology with relatively high AI score. AI = artificial intelligence

https://doi.org/10.3348/kjr.2023.0907

Korean Journal of Radiology
KJR

to the likelihood of breast cancer. In addition, the PPVs were within the acceptable range of medical audit recommendations for the BI-RADS.

In multivariable regression, imaging findings, especially masses with microcalcifications, followed by masses, distortions, and microcalcifications affected the abnormality score the most. The diagnostic indications were also associated with higher abnormality scores. Since a large overlap existed between the diagnostic indications of examinations and palpable symptoms, we excluded palpability from the multivariable regression analysis. When breast density was classified into fatty and dense groups, the scores showed an increasing trend in fatty breasts. As expected, easily discernible findings in fatty backgrounds seemed to have higher scores for AI-CAD, which was also observed in a previous study [14].

For screening mammography, the overall PPV of AI-CAD diagnosis was 18.6%, which is between the recommended values for PPV_1 (3%–8%) and PPV_2 (20%–40%) in BI-RADS [15]. Although we could not evaluate the recall or abnormal interpretation rate of AI-CAD in our study population, it is known to be similar to that of radiologists in previous studies that analyzed historic cohorts [3]. When we divided the score groups into 1, 2, 3, and 4 with cutoff values calculated using the optimal binning method, the PPV were 5.1%, 29.0%, 57.9%, and 96.3%, respectively, which corresponded to the recommended PPV for BI-RADS 4a, 4b, 4c, and 5.

For diagnostic mammography, the overall PPV of AI-CAD diagnosis was 45.1% and 51.9% for women with palpable lumps, which was much higher than that of the screening population. This was in line with the BI-RADS recommendation of 15%–40% for PPV_2 in the case of diagnostic mammography and 25%–50% for PPV_2 for palpable lumps. Even in the lower score groups of 1 or 2, which suggest equivocal imaging findings, we could rely more on the AI score for diagnostic mammography or mammography in patients with palpable lesions.

Recent meta-analyses have reported that standalone AI-CAD showed a performance similar to or better than that of radiologists [3,16]. In our study, we showed that the scale of abnormality scores correlated well with the PPVs, and their values satisfied the BI-RADS recommendations. In addition, we confirmed that when a radiologist finds no explainable imaging findings for AI-CAD detection, the actual likelihood of a cancer diagnosis could be extremely low. We hope that these findings will provide evidence for standalone AI-CAD and contribute to our understanding of AI-CAD scores.

This study had some limitations. First, the use of a single AI-CAD software may limit the broader applicability of this study's findings to AI from other vendors and developers. Secondly, our study pertains to the calibration analysis of the "abnormality score," which is the probability generated by an algorithm-specific AI influenced by its training data. Our primary objective was to group these scores to provide a clear understanding of their significance. We discovered that our score groupings aligned with the BI-RADS recommendations. Although this result can be a useful reference for radiologists, its generalizability may have limitations. Additionally, although calibration is a critical aspect in interpreting AI results, a broader perspective should encompass the quantification of uncertainty. Uncertainty quantification measures the reliability of an AI prediction and is distinct from calibration. Unfortunately, for this study, we did not have access to algorithm-based data for uncertainty analysis. Future studies should delve further into calibration analysis and its implications to enhance the reliability of AI predictions in clinical decision making. Third, the dataset was collected retrospectively from a single institution and included only mammography with abnormal results identified by AI-CAD to focus on suspicious characteristics. Additionally, to include as many consecutive studies as possible, we used diagnostic study results from the same day, including US and an additional view of mammography, as a benign standard reference without pursuing additional follow-up. This approach may have led to an underestimation of cancer, with some cases possibly being overlooked. Finally, our study sample included all screening and diagnostic indications, and the screening population comprised only 48% of all mammography studies. The number of patients categorized by certain imaging findings, such as distortion, may not be sufficient to generalize the PPVs for each imaging finding.

In conclusion, the PPVs increased with increasing AI-CAD abnormality scores. The PPVs of AI-CAD were within the acceptable performance ranges suggested by the BI-RADS for screening mammography and were higher for diagnostic mammography.

Availability of Data and Material

The datasets generated or analyzed during the study are not publicly available since IRB approval is valid only for this research but are available from the corresponding author on reasonable request.

https://doi.org/10.3348/kjr.2023.0907

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Eun-Kyung Kim. Data curation: Si Eun Lee. Formal analysis: Hanpyo Hong. Funding acquisition: Si Eun Lee. Investigation: Si Eun Lee, Hanpyo Hong. Methodology: Eun-Kyung Kim, Si Eun Lee. Project administration: Si Eun Lee. Resources: Si Eun Lee. Supervision: Eun-Kyung Kim. Visualization: Si Eun Lee, Hanpyo Hong. Writing—original draft: Si Eun Lee. Writing—review & editing: all authors.

ORCID IDs

Si Eun Lee
https://orcid.org/0000-0002-3225-5484
Hanpyo Hong
https://orcid.org/0000-0002-0573-4527
Eun-Kyung Kim
https://orcid.org/0000-0002-3368-5013

Funding Statement

This work was supported by Medical AI Clinic Program through the National IT Industry Promotion Agency (NIPA), funded by the Ministry of Science and ICT (MSIT).

REFERENCES

1. Salim M, Dembrower K, Eklund M, Lindholm P, Strand F. Range of radiologist performance in a population-based screening cohort of 1 million digital mammography examinations. *Radiology* 2020;297:33-39
2. Checka CM, Chun JE, Schnabel FR, Lee J, Toth H. The relationship of mammographic density and age: implications for breast cancer screening. *AJR Am J Roentgenol* 2012;198:W292-W295
3. Yoon JH, Strand F, Baltzer PAT, Conant EF, Gilbert FJ, Lehman CD, et al. Standalone AI for breast cancer detection at screening digital mammography and digital breast tomosynthesis: a systematic review and meta-analysis. *Radiology* 2023;307:e222639
4. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89-94
5. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw Open* 2020;3:e200265
6. Lee JH, Kim KH, Lee EH, Ahn JS, Ryu JK, Park YM, et al. Improving the performance of radiologists using artificial intelligence-based detection support software for mammography: a multi-reader study. *Korean J Radiol* 2022;23:505-516
7. Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health* 2020;2:e138-e148
8. Gao Y, Geras KJ, Lewin AA, Moy L. New frontiers: an update on computer-aided diagnosis for breast imaging in the age of artificial intelligence. *AJR Am J Roentgenol* 2019;212:300-307
9. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep learning in radiology: does one size fit all? *J Am Coll Radiol* 2018;15(3 Pt B):521-526
10. Yoon JH, Kim EK. Deep learning-based artificial intelligence for mammography. *Korean J Radiol* 2021;22:1225-1239
11. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 2001;358:1389-1399
12. Kim EK, Kim HE, Han K, Kang BJ, Sohn YM, Woo OH, et al. Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. *Sci Rep* 2018;8:2762
13. Salim M, Wåhlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncol* 2020;6:1581-1588
14. Lee SE, Han K, Yoon JH, Youk JH, Kim EK. Depiction of breast cancers on digital mammograms by artificial intelligence-based computer-assisted diagnosis according to cancer characteristics. *Eur Radiol* 2022;32:7400-7408
15. 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
16. Hickman SE, Woitek R, Le EPV, Im YR, Mouritsen Luxhøj C, Aviles-Rivero AI, et al. Machine learning for workflow applications in screening mammography: systematic review and meta-analysis. *Radiology* 2022;302:88-104