Clinical Radiology 79 (2024) e885–e891

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

The efficacy of artificial intelligence (AI) in detecting interval cancers in the national screening program of a middle-income country

L. Çelik [a], E. Aribal [b,c,*]

[a] Maltepe University Hospital, Feyzullah cad 39, Maltepe, 34843, Istanbul, Turkey
[b] Acibadem University, School of Medicine, 34752, Istanbul, Turkey
[c] Acibadem Altunizade Hospital, Tophanelioglu cad 13, Altunizade, 34662, Istanbul, Turkey

ARTICLE INFORMATION

Article history:
Received 15 January 2024
Received in revised form
14 March 2024
Accepted 21 March 2024

AIM: We aimed to investigate the efficiency and accuracy of an artificial intelligence (AI) algorithm for detecting interval cancers in a middle-income country's national screening program.

MATERIAL AND METHODS: A total of 2,129,486 mammograms reported as BIRADS 1 and 2 were matched with the national cancer registry for interval cancers (IC). The IC group consisted of 442 cases, of which 36 were excluded due to having mammograms incompatible with the AI system. A control group of 446 women with two negative consequent mammograms was defined as time-proven normal and constituted the normal group. The cancer risk scores of both groups were determined from 1 to 10 with the AI system. The sensitivity and specificity values of the AI system were defined in terms of IC detection. The IC group was divided into subgroups with six-month intervals according to their time from screening to diagnosis: 0–6 months, 6–12 months, 12–18 months, and 18–24 months. The diagnostic performance of the AI system for all patients was evaluated using receiver operating characteristics (ROC) curve analysis. The diagnostic performance of the AI system for major and minor findings that expert readers determined was re-evaluated.

RESULTS: AI labeled 53% of ICs with the highest score of 10. The sensitivity of AI in detecting ICs was 53.7% and 38.5% at specificities of 90% and 95%, respectively. Area under the curve (AUC) of AI in detecting major signs was 0.93 (95% CI: 0.90–0.95) with a sensitivity of 81.6% and 72.4% at specificities of 90% and 95%, respectively (95% CI: 0.73–0.88 and 95% CI: 0.60–0.82 respectively) and minor signs was 0.87 (95% CI: 0.87–0.92) with a sensitivity of 70% and 53% at a specificity of 90% and 95%, respectively (95% CI: 0.65–0.82 and 95% CI: 0.52–0.71 respectively). In subgroup analysis for time to diagnosis, the AUC value of the AI system was higher in the 0–6 month period than in later periods.

CONCLUSION: This study showed the potential of AI in detecting ICs in initial mammograms and reducing human errors and undetected cancers.

* Guarantor and correspondent: E. Aribal.
E-mail addresses: leventcelik@hotmail.com (L. Çelik), earibal@gmail.com, erkin.aribal@acibadem.edu.tr (E. Aribal).

https://doi.org/10.1016/j.crad.2024.03.012
0009-9260/

Introduction

Breast cancer (BC) is globally known as the most frequent female cancer and one of the leading causes of cancer-related mortality.[1] Mortality rates are reported to decrease, particularly in Western countries where nationwide mammography screening programs were established, and modern treatment options were delivered despite increased incidence.[2,3] However, breast density becomes a significant drawback for mammography, seen in almost 50% of the screening population between ages 40 and 69 in various countries.[4,5] The sensitivity of mammography decreases by half in American College of Radiology (ACR) C and D-type breast tissue.[6]

The dense trial[7] showed the effect of MRI in women with extremely dense breast tissue (ACR D type), which consists of approximately 10% of the screening population, by a cancer detection rate of 16.5 per 1000 and decreasing the interval rate from 5 to 0.5 per 1000 screenings. This study confirms the hypothesis that some interval cancers are depictable during mammography screening. However, setting up a national program with supplemental MRI screening is challenging due to some drawbacks of breast MRI, such as the necessity of intravenous administration of gadolinium-based contrast media, the cost of the examination, the limited capacity of MRI centers, limited access to MRI-guided biopsy, claustrophobia, and having MRI-incompatible devices. Furthermore, interpretive errors cause almost one-fifth of nondetected cancers.[8] On the other hand, a lack of awareness results in low levels of compliance with the screening calls. It is of utmost importance to reach the highest possible detection rate at the screening and eliminate human errors, as the women may not comply with the next screening round or present with a more advanced cancer due to a lack of awareness.

Interval cancers are one of the benchmarks of mammography screening. An interval cancer is defined as detecting de novo breast cancer in a woman within the screening period.[9] The risk of not identifying these cancers that could potentially be detected through screening is more significant in countries where screening occurs every two years. As a result, patients may present with more advanced stages of cancer. The interval cancers observed exhibit greater biological significance and are characterized by larger tumor sizes.[10–12] It is important to note that countries with limited resources present more issues, namely a lack of awareness of women compared to higher-income countries and limited human resources for screening.[13–15] Moreover, it is noteworthy that in many countries with limited resources, the median age at which breast cancer is diagnosed tends to be ten years earlier than in Western countries.[16,17] This observation suggests that breast tissue in these regions may exhibit a higher density at the age of diagnosis.

Artificial intelligence (AI) is an algorithmic system that has been used in different branches of radiology. AI also becomes an adjunct method for diagnosis used in breast imaging, especially for BC. These systems have already been used in breast screening programs.[18,19] Recent studies showed the performance of AI algorithms in detecting equal or better than an average radiologist in a screening mammography setting and augmenting the radiologist's performance as a second reader.[18–20] Prospective studies also demonstrate that AI has the potential to replace one human reader, establishing noninferiority when compared to the combined performance of two human readers.[21,22] There are numerous reports on the efficacy of AI in reducing interval cancers.[23–28] These studies demonstrate promising results, suggesting that an AI detection algorithm can identify between 12% and 50% of these cancers. Success depends on the chosen threshold for AI and its implementation.[23–28]

Middle-income countries (MICs) are nations positioned between low- and high-income countries based on per capita incomes, as defined by the World Bank thresholds. As of July 1, 2023, this group encompasses 108 countries with a gross national income per capita ranging from $1,136 to $13,845. Collectively, MICs contribute approximately 30% to global gross domestic product (GDP) and constitute 75% of the world's population, encompassing 60% of the global impoverished population.[29]

In this study, we aimed to investigate the efficiency and accuracy of an AI algorithm for detecting interval cancers in a national screening program in a middle-income country.

Material and methods

This retrospective study was approved by the local ethics committee with the date 04/21/2021 and the number 2021-08/14.

The national screening program is set at 2-year intervals, and all women aged between 40 and 69 years are invited for screening. Every attendant is examined with standard two-view mammograms (craniocaudal (CC) and mediolateral oblique (MLO)). All images from nationwide screening centers are sent to the national picture archiving system, and two readers read all images in one central reading environment. The results of both readers are recorded as a BI-RADS score. Only three BI-RADS scores are used, namely BI-RADS 0, 1, and 2. BI-RADS 0 is assigned for recalls necessitating additional imaging, such as magnification views, spot mammograms, or ultrasound imaging for a mammographic finding, and the biopsy decision was made after this additional evaluation. All these cases were documented as recalls. However, readers have given BI-RADS 4 and 5 in very few evident cases, such as typical spiculated masses or casting calcifications where additional imaging was not required but the patient was recalled for biopsy. All the patient information and screening findings are recorded in the national screening database. The recall rate during the study period was 5.3%.

Recruitment

The national database for the screening program was filtered for all women screened between April 2016 and

March 2019 for the cases reported with BI-RADS scores of 1 and 2 by both readers. A total of 2.248.665 mammograms were reached, and of these, 2.129.486 women were recorded as BI-RADS 1 or 2 by each reader. These women were matched with the national cancer registry program for a de novo BC diagnosis within the 5th and 730th days after their screening. A total of 442 women diagnosed with BC within the given interval dates were reached. These cases were labeled as interval cancers (IC). The mammograms of these women were retrieved from the picture archiving and communication system (PACS) of the National Screening Program. Mammograms of the 36 women in the IC group were excluded because they were digitalized with a computerized radiography (CR) system and were incompatible with AI. Consequently, the IC group consisted of 406 patients. The flowchart of the study method is given in Fig 1. The time between each case's last screening and IC diagnosis was recorded as "time to diagnosis" (TTD). All IC cases were grouped with six-month periods of TTD, such as the first six months, 6–12 months, 12–18 months, and 18–24 months.

Another set of normal mammograms was created from the initial mammograms of randomly selected 446 women who did not have any BC-related diagnosis after two consecutive screening mammograms, which were reported as BI-RADS 1 or 2 by each of the two readers. Their data were also searched in the national cancer registry for BC diagnosis, resulting in negative.

Expert evaluation

Two expert readers with more than 30 years of experience in mammography retrospectively reevaluated the mammograms in the IC group with the knowledge of the interval cancer findings. They assessed these mammograms in consensus and grouped them into three categories as negative, with major and minor findings. Major findings were the evident findings as any mammography reader could identify at the location where BC was detected. In contrast, minor findings were the subtle findings that could be detected only with the knowledge of a clinical finding or in a diagnostic reading environment.

Artificial intelligence evaluation

The IC and control group mammograms were reevaluated by an AI system (Transpara™ ScreenPoint™ Medical v1.6). These lesions were combined for labeling suspicious lesions. Consequently, the AI system labeled cases 1 to 10 for malignancy risk. The accuracy of the data from the AI system was evaluated and compared to pathological findings.

Statistical analysis

The statistical analyses were performed with R statistical software, version 2021.09.0. Sensitivity and specificity

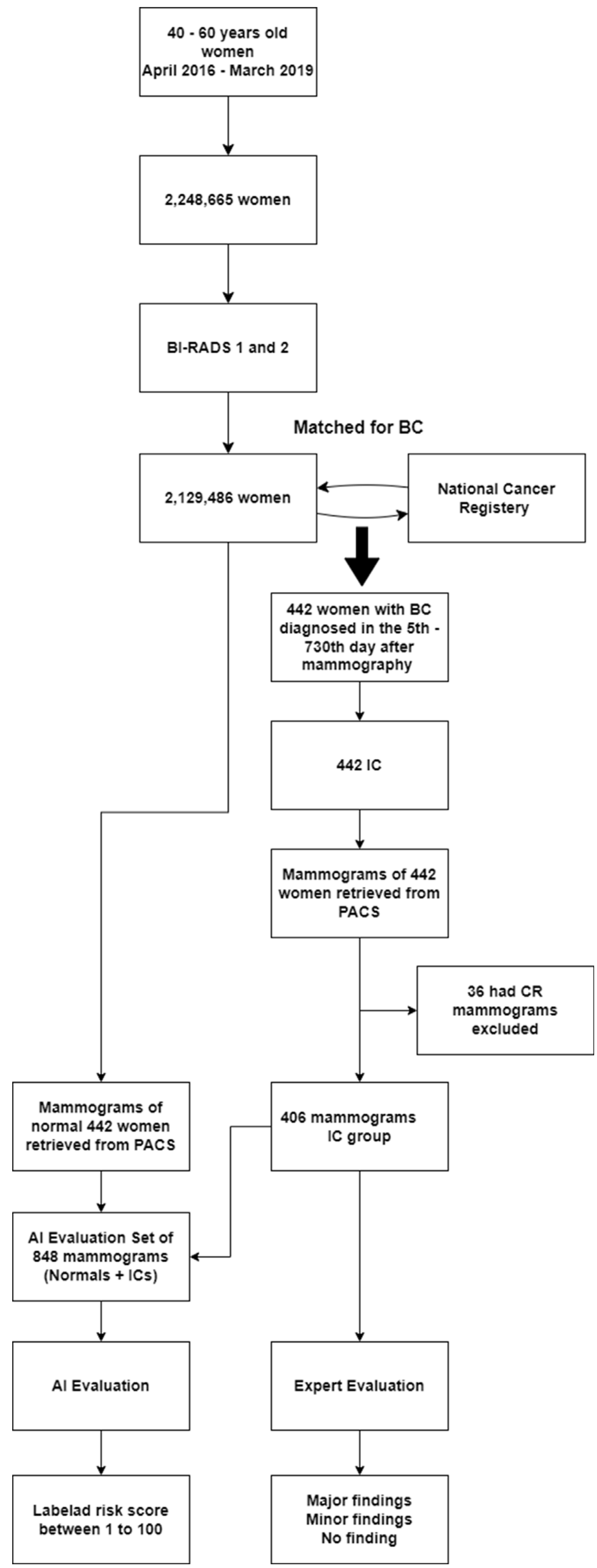

Figure 1 Flowchart of the study method. BC: Breast Cancer, CR: Computed Radiography IC: Interval Cancer.

values were determined according to pathologic diagnosis for the diagnostic performances of Transpara™ v1.6. Area under the curve (AUC) values were determined by the receiver operating characteristics (ROC) curve at a 95% confidence interval (CI) with specificities at 90% and 95%. Then, the cases were divided into 1–6 months, 6–12 months, 12–18 months, and 18–24 months according to the IC diagnosis times, and the diagnostic performance of the AI system was evaluated separately with the ROC curve. A p-value of 0.05 was accepted to be statistically significant.

Results

AI systems labeled 53% of ICs with a score of 10. The AUC of AI to detect signs of IC on negative screening exams was 0.80 (95% CI = 0.77–0.83) (Fig 2). The sensitivity of AI in detecting IC was 53.7% and 38.5% at specificities of 90% and 95%, repectively, corresponding to a recall rate of 10% and 5%.

An expert retrospective reading of the 406 confirmed cancers showed major findings in 109 cases (26.8%), minor findings in 116 cases (28.6%), and 'undetectable' findings in 181 cases. With the AI system, 92.7% of major findings were labeled with a score > 7, and 78.89% were labeled as 10. The AUC value of the AI system was 0.93 (95% CI: 0.90–0.95) for detecting major findings with a sensitivity of 81.6% and 72.4% at specificities of 90% and 95%, respectively (95% CI: 0.73–0.88 and 95% CI: 0.60–0.82, respectively) (Figs 3 and 4). On the other hand, 87.55% of both major and minor findings were labeled with a score > 7, and 70.22% were labeled as 10. The AUC value of the AI system was 0.90 (95% CI: 0.87–0.92) for detecting both major and minor findings with a sensitivity of 75.1% and 62.2% at a specificity of 90% and 95%, respectively (95% CI: 0.65–0.82 and 95% CI: 0.52–0.71, respectively). Analysis including minor findings only showed that 20.6% of minor findings were labeled with a score > 7. 61.5% were labeled as 10. The AUC value of the AI system was 0.87 (95% CI: 0.87–0.92) for detecting only minor findings with a sensitivity of 70% and 53% at a specificity of 90% and 95%, respectively (95% CI: 0.65–0.82 and 95% CI: 0.52–0.71, respectively) (Fig 3 and 4).

AI's ability to detect ICs at various intervals after mammography screening was evaluated by dividing the 24 months between two screenings into four six-month intervals. AUC values of the AI system according to the time of IC diagnosis were found to be 0.88 for 0–6 months, 0.76 for 6–12 months, 0.81 for 12–18 months, and 0.74 for 18–24 months, respectively (Fig 5).

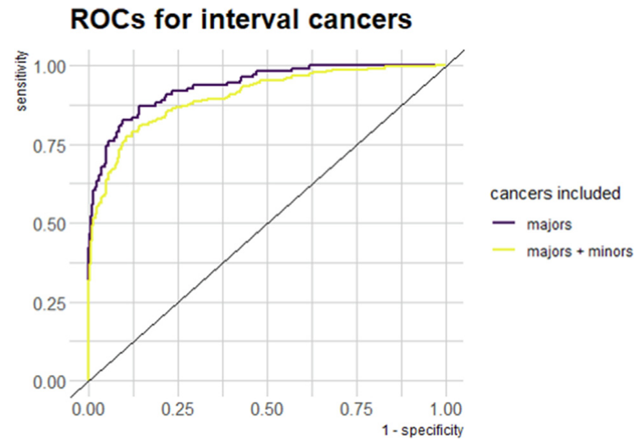

Figure 3 Area under the curve ROC analysis for AI performance in detecting the interval cancers that were labeled with major findings and minor findings regarding to expert retrospective reading. AI, artificial intelligence; ROC, receiver operating characteristics.

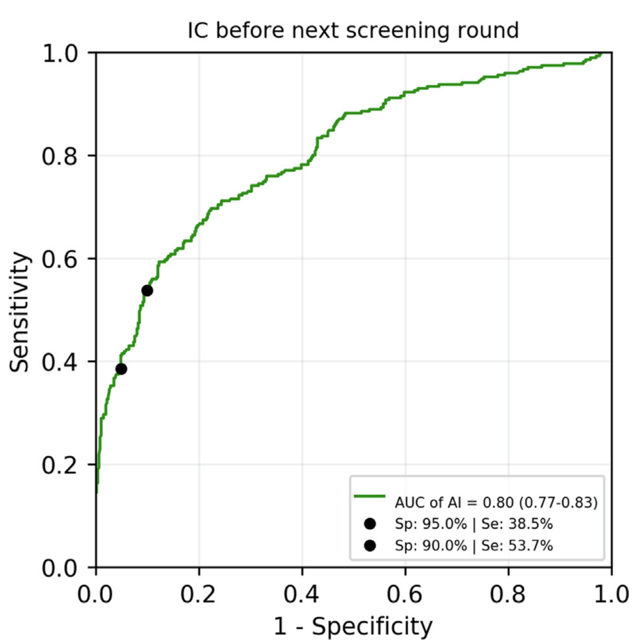

Figure 2 Area under the curve ROC analysis for AI performance in detecting interval cancers. AI, artificial intelligence; ROC, receiver operating characteristics.

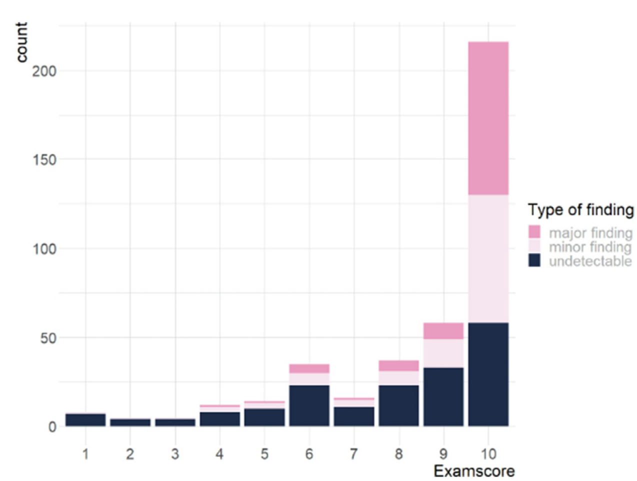

Figure 4 Scores given by AI for the interval cancers that were labeled with major findings and minor findings regarding to expert retrospective reading. AI, artificial intelligence.

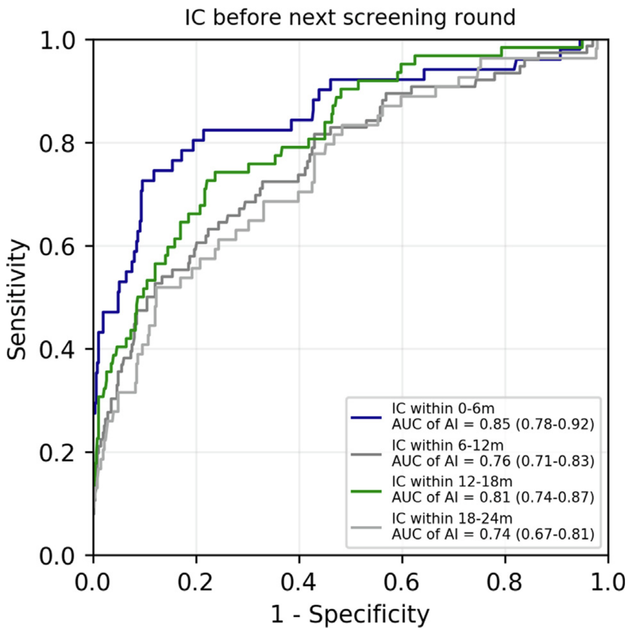

Figure 5 Area under the curve ROC analysis for AI performance in detecting the interval cancers for different 6 month periods of the 24 month interval time. AI, artificial intelligence; ROC, receiver operating characteristics.

Discussion

The AI system showed high sensitivity and specificity in detecting IC in screening mammograms and was able to flag more than half of the cases (53%), with a score of 10, which translates to a recall rate of 10%. The expert reading showed that more than one-fourth (26.8%) of these ICs had major findings that both readers missed. AI detected 78.9% of these cases with a score of 10. In line with our results, recent studies showed that nearly 80% of undetected cancers were missed due to the superposition of the adjacent fibroglandular tissue, where these cancers showed subtle nonspecific findings or no findings at all, and the yield of interpretive errors was reported to be approximately 20%.[8,30–32] Missed cancers, considered preventable human errors, can be mitigated through double reading. However, studies show that missed cancers with particular major findings are still being recorded.[8,30–32] This study showed that the AI system could have prevented 78.9% of missed cancers with major malignancy findings that both readers overlooked.

On the other hand, 28.6% of interval cancers showed minor findings that were unlikely to be detected during screening, as their recall would raise the false-positive rate beyond the recommended European threshold of 5%. Despite the low likelihood of detecting minor findings in screening, the AI algorithm identified 61.5% with a score of 10, corresponding to 17.6% of all interval cancers. Furthermore, the AI algorithm flagged 53 (44.9%) out of 118 (13.1% of all ICs) undepictable interval cancers with a score of 10 that did not exhibit any signs on expert readings and were recorded as undetectable. This indicates that AI can identify almost half of the interval cancers that are either obscured by adjacent fibroglandular tissue or undetectable. Similar studies with different cohorts showed AI's ability to detect IC with minor findings at 21–49% with specificities over 90%.[20,24,25,33,34] However, using AI algorithms to detect undepictable cancers carries the risk of increasing false-positive results. This study showed a high detection rate of interval cancers with a score of 10, corresponding to 10% recall, which would be a high false positivity rate for a national screening program. It is shown that the cancer detection rate of a score of 10 is 44 per 1000 screenings, implying that approximately 95% of these are false positives.[34] Moreover, the score 10 risk group has been reported to have the highest false positivity rate among all scores.[35] Radiologists may find it difficult to distinguish between such cases, and there could be a bias towards cases flagged by AI, which may lead to higher recall rates and false positives. In this case, there is a need to create a strategy for implementing AI in the reading setting. However, employing various thresholds can reduce false positivity and recall rates while effectively detecting interval cancers.[25,28] Establishing an optimal balance between false positivity and cancer detection rates is crucial. Given the diversity of thresholds among different AI algorithms, thorough research is essential before integrating AI into screening practices. A reading strategy study based on 122,969 mammograms concluded that different scenarios where AI is implemented in the reading process as a triage method, second reader, or decision maker showed a potential to reduce the workload of radiologists with a similar cancer detection rate.[36] The study also demonstrated a potential decrease in ICs and false-positive rates.[36] In two other studies employing AI for triaging negative cases, one reported a detection rate of 12% to 27% of interval cancers, while the second study demonstrated an even higher rate of 51.7% of interval cancers detectable by AI algorithms.[26,28] We believe it is necessary to conduct additional real-life studies to determine the appropriate approach for utilizing the suitable AI algorithm in national screening programs. A recent real-life study, adding AI as a second reader, demonstrated the noninferiority of AI to two human readers. However, further studies on different AI implementation strategies are still needed.[21]

Our study showed a higher rate of ICs with major findings (26.8%), which can be labeled as missed cancers or false negatives, compared to European guidelines recommendations (<20%).[9] A retrospective study with 231 ICs revealed significant variations in the missed cancer rate based on the review design of mammograms that ranged from 19.9% to 35.9%.[31] However, one of the reading designs they utilized that closely resembled our retrospective analysis strategy resulted in a missed cancer rate of 33.8%.[31] This may explain our comparatively higher missed cancer rate. AI could potentially reduce these human errors by 78.9%. Reading screening mammograms in a national screening program is an important challenge for human resources. Countries with limited resources appear to face a greater challenge in this regard and need more quality in their mammography practices.[13,14] Enhancing the efficacy of mammography screening by mitigating human errors is a significant improvement. AI can improve screening mammogram outcomes, particularly in countries with limited resources, by reducing these human errors.

The National Screening Program interval was two years. The AUC of AI in detecting ICs in different time frames after mammography screening to time to diagnosis was 0.88 in the first six months and 0.74 between the 18th and 24th months. This high AUC finding for cancers detected after six months showed the efficacy of AI even in detecting clinically late-presenting cancers. However, there is little data regarding the characteristics of these cancers that are detectable on previous mammograms, particularly regarding their biological significance. Although ICs are reported to be biologically more hazardous than screen-detected cancers, a recent study analyzing prior mammograms of screen-detected cancers showed that the cancers that could have been identified on prior mammograms, either with major or minor findings, had a more favorable histology compared to the cancers that were not identified on prior mammograms.[10–12,37] The lack of robust data on the biology of interval cancers (ICs) that could have been detected during the initial screening is a significant concern, particularly given that AI benefits these women through early detection at the expense of a possible higher recall rate.

There are several limitations of this study. First, the findings are for a single AI and cannot be generalized to all AI systems. Although AI can reduce human errors and detect ICs with minimal or no signs, there is no robust evidence of implementing this system into readings to achieve the best performance without increasing the false positives. Second, the AI algorithm utilized for this study has an upgraded version, which may show a higher detection rate with lower false-positive results. Third, this study was conducted with the xxxxxx national screening mammograms, where the start age of screening is 40 and the median age for cancers (51 years)[5] is below the median age of cancers in Europe, which is one decade later. Countries with limited resources and such a younger population may result in more mammograms with dense tissue and benefit more from AI systems. Fourth, the expert readers were not blinded to the findings and were informed about the tumor's diagnosis and location. Fifth, the AI evaluation was not based on lesion-level detection.

Conclusion

This study showed the efficacy of AI in detecting ICs in initial mammograms and reducing human errors. AI has the potential to enhance screening mammogram outcomes, particularly in countries with limited resources, by reducing human errors and detecting more undepictable cancers, at the expense of a possible higher recall rate. Further studies are needed to optimize the combination of AI and human reading.

Author contribution

1. Guarantor of integrity of the entire study: Both authors contributed equally.
2. Study concepts and design: LC 70%, EA 30%.
3. Literature research: EA 80%, LC 20%.
4. Clinical studies: Both authors contributed equally.
5. Experimental studies/data analysis: LC60%, EA 40%.
6. Statistical analysis: Both authors contributed equally.
7. Manuscript preparation: EA 80%, LC20% 8. Manuscript editing: EA60%, LC 40%.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgment

We want to thank the Ministry of Health and Suayip Birinci for facilitating our access to the National Screening Database and enabling us to retrieve the data, and Chief Software Architect at National Teleradiology System and Yaşar Alper Ozkaya to retrieve and anonymization of the dataset.

The preliminary data of the study was presented in Radiologic Society of North America (RSNA) meeting in 2021.

The contribution of the first and second authors is equal.

The authors declare that they have nothing to disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work.

References

1. Siegel RL, Miller KD, Wagle NS, *et al*. Cancer statistics, 2023. *CA Cancer J Clin* 2023;**73**:17–48. https://doi.org/10.3322/caac.21763.
2. Kohler BA, Sherman RL, Howlader N, *et al*. Annual report to the nation on the status of cancer, 1975-2011, featuring incidence of breast cancer subtypes by race/ethnicity, poverty, and state. *J Natl Cancer Inst* 2015;**107**:djv048, https://doi.org/10.1093/jnci/djv048.
3. de Gelder R, Heijnsdijk EAM, Fracheboud J, *et al*. The effects of population-based mammography screening starting between age 40 and 50 in the presence of adjuvant systemic therapy. *Int J Cancer* 2015;**137**:165–72. https://doi.org/10.1002/ijc.29364.
4. Sprague BL, Gangnon RE, Burt V, *et al*. Prevalence of mammographically dense breasts in the United States. *J Natl Cancer Inst* 2014;**106**:dju255, https://doi.org/10.1093/jnci/dju255.
5. Ozkan Gurdal S, Ozaydın AN, Aribal E, *et al*. Bahcesehir long-term population-based screening compared to National Breast Cancer Registry Data: effectiveness of screening in an emerging country. *Diagn Interv Radiol* 2021;**27**:157–63. https://doi.org/10.5152/dir.2021.20486.
6. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006;**15**:1159–69. https://doi.org/10.1158/1055-9965.EPI-06-0034.
7. Bakker MF, de Lange SV, Pijnappel RM, *et al*. Supplemental MRI screening for women with extremely dense breast tissue. *N Engl J Med* 2019;**381**:2091–102. https://doi.org/10.1056/NEJMoa1903986.
8. Bae MS, Moon WK, Chang JM, *et al*. Breast cancer detected with screening US: reasons for nondetection at mammography. *Radiology* 2014;**270**:369–77. https://doi.org/10.1148/radiol.13130724.
9. European CommissionDirectorate-General for Health and Consumers, Karsa L, Holland R, Broeders M, *et al*. European guidelines for quality assurance in breast cancer screening and diagnosis - Fourth edition, supplements. In: Karsa L, Holland R, Broeders M, *et al*., editors. Publications Office; 2013 https://data.europa.eu/doi/10.2772/13196.

10. Nagtegaal ID, Allgood PC, Duffy SW, *et al*. Prognosis and pathology of screen-detected carcinomas: how different are they? *Cancer* 2011;**117**:1360–8. https://doi.org/10.1002/cncr.25613.
11. Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer* 2017;**3**:12. https://doi.org/10.1038/s41523-017-0014-x.
12. Niraula S, Biswanger N, Hu P, *et al*. Incidence, characteristics, and outcomes of interval breast cancers compared with screening-detected breast cancers. *JAMA Netw Open* 2020;**3**:e2018179, https://doi.org/10.1001/jamanetworkopen.2020.18179.
13. Aribal E, Mora P, Chaturvedi AK, *et al*. Improvement of early detection of breast cancer through collaborative multi-country efforts: observational clinical study. *Eur J Radiol* 2019;**115**:31–8. https://doi.org/10.1016/j.ejrad.2019.03.020.
14. Ginsburg O, Yip C-H, Brooks A, *et al*. Breast cancer early detection: a phased approach to implementation. *Cancer* 2020;**126**:2379–93. https://doi.org/10.1002/cncr.32887.
15. da Costa Vieira RA, Biller G, Uemura G, *et al*. Breast cancer screening in developing countries. *Clinics (Sao Paulo)* 2017;**72**:244–53. https://doi.org/10.6061/clinics/2017(04)09.
16. Lei S, Zheng R, Zhang S, *et al*. Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020. *Cancer Commun (Lond)* 2021;**41**:1183–94. https://doi.org/10.1002/cac2.12207.
17. Özmen V, Özmen T, Doğru V. Breast cancer in Turkey; an analysis of 20.000 patients with breast cancer. *Eur J Breast Health* 2019;**15**:141–6. https://doi.org/10.5152/ejbh.2019.4890.
18. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, *et al*. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. *J Natl Cancer Inst* 2019;**111**:916–22. https://doi.org/10.1093/jnci/djy222.
19. Schaffter T, Buist DSM, Lee CI, *et al*. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw Open* 2020;**3**:e200265, https://doi.org/10.1001/jamanetworkopen.2020.0265.
20. Kizildag Yirgin I, Koyluoglu YO, Seker ME, *et al*. Diagnostic performance of AI for cancers registered in A mammography screening program: a retrospective analysis. *Technol Cancer Res Treat* 2022;**21**:15330338221075172, https://doi.org/10.1177/15330338221075172.
21. Lång K, Josefsson V, Larsson A-M, *et al*. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *Lancet Oncol* 2023;**24**:936–44. https://doi.org/10.1016/S1470-2045(23)00298-X.
22. Dembrower K, Crippa A, Colón E, *et al*. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health* 2023;**5**:e703–11. https://doi.org/10.1016/S2589-7500(23)00153-X.
23. Wanders AJT, Mees W, Bun PAM, *et al*. Interval cancer detection using a neural network and breast density in women with negative screening mammograms. *Radiology* 2022;**303**:269–75. https://doi.org/10.1148/radiol.210832.
24. Graewingholt A, Rossi PG. Retrospective analysis of the effect on interval cancer rate of adding an artificial intelligence algorithm to the reading process for two-dimensional full-field digital mammography. *J Med Screen* 2021;**28**:369–71. https://doi.org/10.1177/0969141320988049.
25. Larsen M, Aglen CF, Lee CI, *et al*. Artificial intelligence evaluation of 122 969 mammography examinations from a population-based screening program. *Radiology* 2022;**303**:502–11. https://doi.org/10.1148/radiol.212381.
26. Dembrower K, Wåhlin E, Liu Y, *et al*. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *Lancet Digit Health* 2020;**2**:e468–74. https://doi.org/10.1016/S2589-7500(20)30185-0.
27. Hickman SE, Woitek R, Le EPV, *et al*. Machine learning for workflow applications in screening mammography: systematic review and meta-analysis. *Radiology* 2022;**302**:88–104. https://doi.org/10.1148/radiol.2021210391.
28. Seker ME, Koyluoglu YO, Ozaydin AN, *et al*. Diagnostic capabilities of artificial intelligence as an additional reader in a breast cancer screening program. *Eur Radiol* 2024, https://doi.org/10.1007/s00330-024-10661-3.
29. UN DESA Policy Brief No. 155: Accelerating middle-income countries' progress towards sustainable development | Department of Economic and Social Affairs. https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-no-155-accelerating-middle-income-countries-progress-towards-sustainable-development/. Accessed 21 Feb 2024.
30. Ikeda DM, Birdwell RL, O'Shaughnessy KF, *et al*. Analysis of 172 subtle findings on prior normal mammograms in women with breast cancer detected at follow-up screening. *Radiology* 2003;**226**:494–503. https://doi.org/10.1148/radiol.2262011634.
31. Hofvind S, Skaane P, Vitak B, *et al*. Influence of review design on percentages of missed interval breast cancers: retrospective study of interval cancers in a population-based screening program. *Radiology* 2005;**237**:437–43. https://doi.org/10.1148/radiol.2372041174.
32. Hoff SR, Abrahamsen A-L, Samset JH, *et al*. Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography– results from a retrospective review. *Radiology* 2012;**264**:378–86. https://doi.org/10.1148/radiol.12112074.
33. Byng D, Strauch B, Gnas L, *et al*. AI-based prevention of interval cancers in a national mammography screening program. *Eur J Radiol* 2022;**152**:110321, https://doi.org/10.1016/j.ejrad.2022.110321.
34. Lång K, Hofvind S, Rodríguez-Ruiz A, *et al*. Can artificial intelligence reduce the interval cancer rate in mammography screening? *Eur Radiol* 2021;**31**:5940–7. https://doi.org/10.1007/s00330-021-07686-3.
35. Lång K, Dustler M, Dahlblom V, *et al*. Identifying normal mammograms in a large screening population using artificial intelligence. *Eur Radiol* 2021;**31**:1687–92. https://doi.org/10.1007/s00330-020-07165-1.
36. Larsen M, Aglen CF, Hoff SR, *et al*. Possible strategies for use of artificial intelligence in screen-reading of mammograms, based on retrospective data from 122,969 screening examinations. *Eur Radiol* 2022;**32**:8238–46. https://doi.org/10.1007/s00330-022-08909-x.
37. Hovda T, Tsuruda K, Hoff SR, *et al*. Radiological review of prior screening mammograms of screen-detected breast cancer. *Eur Radiol* 2021;**31**:2568–79. https://doi.org/10.1007/s00330-020-07130-y.