

False Negative and False Positive Cases of AI in Mammography

You Me Kim

Department of Radiology, Donggunsan Hospital, Korea

Introduction

Breast cancer remains a significant global health issue, and early detection is pivotal for improving patient outcome (1-4). Breast cancer screening using mammography has been increasing continuously over the past decade and it places a substantial workload on breast radiologists (5-6). In recent years, many retrospective studies have indicated that employing artificial intelligence (AI) considerably reduced the workload of breast radiologists while improving screening performance (8-14). However, the reliability of such AI tools demands scrutiny, particularly in real-world clinical settings.

In mammography utilizing AI, diagnostic errors—false negatives (FNs) and false positives (FPs) are critical challenges that significantly impact diagnostic accuracy. Understanding these issues helps to discern both the strengths and limitations of AI tools in breast cancer detection and screening.

This review will focus on cases of FNs and FPs in mammography interpretation using Lunit AI system.

1. Overview of AI Mammography

AI algorithms, especially those using deep

learning, are increasingly used to interpret mammograms. Their goal is to improve early cancer detection by identifying patterns associated with malignancy, including microcalcifications, architectural distortions, and masses (15-19). However, despite their power, AI models are prone to errors, specifically FNs and FPs.

- False Negative (FN): AI misses a malignant lesion, reporting no abnormal findings despite cancer being present.
- False Positive (FP): AI identifies a suspicious lesion incorrectly flagged as malignant when it's actually benign.

2. False Negative Cases

FN is a serious error in mammography since it delays diagnosis and treatment, potentially leading to worse outcomes for patients.

(1) Causes of False Negatives

1) Dense Breast Tissue

- Dense breast tissue makes it challenging to differentiate between cancerous lesions and surrounding tissue. AI algorithms, like radiologists, struggle with reduced sensitivity in dense breasts (Fig. 1).
- Breast density is a major risk factor, and about 40-50% of Asian women have dense breasts,

통신저자: 김유미

Department of Radiology, Donggunsan Hospital, Korea
Tel. +82-63-440-0300 Fax. +82-63-440-0310
E-mail: sirenos@naver.com

increasing the chance of missed cancers.

2) Subtle or Inconspicuous Lesions

- Early-stage cancers such as ductal carcinoma in situ (DCIS) or small invasive tumors may not have clear visual markers. AI models trained primarily on obvious cancers might miss subtle findings (Fig. 2).
- Lesions near the pectoral muscle or behind the nipple may be hard to detect.

3) Algorithmic Bias and Training Gaps

- AI models may be trained on data from Western populations, limiting generalizability across ethnic groups (e.g., Asian populations).
- Models may fail to recognize atypical patterns of malignancy or rare forms like triple-negative breast cancer, which behave differently on imaging (Fig. 3).

4) Lack of comparison with prior images

- Although the AI program has shown excellent performance, there is currently no AI program capable of comparing results with prior images. Therefore, newly detected subtle findings may

be missed (Fig. 4).

(2) Clinical Impact of False Negatives

Missed cancers can significantly delay the initiation of critical treatments, such as surgery, chemotherapy, or radiation. In addition, FN cases may erode patient trust in screening programs, potentially discouraging future participation. Moreover, interval cancers—those diagnosed between scheduled screenings—are often more aggressive and typically discovered at more advanced stages, making them harder to treat effectively (19–21).

3. False Positive Cases

FP occurs when AI detects an abnormality that is ultimately benign. While not life-threatening, FPs can create psychological, financial, and clinical burdens.

(1) Causes of False Positives

1) Overinterpretation of Benign Lesions

- AI models trained for high sensitivity may flag fibroadenomas, cysts, or calcifications as

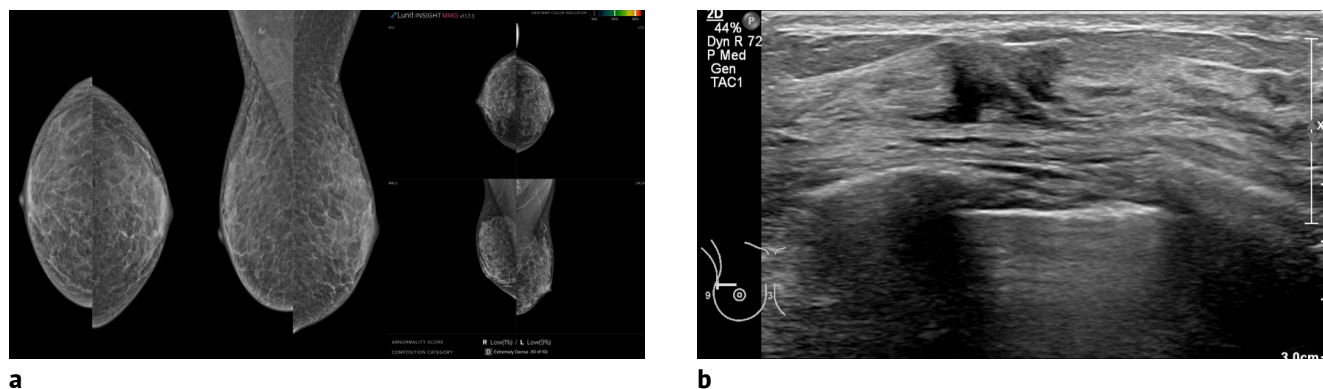
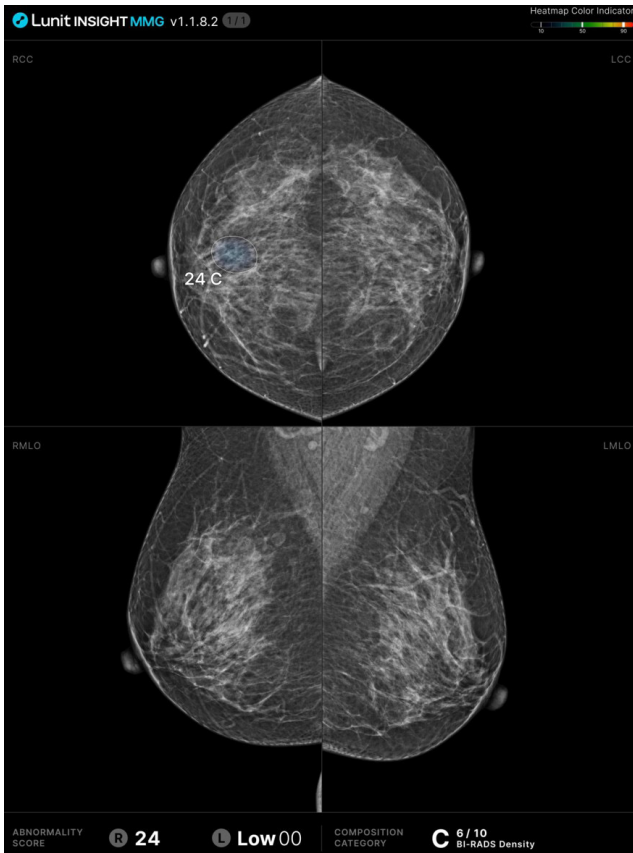
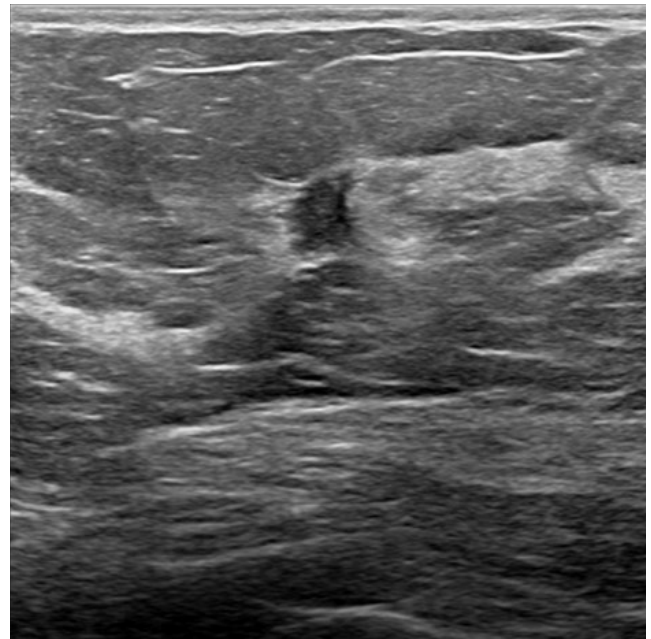


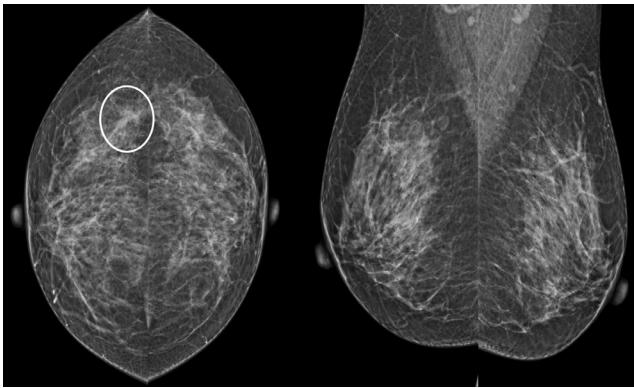
Fig. 1. False negative case in a 61-year-old woman who underwent screening with mammography and an artificial intelligence (AI) system. (a) Screening mammography and AI system show no abnormalities in dense breasts. (b) Screening ultrasound reveals a 11 mm invasive ductal carcinoma in the right breast at the 10 o'clock position.



a



c



b

Fig. 2. Concurrent false negative and false positive findings on AI system in 53-year-old woman. (a) Right craniocaudal (CC) view shows a low AI score of calcifications. The AI system did not detect other suspicious lesions. (b) Right CC view shows subtle asymmetry in the outer portion, identified by the radiologist (circle). (c) Ultrasound image reveals a 10 mm invasive ductal carcinoma observed during the diagnostic workup. No malignancy was found in the right retroareolar calcifications flagged by the AI system.

suspicious, increasing false positives (Fig. 2).

- This "over-call" behavior is designed to ensure minimal FN but increases the FP rate.

2) Benign Mimics of Malignancy

- Scarring, benign ductal hyperplasia, or intramammary lymph nodes can appear

suspicious on mammography. AI may lack the nuanced judgment radiologists develop through clinical experience.

3) Variability in AI Model Calibration

- Some AI models are tuned for higher sensitivity to avoid missing cancers but sacrifice

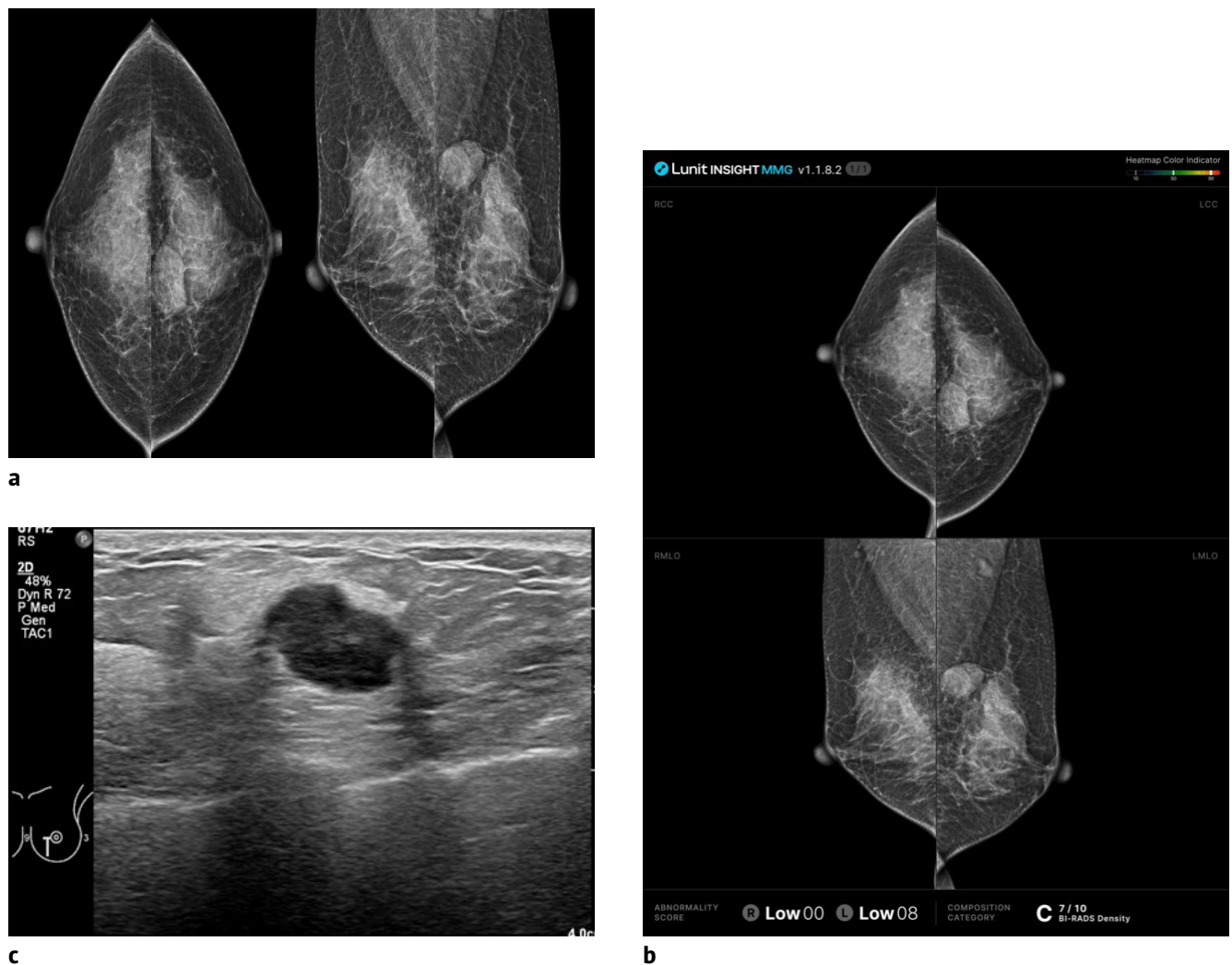


Fig. 3. False negative case in a 39-year-old woman presenting with a left breast mass, which the AI system interpreted as a benign tumor and did not flag. (a) Mammography shows a well-defined, oval-shaped hyperdense mass in the left lower inner breast. (b) The AI system did not detect any suspicious lesion. (c) Ultrasound images reveal a 29 mm invasive ductal carcinoma.

specificity. This is especially problematic in screening programs, where most women are healthy.

4) Lack of Contextual Interpretation

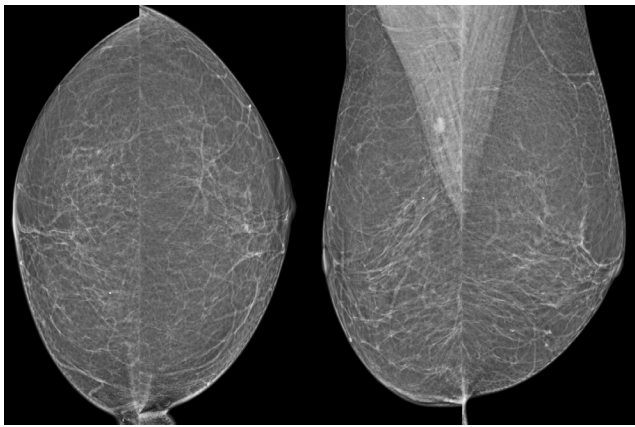
- AI models typically analyze static images and may not incorporate other relevant clinical data (e.g., family history, hormonal status). Radiologists, on the other hand, use such context to reduce unnecessary recalls.

5) Post-treatment changes or Scarring

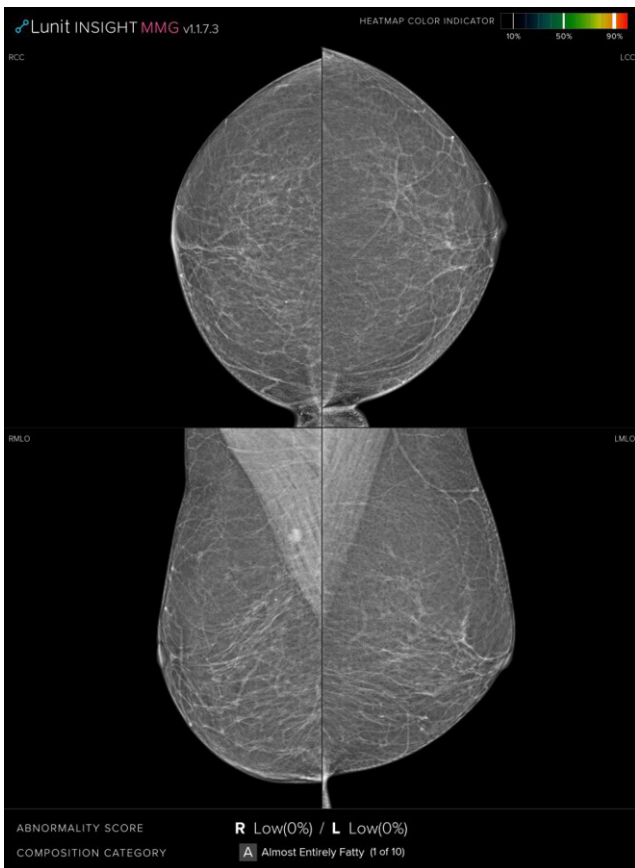
- Patients with prior surgery or radiation therapy might have scar tissue that resembles abnormal breast tissue, potentially confusing imaging system, including AI and leading to missed diagnoses of recurrent cancers (Fig. 5).

6) Lack of comparison with prior images

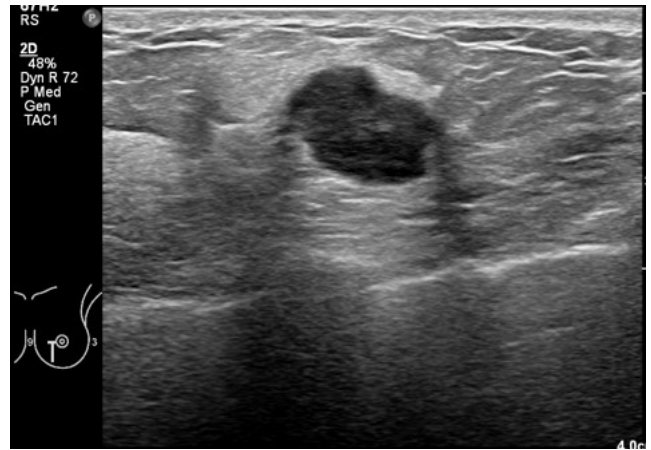
- For similar reasons seen in FN case, mass or asymmetry and calcifications that show no change over several years may be incorrectly



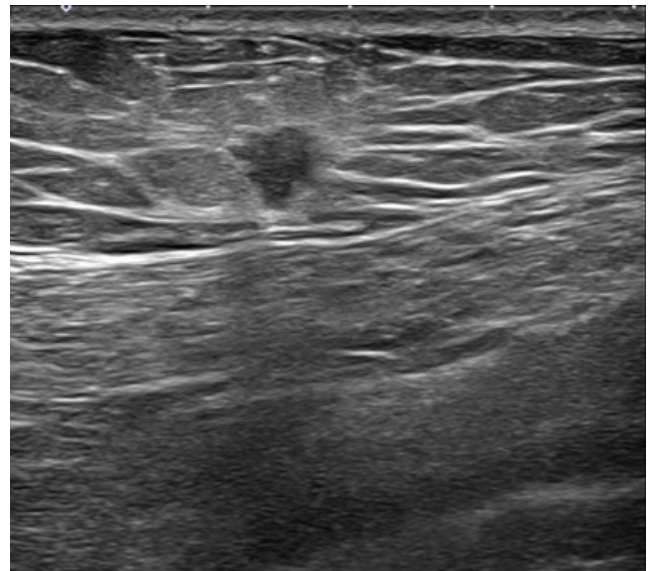
a



b



c



d

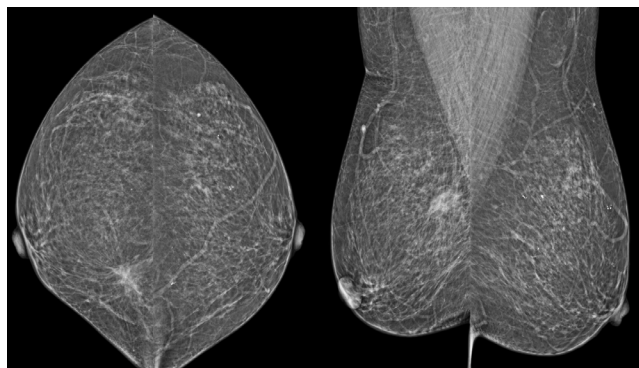
Fig. 4. False negative case in a 63-year-old woman, who underwent a right lumpectomy for invasive ductal carcinoma 6 years ago. (a) Mammography shows a small, high-density asymmetry in the right upper posterior area, overlapping the pectoralis muscle. (b) The AI system did not detect any suspicious lesion. (c) A prior view from 1 year earlier shows that the lesion was not visible at that time. (d) Ultrasound image shows an 8mm invasive ductal carcinoma.

marked as FPs (Fig. 6).

(2) Clinical Impact of False Positives

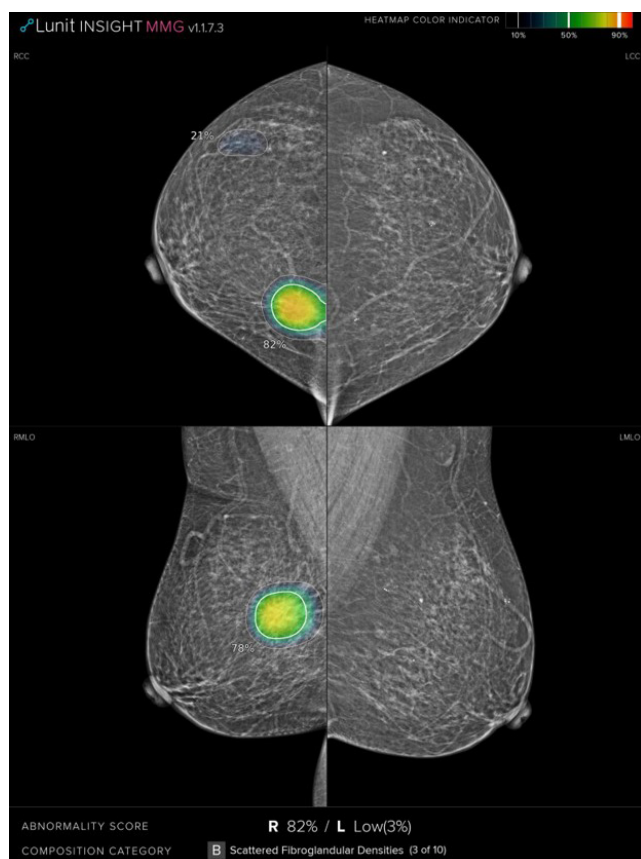
FP results can lead to significant patient anxiety, as women may experience unnecessary stress from additional imaging or biopsies. These follow-up

procedures, such as diagnostic mammography, ultrasound, or biopsy, also contribute to increased healthcare costs. Furthermore, FPs can result in overdiagnosis, where indolent or slow-growing cancers that may never cause harm are detected, potentially leading to overtreatment (22–24).

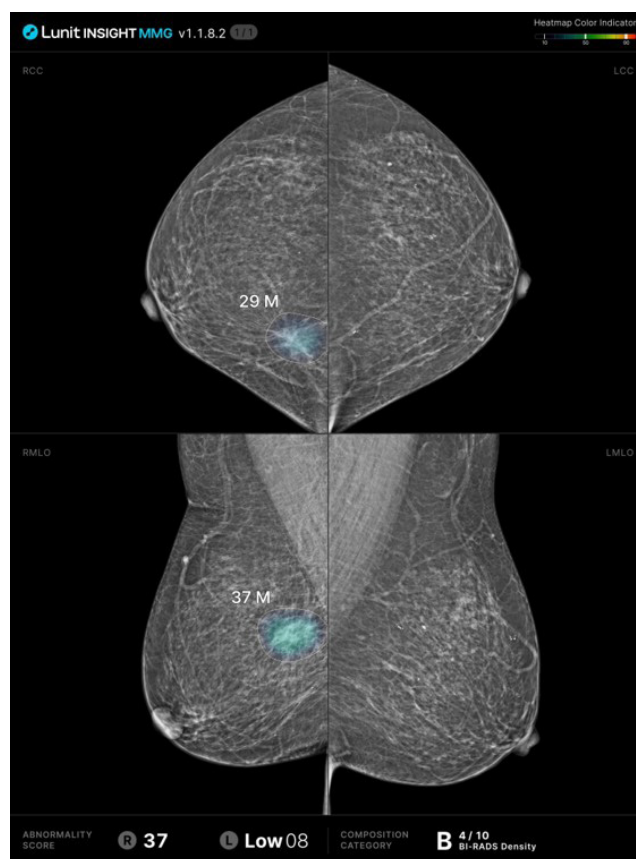


a

Fig. 5. False positive case in a 52-year-old woman who underwent a right lumpectomy 1 year ago. (a, b) Mammography shows postoperative changes in the right upper inner breast, and the AI system indicates a high probability of abnormality. (c) Follow-up mammography at 6 months shows a decreased extent of postoperative changes, and the AI system also shows a reduced probability of abnormality.



b



c

4. Balancing Sensitivity and Specificity in AI Models

AI systems must find a balance between sensitivity, which ensures that all cancers are detected, and specificity, which minimizes FPs. High sensitivity reduces the risk of missed cancers but may lead to a higher rate of FPs, while

high specificity decreases FPs but increases the likelihood of missing subtle cancers. To reduce errors, the use of multiple imaging modalities, such as breast ultrasound and MRI, can complement mammography, especially for patients with dense breasts or those at high risk (25-27). AI-based multi-modality systems that combine mammography, ultrasound, and MRI findings

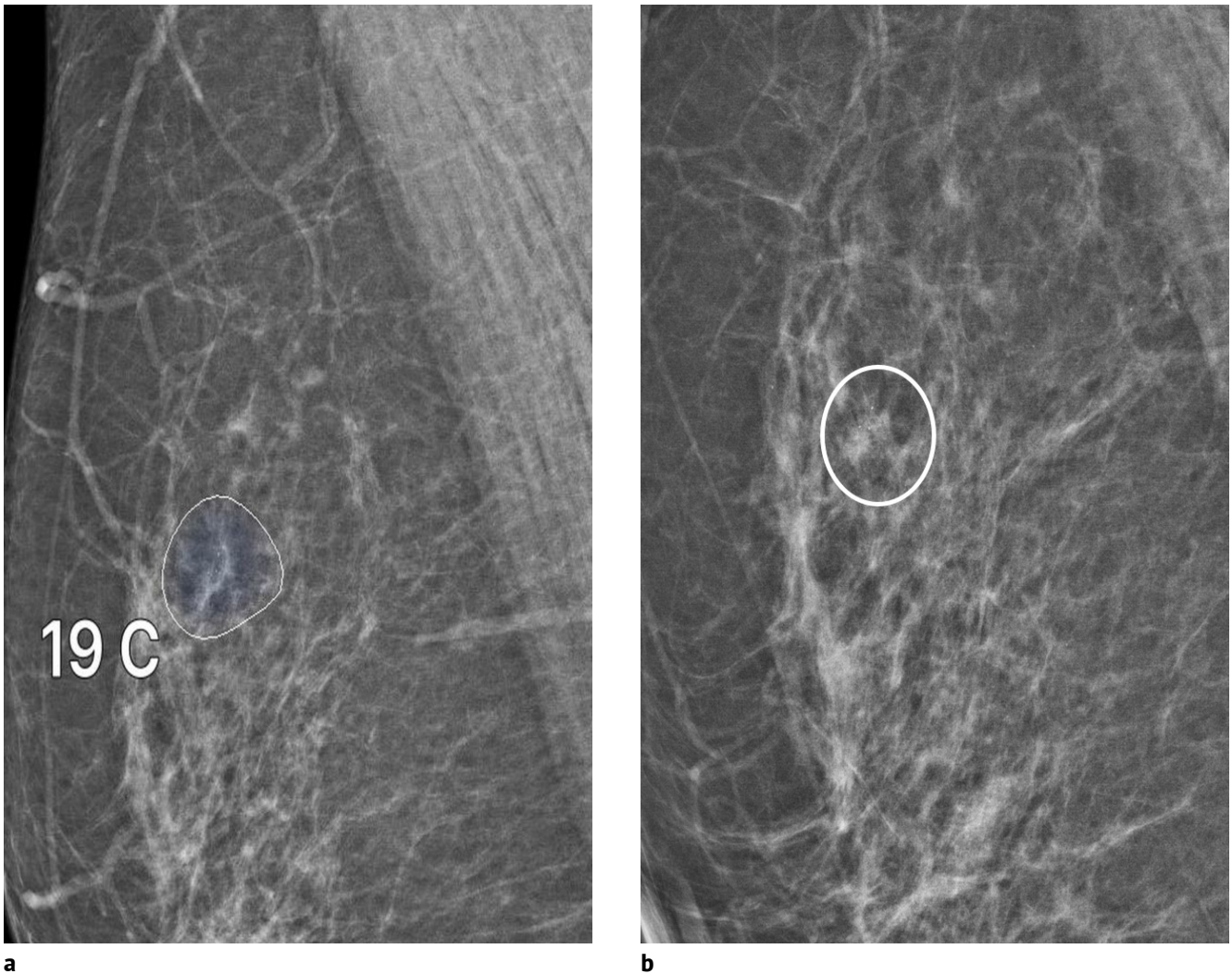


Fig. 6. False positive case in a 62-year-old woman who underwent a left mastectomy due to invasive ductal cancer. (a) The image shows AI-provided marking for microcalcifications in the right breast. (b) A prior image from 1 year earlier shows that the calcifications were noted (circle), with no interval change, confirmed by the radiologist.

show promise in lowering both FN and FP rates. Additionally, studies demonstrate that human-AI collaboration—where AI serves as a second reader, offering suggestions for radiologists to review and confirm—improves diagnostic accuracy (1,4,28–29). Radiologists can also provide critical clinical context, especially in complex cases with prior imaging histories, helping to address limitations in AI’s decision-making process.

5. Future Directions and Solutions

To enhance AI performance, it is essential to use more diverse datasets that represent a variety of populations, imaging technologies, and rare cancers. Additionally, continuous learning systems that adapt based on feedback from radiologists’ corrections can improve the accuracy of AI models over time. Personalized screening protocols, where AI models are tailored to specific populations—such as by breast density or ethnicity—could help reduce both FNs and FPs. Future AI systems may also

integrate clinical data, including genetic information and hormonal status, to provide more precise and individualized predictions. Finally, regular quality assurance programs, such as audits comparing AI performance with radiologists' interpretations, will be critical in identifying biases or blind spots in the models, ensuring ongoing improvement in diagnostic accuracy (13,14 30).

Conclusion

AI in mammography has the potential to enhance breast cancer detection, but it is not without limitations. False negatives and false positives pose significant challenges, each with different clinical and psychological impacts. An ideal approach combines AI tools with expert radiologist oversight, ensuring a balance between sensitivity and specificity. Continuous advancements in AI technology, along with diverse training datasets and multi-modality systems, will be key to minimizing these errors in the future.

References

1. Yala A, Barzilay R, Wang X, et al. A deep learning algorithm for breast cancer detection in mammography and ultrasound: An international validation study. *Lancet Digital Health* 2020;2(8).
2. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577(7788):89-94.
3. Lehman CD, Warshawsky R, Yau C, et al. Comparison of AI to Radiologists in Breast Cancer Screening Using Digital Mammography. *JAMA* 2020;323(9):858-866.
4. Lee C, Cirelli S, Geras KJ, et al. Artificial intelligence in breast imaging: A review of current applications and future directions. *Cancer Imaging* 2021;21(1):1-15.
5. Dembrower, K., Pölva, P., Zederfeldt, M., et al. AI-based mammography analysis and its impact on screening accuracy: A study of 18,000 cases. *Radiology* 2021;299(2):301-308.
6. Wootton, A., Sheth, V., Wark, J., et al. Evaluating AI performance for detection of breast cancer: A systematic review. *BMC Med Imaging* 2020;20(1):61.
7. Lauritzen AD, Lillholm M, Lynge E, et al. Early indicators of the impact of using AI in mammography screening for breast cancer. *Radiology* 2024;232(4):479-488.
8. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577(7788):89-94.
9. Yip S, Zhi Y, Yu G, et al. The role of AI in improving breast cancer screening: A retrospective analysis. *Breast Cancer Res Treat* 2021;178(2):403-411.
10. Liao X, Zhang J, Hu W, et al. The impact of AI on the reduction of false negatives in mammography. *Radiology* 2021;300(3):618-626.
11. Wang Y, Wang X, Lee J, et al. Application of AI for mammographic screening in high-risk patients: A comprehensive review. *Breast Cancer Res Treat* 2020;182(1):31-40.
12. Blake D, Moore M, Khan M, et al. An investigation into the causes of false positives in mammography and AI screening systems. *J Natl Cancer Inst* 2020;112(9):867-874.
13. Liu Y, Xiao W, Li X, et al. How AI impacts breast cancer detection and reduces errors: A meta-analysis. *Eur J Radiol* 2020;126:108930.
14. Dinesh B, Ramasamy A, Khan G, et al. Improving mammography interpretation using deep learning algorithms. *Radiology* 2020;298(1):213-222.
15. Manisha Bahl, Jay A. Baker, Emily N. Kinsey, and Sujata V. Ghate. Architectural Distortion on Mammography: Correlation With Pathologic Outcomes and Predictors of Malignancy. *Am J Roentgenology* 2015;205:1339-1345.
16. Wu, G., Zhang, Y., Xu, T., et al. Review of AI models for breast cancer detection in mammography. *Cancer Medicine* 2020;9(11):3797-3805.
17. Yala, A., Lehman, C., Schuster, T., et al. Deep learning and breast cancer screening performance: Reducing radiologist workload while improving diagnostic accuracy. *Breast Cancer Research* 2021;23(1):36.
18. Kooi, T., Litjens, G., van Ginneken, B., et al. Large-scale deep learning for computer-aided detection: Providing radiologists with precise lesion localization. *Radiology* 2017;285(2):286-295.
19. Wald, N. J., Chamberlain, J., Hackshaw, A. K.,

- & Frost, C. Reported Rates of Interval Cancers in Mammography Screening: A Meta-Analysis. *J Med Screen* 1997;4(1):17-21.
20. Houssami, N., Irwig, L., & Simpson, J. M. The nature and outcome of interval breast cancers in population mammographic screening. *The Lancet Oncology* 2006;7(1):58-66.
21. Gilliland, F. D., Joste, N., Stauber, P. M., et al. Biologic characteristics of interval and screen-detected breast cancers. *JAMA* 2000;284(8):887-893.
22. Houssami, N., Ciatto, S., Macaskill, P., et al. False negative and false positive mammography findings: Implications for AI screening. *Breast Cancer Res Treat* 2019;176(2):269-276.
23. Berry, D. A., Cronin, K. A., Plevritis, S. K., et al. The impact of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 2005;353(17):1784-1792.
24. Gøtzsche, P. C., & Jørgensen, K. J. Mammography screening: The impact of overdiagnosis and false positives. *J Natl Cancer Inst* 2013;105(9):707-712.
25. Kuhl, C. K., Schrading, S., Morakkabati-Spitz, N., et al. Breast cancer screening using dynamic contrast-enhanced MRI: Comparative effectiveness of mammography and ultrasound. *Eur Radiol* 2005;15(3):464-474.
26. Berg, W. A., Zhang, Z., Lehrer, D., et al. Mammography, ultrasound, and magnetic resonance imaging for breast cancer screening in women at elevated risk. *Ann Intern Med* 2012;156(11):601-610.
27. Sickles, E. A., & D' Orsi, C. J. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. American College of Radiology 2013.
28. Liu, Y., Santos-Magalhaes, N., Xie, L., et al. Radiologists vs artificial intelligence: Comparison of the accuracy of AI models in detecting breast cancer. *JAMA Network Open* 2021;4(12).
29. McKinney, S. M., Sieniek, M., Godbole, V., et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577(7788):89-94.
30. Zhang Y, Deng W, Xu J, et al. A deep learning algorithm to predict the risk of breast cancer in women. *Cancer Imaging* 2021;21(1):1-8.