




Diagnostic Performance of AI for Cancers Registered in A Mammography Screening Program: A Retrospective Analysis

Technology in Cancer Research & Treatment
Volume 21: 1-11
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15330338221075172
journals.sagepub.com/home/tct


Inci Kizildag Yirgin, MD¹ , Yilmaz Onat Koyluoglu, MD², Mustafa Ege Seker, MD² , Sibel Ozkan Gurdal, MD³, Ayse Nilufer Ozaydin, MD⁴, Beyza Ozcinar, MD⁵, Neslihan Cabioglu, MD⁵, Vahit Ozmen, MD⁵, and Erkin Aribal, MD²

Abstract

Purpose: To evaluate the performance of an artificial intelligence (AI) algorithm in a simulated screening setting and its effectiveness in detecting missed and interval cancers. **Methods:** Digital mammograms were collected from Bahcesehir Mammographic Screening Program which is the first organized, population-based, 10-year (2009-2019) screening program in Turkey. In total, 211 mammograms were extracted from the archive of the screening program in this retrospective study. One hundred ten of them were diagnosed as breast cancer (74 screen-detected, 27 interval, 9 missed), 101 of them were negative mammograms with a follow-up for at least 24 months. Cancer detection rates of radiologists in the screening program were compared with an AI system. Three different mammography assessment methods were used: (1) 2 radiologists' assessment at screening center, (2) AI assessment based on the established risk score threshold, (3) a hypothetical radiologist and AI team-up in which AI was considered to be the third reader. **Results:** Area under curve was 0.853 (95% CI = 0.801-0.905) and the cut-off value for risk score was 34.5% with a sensitivity of 72.8% and a specificity of 88.3% for AI cancer detection in ROC analysis. Cancer detection rates were 67.3% for radiologists, 72.7% for AI, and 83.6% for radiologist and AI team-up. AI detected 72.7% of all cancers on its own, of which 77.5% were screen-detected, 15% were interval cancers, and 7.5% were missed cancers. **Conclusion:** AI may potentially enhance the capacity of breast cancer screening programs by increasing cancer detection rates and decreasing false-negative evaluations.

Keywords

artificial intelligence, breast cancer, deep learning, mammography, screening

Abbreviations

AI, Artificial intelligence; ANNs, Artificial neural networks; BIRADS, Breast Imaging-Reporting and Data System of the American College of Radiology; BMSP, Bahcesehir Mam mographic Screening Program; CAD, Computer-aided detection; CC, Craniocaudal; CNN's, Convolutional neural networks; DBT, Digital breast tomosynthesis; DCIS, Ductal carcinoma in situ; DL, Deep learning; MRI, Magnetic resonance imaging; ML, Machine learning; MLO, Mediolateral oblique; NN, Neural network; ROC, Receiver operating characteristic

Received: October 11, 2021; Revised: December 20, 2021; Accepted: January 5, 2022.

Introduction

Breast cancer is the most common cancer type and the second cause of cancer-based mortality in women according to the 2020 global cancer statistics.¹ Screening for breast cancer with mammography has shown a reduction in breast cancer mortality by many randomized trials and incidence-based

¹ Oncology Institute, Istanbul University, Istanbul, Turkey

² Acibadem M.A.A University School of Medicine, Istanbul, Turkey

³ Namik Kemal University, School of Medicine, Tekirdag, Turkey

⁴ Marmara University, School of Medicine, Istanbul, Turkey

⁵ Istanbul University, School of Medicine, Istanbul, Turkey

Corresponding Author:

Inci Kizildag Yirgin, MD, Department of Radiology, Oncology Institute, Istanbul University, Istanbul, 34390, Capa, Turkey.
Email: inci.kizildagyirgin@istanbul.edu.tr



mortality studies.²⁻⁵ Therefore, many developed countries have implemented large-scale mammography screening programs in the last 3 decades. But, despite these successful screening programs and improved treatment options, breast cancer is still one of the major causes of cancer-related death in women around the world. The efficiency of mammography remains controversial.^{6,7} The main reported disadvantages of mammography are high rates of false positives and false negatives.⁸ Studies have shown that up to 30% to 40% of cancers can be missed during mammography screening and only 10% of women are recalled for further diagnostic workup and are diagnosed with breast cancer.^{9,10} This fact can be explained by many reasons as follows; dense breast tissue, false positioning, human interpretation error. On the other hand, important consequences of high recall rate and false positivity in daily practice are increased patient anxiety, excessive follow-up, and invasive diagnostic procedures. Due to these disadvantages, there has been a need for methods and techniques that will increase the sensitivity and specificity, and correct reading rates of mammography evaluation. Other radiological methods including ultrasound, digital breast tomosynthesis, and magnetic resonance imaging have been introduced for screening, but mammography is still the frontline and most common modality used around the globe.

Double reading by 2 radiologists independently has been implemented in screening programs to improve cancer detection rates. Successful rates were achieved in cancer detection, there has been a decrease in recall rates and positive predictive values for cancer detection.^{11,12} Computer-aided detection (CAD) software was designed in 1998 to improve and assist mammography readings.¹³ The literature on CAD's efficacy is still under discussion despite being used for 2 decades. Early studies had shown improved cancer detection.^{14,15} However, large-scale studies have shown a high false-positive rate, low specificity, and failure to improve radiologists' performance due to increased additional review and number of marked areas.^{16,17}

Artificial intelligence (AI) is a rapidly growing branch of computer science that has created great excitement this century with breakthroughs in the development of many applications and its potential to change paradigms in breast imaging.¹⁸ CAD is based on human decisions like density or shape and results as negative or positive, however, AI algorithms can find the new characteristics which enables classification of lesions which are unknown and undetectable by human eyes. Many studies have shown that AI can increase the sensitivity for detection of breast cancer and decrease false-positive evaluations, so indicates a great potential in improving radiologists' contributions to patients' care.¹⁹⁻²³ Machine learning (ML) and deep learning (DL) models have been used in personal breast cancer risk assessment, predicting pathologic upgrade of high-risk lesions, detecting the negative screening mammography, estimating the presence of invasive component accompanying ductal carcinoma in situ, early prediction of response after neoadjuvant chemotherapy or prediction of lymph node metastases in primary breast cancer.²⁴⁻³⁴

AI is defined as a large spectrum description for many different sections and training models, including artificial neural networks, ML, and DL. ML is based on a training model which learns to identify the characteristics and associated variables that are described and observed in the input data.³⁵ DL is a learning model based on multiple layers of deep neural networks (NN), that is similar to human neural tissue in the brain.³⁶ DL is more important for radiology, especially for breast imaging due to its ability to learn the characteristics which are essential to categorize the mammograms as positive or negative and has a potential to find new correlations which were not evident for human interpretation. Last year Kim et al developed and validated an AI algorithm by using large-scale data and showed better diagnostic efficiency than radiologists in breast cancer detection.³⁷

In this study, we aimed to evaluate the performance of an AI algorithm in a simulated screening setting and its effectiveness in detecting missed and interval cancers.

Materials and Methods

Population

Digital mammograms were collected from Bahcesehir Mammographic Screening Program (BMSP) which is the first organized, population-based, 10-year (2009-2019) mammography screening program in Turkey. During the 10-year period, biennially, women between ages 40 and 69 in the region were invited to the screening. The mammograms taken biennially were recorded in the archive system. The study was approved by the ethics committee of Acibadem M.A.A University school of medicine institutional review board number with 2020-22/23 (location: Istanbul, date: 15.10.2020). Each eligible woman signed a written informed consent form when they were enrolled to the BMSP. The reporting of this study conforms to STROBE guidelines.³⁸ Patient and tumor characteristics are listed in Table 1.

Mammograms

During the 10-year screening period, a total of 22 621 screening examinations were performed. All cancers detected in the screening program during this period were included into the study without an exclusion criterion. In total, 211 mammograms were extracted from the archive of the screening program in this retrospective study. One hundred ten of these were diagnosed as breast cancer (74 screen-detected, 27 interval, 9 missed), 101 were negative mammograms. The negative mammograms were chosen from the mammograms of women who did not have any breast related diagnosis in the 2 years following the initial mammogram and who matched the ages and densities of the cancer patients. Power analysis was performed using the Open Epi program. It was found that the sample size was sufficient for this study. Definitions of diagnosed breast cancers were as follows: (1) interval cancer was described as finding a primary breast cancer following a negative mammographic evaluation within 2 years; (2) missed cancer was

Table 1. Patient and Tumor Characteristics.

Characteristics	
Age (median, range)	53.4(40-73)
Breast density	
A: Fatty	7
B: Scattered fibro glandular density	24
C: Heterogeneously dense	59
D: Extremely dense	20
BI-RADS	
BI-RADS 0: incomplete	
BI-RADS 1: negative.	6
BI-RADS 2: benign.	15
BI-RADS 3: probably benign	15
BI-RADS 4a: low suspicion for malignancy	33
BI-RADS 4b: moderate suspicion for malignancy	1
BI-RADS 4c: high suspicion for malignancy	7
BI-RADS 5: highly suggestive of malignancy	33
Definition of cancer	
Screen detected	74
Interval	27
Missed	9
Histological type	
Invasive ductal carcinoma	69
Invasive lobular carcinoma	19
Ductal carcinoma in-situ	11
Tubular carcinoma	3
Mixed carcinoma	2
Microinvasive carcinoma	2
Papillary carcinoma	2
Mucinous carcinoma	2
Tumor size (mm) (median, range)	15 (3-37)
Cancer stage	
Stage 0	12
Stage 1	57
Stage 2a	18
Stage 2b	10
Stage 3a	10
Stage 3b	1
Stage 3c	1
Stage 4	1

Abbreviations: BIRADS, Breast Imaging Reporting and Data System.

described as the detection of breast cancer after a false-negative mammogram but detected in the first 30 days with another imaging modality or had clinical findings; (3) screen-detected cancer was defined as detection of cancer with a routine screening mammogram. All cancer cases that were diagnosed during the BMSP were included into our retrospective study. Negative mammograms were used as a control group by AI evaluation.

Image Analysis

Digital mammography images were obtained with a full-field digital mammographic device (Selenia, Hologic) from the screening center. Two projections, mediolateral oblique and craniocaudal were obtained for each woman. Two breast radiologists with more than 5 years of experience read the

mammograms in the screening center independently. In case of inconsistency between the readers, a third radiologist with more than 20 years of experience interpreted the findings for the final decision. Radiological findings were evaluated under the guidance of 4th edition of Breast Imaging-Reporting and Data System of the American College of Radiology (BIRADS).³⁹ The BMSP had already started before the last updated 5th version of BIRADS.

Artificial Intelligence System

We used a recently developed diagnostic support software (Lunit INSIGHT MMG, Seoul, South Korea) on a free website (<https://insight.lunit.io/mmg/login>).³⁷ The AI algorithm of this software uses deep convolutional neural networks (CNNs) and highlights areas in the mammograms where the suspicion of malignancy is above a certain threshold.⁴⁰ The system calculates an abnormality score which reflects the likelihood of malignancy of the detected lesion. The score between 1% and 100% likelihood of malignancy flagged by AI is recorded. In this study, we did not use the images, but instead used the underlying prediction score of the algorithm. In the case of multiple findings with different values, the highest score is considered final. The images used in this study have never been used to train, validate, or test a previously developed AI algorithm.

Statistical Analysis

The breast cancer detection rates of radiologists in the screening program compared with AI system in a simulation scenario. Receiver operating characteristic (ROC) analysis was done and a threshold for cancer detection was calculated with Youden's index. All mammograms were relabeled based on the threshold. Three different mammography assessment methods were compared in this study: (1) 2 radiologists' assessment at screening center, (2) AI assessment based on the established risk score threshold, (3) a hypothetical radiologist and AI team-up in which AI is defined as the third reader. R systems (R Core Team, 2020) and pROC package (Robin X. et al, 2011) were used for statistical analysis.

Results

In total 211 mammograms, 74 screen-detected cancers (67.3%), 27 interval cancers (24.5%), 9 missed cancers (8.1%), and 101 negative control group mammograms were evaluated by the AI. Area under curve (AUC) was 0.853 (95% CI = 0.801-0.905) and the cut-off value for risk score was 34.5% with a sensitivity of 72.8% (80/110) and a specificity of 88.3% (89/101) for AI cancer detection in ROC analysis (Figure 1).

Risk score distributions for each cancer subgroups were as follows, 83.8% of screen-detected cancers showed a risk score higher than 34.5% while 16.2% of them were below, 44.4% of interval cancers had a risk score higher than 34.5%, while 55.5% of them were below and lastly 66.6% of missed

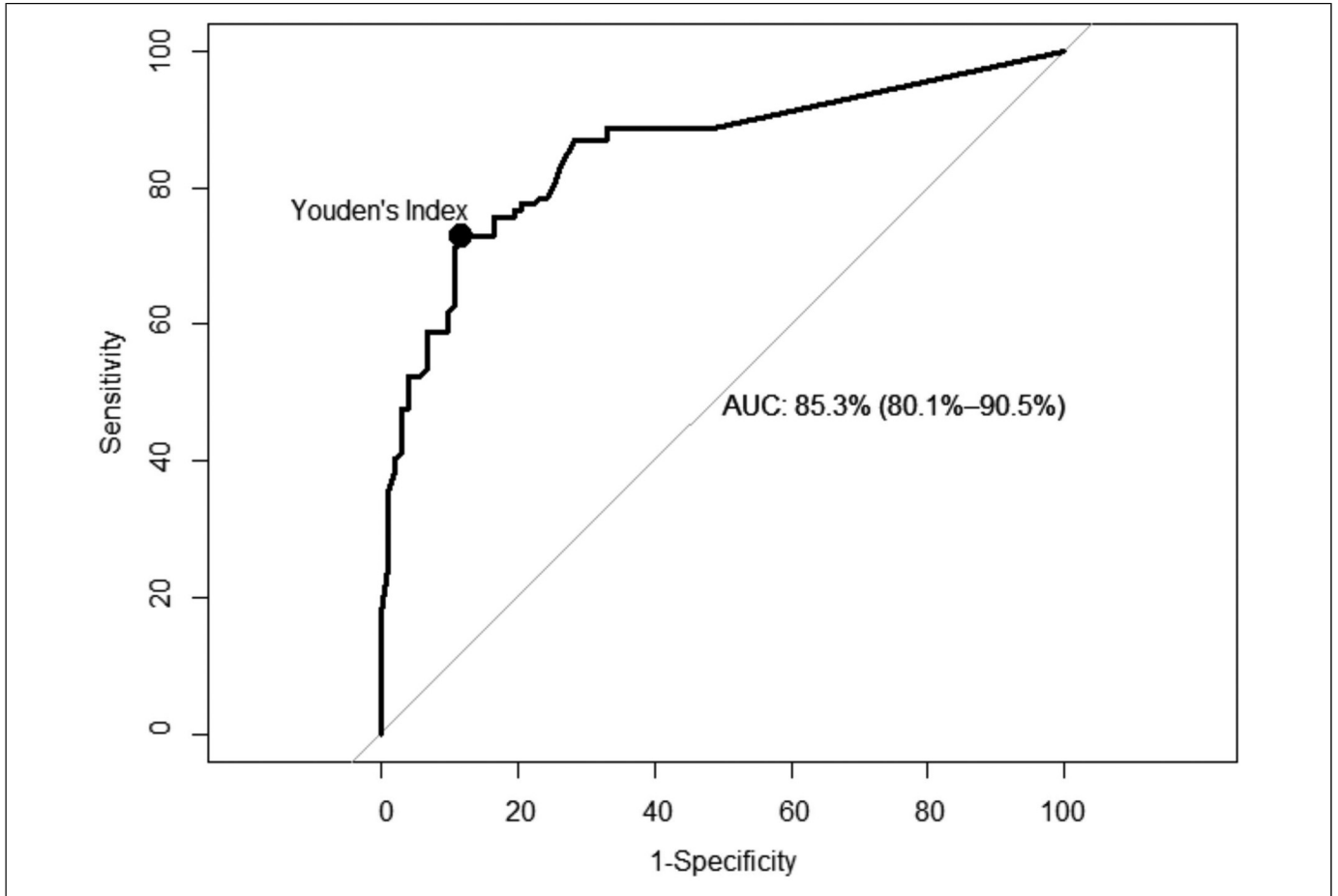


Figure 1. Receiver operating characteristic (ROC) analysis and the threshold calculated with Youden's index.

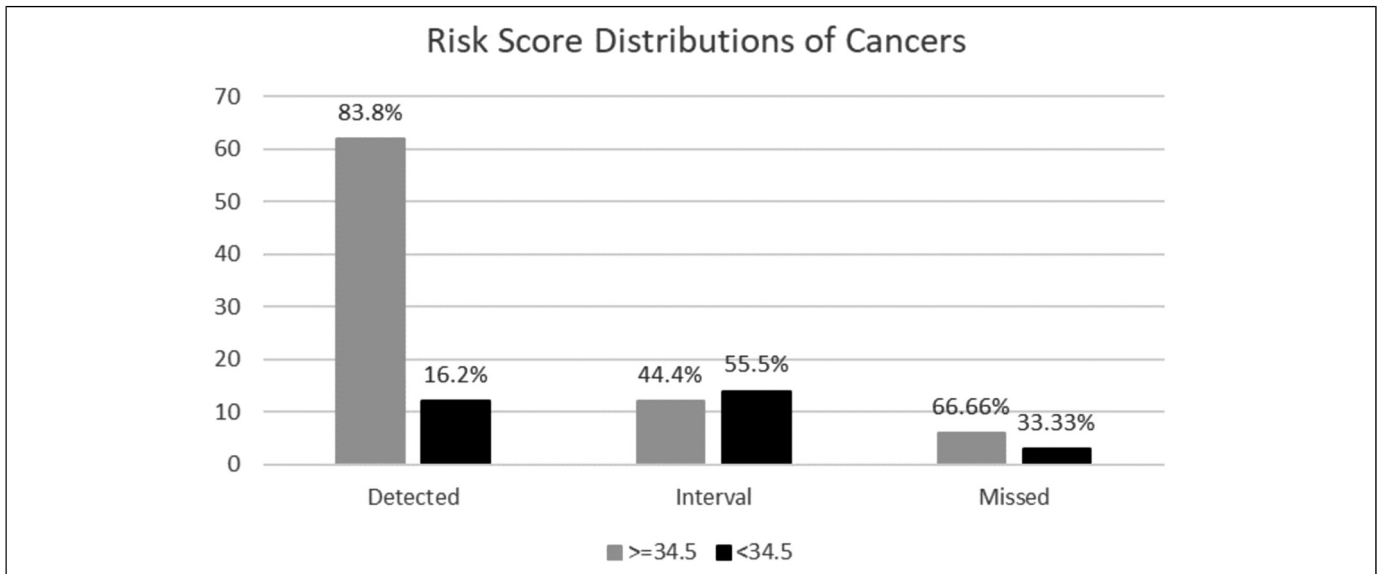


Figure 2. Risk score distributions for each cancer subgroup.

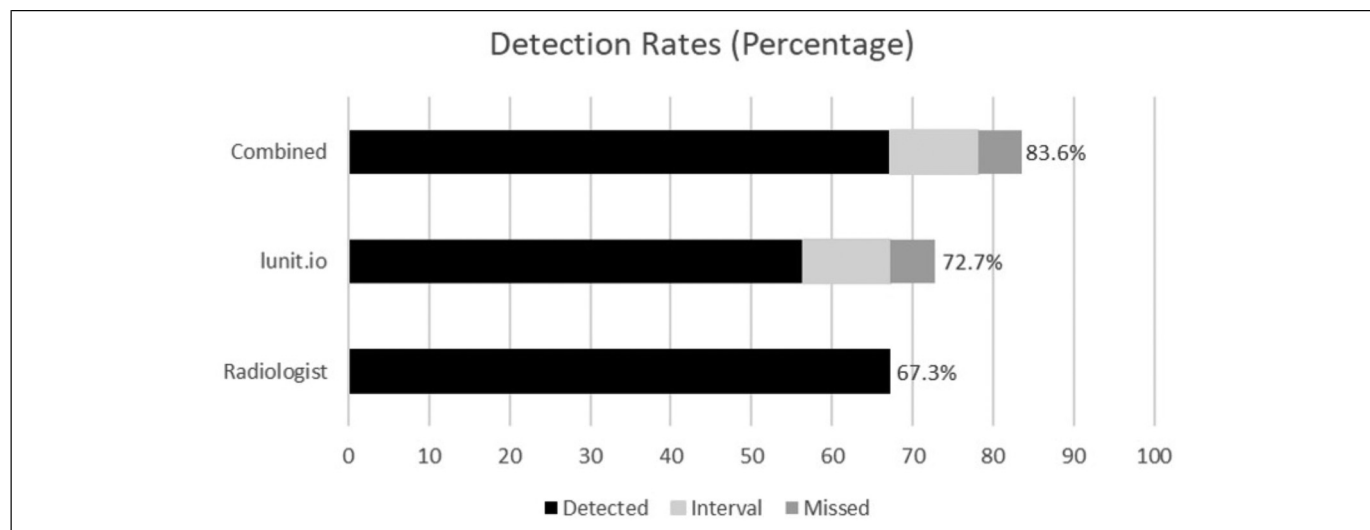


Figure 3. Cancer detection rates for each group.

cancers had a risk score higher than 34.5% while 33.3% of them were below (Figure 2).

Overall cancer detection rates were 67.3% (74/110) for radiologists, 72.7% (80/110) for AI, and 83.6% (92/110) for radiologists and AI team-up (Figure 3). AI detected 72.7% (80/110) of all cancers on its own, of which 62 were screen-detected, 12 were interval cancers and 6 were missed cancers. Hypothetical AI and radiologist team-up detected 83.6% (92/110) of all cancers, of which 74 were screen-detected, 12 were interval cancers, 6 were missed cancers. AI evaluated 16.2% of the true positive mammograms as a negative mammogram (Figure 4). On the other hand, AI detected an additional 44.4% (12/27) of interval 66.7% (6/9) of missed cancers that were not previously detected by radiologists (Figure 5; Table 2).

Discussion

In this study, we evaluated the performance of an AI algorithm in a simulated screening setting and its effectiveness in detecting missed and interval cancers. The cancer detection rate of AI was higher than radiologists, however, it was found to be lower than the hypothetical radiologist and AI team-up. AI was able to detect an additional 44.4% of interval and 66.7% of missed cancers which were not previously detected by radiologists.

In the present study, the AI system that we used in our study is based on CNNs which is the most used NN type in radiologic studies. We have shown that the AI algorithm is a successful diagnostic tool for breast cancer detection with 0.853 of AUC which is in line with the current literature.^{22,29} In 2019, Rodriguez-Ruiz et al published a retrospective, multi-reader, and multi-case study and investigated the performance of radiologists with and without a supporting AI.²¹ Their study included 240 mammograms (consisting of cancers, false-positive cases, and normal mammograms) read by 14 radiologists and resulted in statistically significant higher AUC values with AI support than unassisted reading (0.89 and

0.87, respectively). An improvement was made with less-experienced radiologists but not with experienced radiologists which makes one question the real effect in clinical practice. Therefore, the same group published a subsequent study with the same AI algorithm and compared the performances of AI with 101 radiologists and showed higher AUC values with AI (0.84 vs 0.81). This result was obtained not only with less-experienced radiologists, but also 61.4% of all radiologists.²² Pacile et al published a multi-reader study to evaluate the effectiveness of AI in breast cancer detection with a similar design to a previous study with 240 mammograms (including true-positive, false-negative, true-negative, false-positive cases) read by 14 radiologists with and without AI support and found AUC values of 0.769 and 0.797, respectively.⁴¹ Average sensitivity was also found to be increased with AI assistance for breast cancer detection in the same study. Our study included 211 mammograms consisting of true positive and false negative cases together with normal mammograms from a population based screening program. The difference between AUC values of different studies can be explained by different designs and AI algorithms. An optimal data should contain all types of mammographic evaluations in order to be as similar as possible to real-life or routine screening. Kim et al developed and validated an AI algorithm using 170,230 mammograms derived from 5 centers (South Korea, the USA, UK).³⁷ Then, they designed a multicenter, reader study with 320 mammograms (cancers, benign lesions, and normal mammograms) read by 14 radiologists and found a significant improvement in breast cancer detection rates. Overall AUC values for AI only, AI and radiologists, and radiologists only were 0.959, 0.881, and 0.810, respectively. Unlike the other studies, AI-only performance was better than AI-assisted radiologist's performance. In a detailed analysis, they showed that AI had a better performance especially in detecting early stages of cancer (T1 and node-negative cancers) and also the cases presenting with asymmetry or architectural distortions.

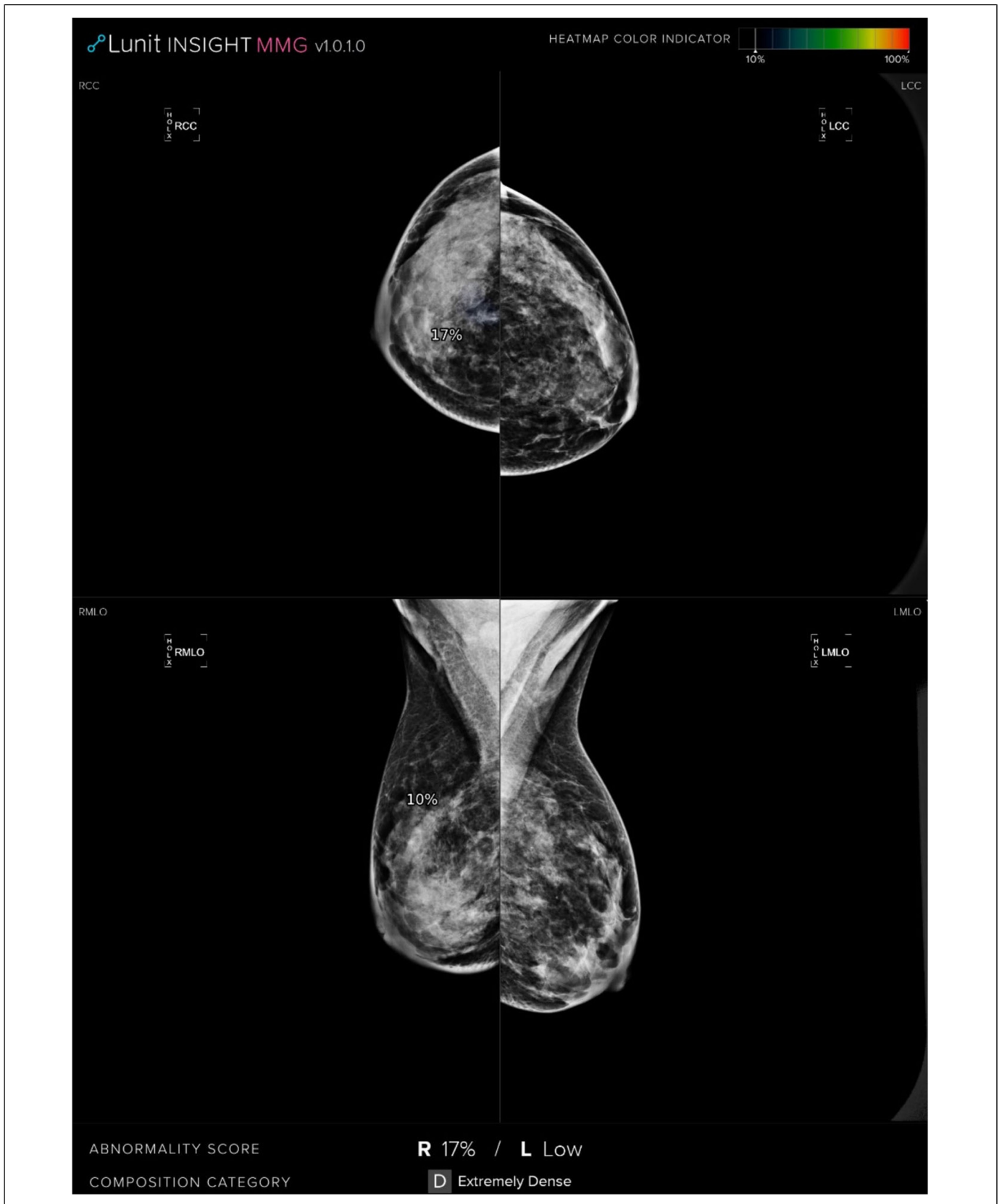


Figure 4. CC and MLO mammograms show a lesion presented with architectural distortion in the retroglandular space of the upper quadrant of the right breast which was assessed as true positive by a radiologist. However, the AI system calculated the risk score as 17% and assessed as a negative mammogram.

Abbreviations: CC, craniocaudal; MLO, mediolateral oblique; AI, artificial intelligence.

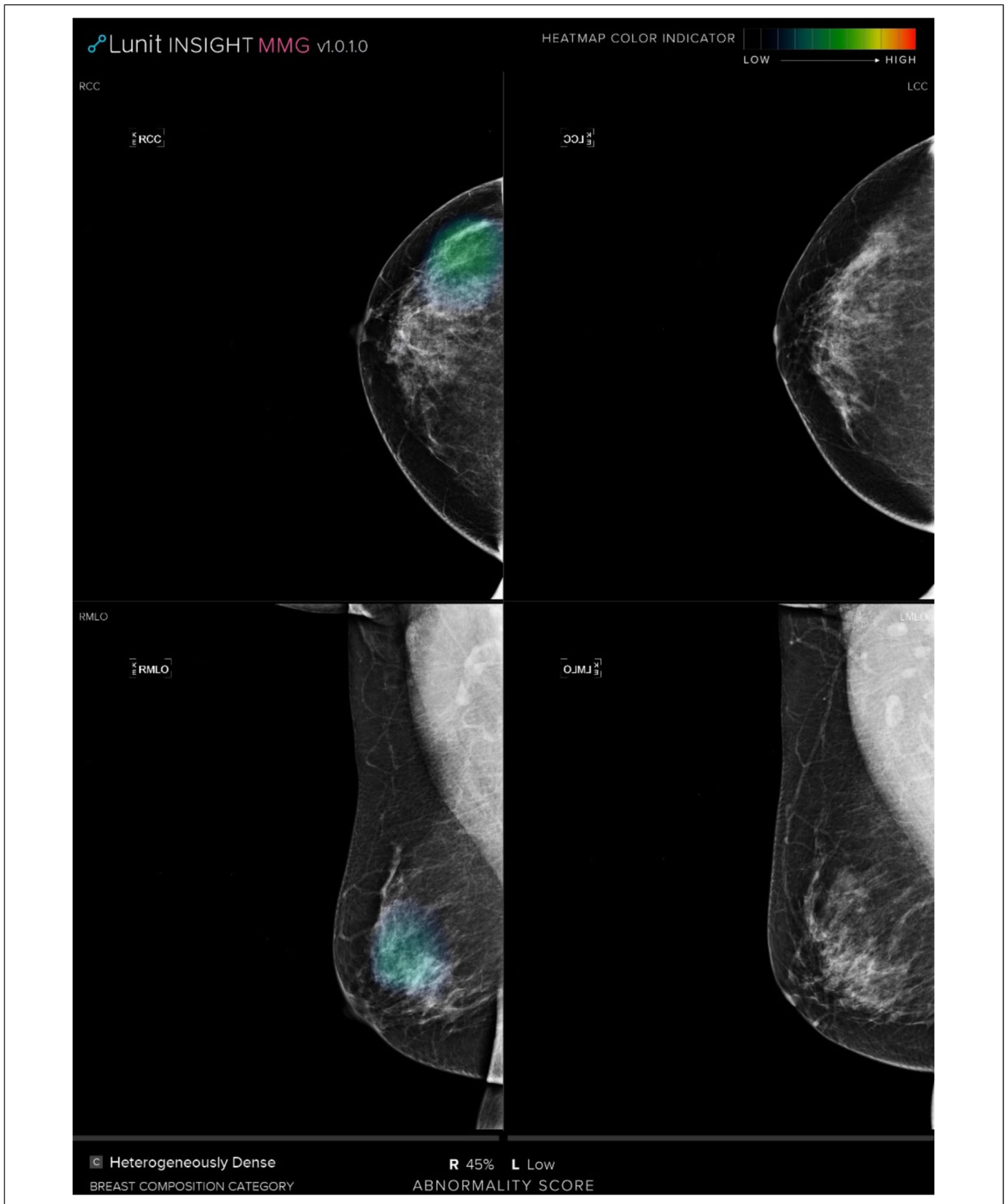


Figure 5. CC and MLO mammograms evaluated as a negative mammogram by radiologist; however, the AI system detected the missed cancer with a risk score of 45%.
Abbreviations: CC, craniocaudal; MLO, mediolateral oblique; AI, artificial intelligence.

Table 2. Mammographic and Clinicopathologic Features of AI-Detected Additional Cancers.

Characteristics	AI-detected additional missed cancer (6 of 9)	AI- detected additional interval cancer (12 of 27)
Age (median, range)	53.1(44-61)	51.8 (41-69)
Breast density		
A: Fatty	0	1
B: Scattered fibro glandular density	1	2
C: Heterogeneously dense	4	6
D: Extremely dense	1	3
BI-RADS		
BI-RADS 0: incomplete	0	0
BI-RADS 1: negative.	2	0
BI-RADS 2: benign	4	6
BI-RADS 3: probably benign	0	6
BI-RADS 4a: low suspicion for malignancy	0	0
BI-RADS 4b: moderate suspicion for malignancy	0	0
BI-RADS 4c: high suspicion for malignancy	0	0
BI-RADS 5: highly suggestive of malignancy	0	0
Histological type		
Invasive ductal carcinoma	4	8
Invasive lobular carcinoma	2	2
Ductal carcinoma in situ	0	0
Tubular carcinoma	0	1
Mixed carcinoma	0	1
Microinvasive carcinoma	0	0
Papillary carcinoma	0	0
Mucinous carcinoma	0	0
Tumor size (mm) (median, range)	19(3-60)	15(7-32)
Cancer stage		
Stage 0	0	0
Stage 1	3	7
Stage 2a	2	3
Stage 2b	0	1
Stage 3a	1	1
Stage 3b	0	0
Stage 3c	0	0
Stage 4	0	0

Abbreviations: BIRADS, Breast Imaging Reporting and Data System; AI, artificial intelligence.

Additionally, AI was not affected by breast density as much as radiologists, according to the same study. These results showed that AI may positively contribute to the prognosis of patients by decreasing the rate of interval breast cancers. However, these studies did not focus on the interval or missed cancers and were designed in a prospective analysis where the readers were expected to detect a high number of positive mammograms than the real-life situation in which the cancer detection rate is less than 8 in 1000. This may cause a biased artificial environment where the reader stays more cautious. Although our study is retrospective the reader performance was real time in a real screening program.

Interval cancer rate for biannual mammography screening is between 0.8 and 2.1 per 1000 screening and these cancers tend to be biologically more aggressive tumors.⁴² Thus, reducing the interval cancer rate should result in a better outcome of a screening program. This study showed a potential decrease in interval cancers by 44.4% in a screening program. A study by Lang et al showed the effect of AI in detecting at 19.3% of the interval

cancers in mammography screening which is less than half of the interval cancer detection rate in our study. However, Lang et al included the interval cancers with the highest AI score of 10 in order not to increase the recall rate.⁴³ On the other hand, in our study, we have included middle and high scores with a threshold at 34.5 and achieved a high specificity at 88.3%. Interval cancers can be stratified as true negative and false negative depending on the presence of an evident finding on the initial mammogram. False negative interval cancers were reported between 25% and 40% in majority of the studies.⁴⁴ In other words, almost one-third of the interval cancers have visible findings in the initial mammograms and evitable. However, detection of subtle changes is challenging and difficult to improve without increasing the recall rates. Although additional information such as prior mammograms, clinical findings, or breast cancer risk can improve the outcomes, it may not be possible to evaluate this additional information in screening programs with high volume of mammograms.⁴⁵⁻⁴⁷ AI, as a second reader could be beneficial

in triage of the suspicious mammograms for a third referee reader.

Watanabe et al published a retrospective study to evaluate the effect of an AI-based CAD software in detecting missed cancers on mammograms.²³ They showed that only 51% of missed cancers could be detected without the assistance of AI while this number jumped to 62% with the assistance of AI. In our study, AI detected 66.6% of the missed cancers which is in line with their study. Both studies show that more than half of the missed cancers can be detected with an AI support. Human errors are the second main cause after overlapping breast tissue for nondetection of cancers at mammography.⁴⁸ Both errors could be decreased with the implementation of AI in screening reading. Our study showed that AI can both increase the detection rate of missed and interval cancers. However, AI detected 16.2% less screen detected cancers than radiologists and the hypothetical radiologist AI combinations showed the highest performance in detecting all the screen detected, missed, and interval cancers. This study shows that adding AI in the reading workflow will improve the outcome of screening. Shortage of human resources particularly in countries with limited resources is one of the main drawbacks of screening.⁴⁹ Implementing AI as a second reader in screening programs will not only help overcoming human resource shortages but will also ameliorate the outcomes.

Our study has several limitations. First, it is a retrospective study, and the performance of radiologists and AI was not correlated in a prospective setting. However, the cases were selected from a population-based screening program and all were evaluated by 2 experienced radiologists. Second, we did not evaluate the histopathological and radiological features of detected cancers that would provide a detailed information about the benefit of the AI system. Third, the AI algorithm evaluates solely the uploaded images but does not consider any other information such as clinical history, family history, or symptoms. Fourth, the Youden's index is computed on an "experimental" ROC curve including many discontinuities. ROC analysis requires appropriate curve fitting before adding any other consideration. The 34.5% threshold value calculated may be affected by "local" effects of the experimental ROC curve basically associated with the small sample size.

Conclusion

In conclusion, AI may potentially enhance the capacity of breast cancer screening programs by increasing cancer detection rates and decreasing false negative evaluations such as missed and interval cancers and may be implemented in the screening reading workflow.

Acknowledgments

The first three authors have contributed to the manuscript equally. This study is presented in the European Congress of Radiology 2021.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Ethics Statement

The study was approved by institutional review board number with 2020-22/23.

Informed consent

Written informed consent was obtained from the patient(s) for their anonymized information to be published in this article.

References

1. Data source: Globocan 2020 Graph production. Global Cancer Observatory (<http://gco.iarc.fr>)
2. Tabár L, Vitak B, Chen TH, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*. 2011;260(3):658-663. doi: 10.1148/radiol.11110469
3. IARC Working Group on the Evaluation of Cancer-Preventive Strategies. *Breast Cancer Screening*. Vol 15. IARC Press; 2016.
4. Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. *Cancer Epidemiol Biomarkers Prev*. 2006;15(1):45-51. doi: 10.1158/1055-9965.EPI-05-0349
5. Njor S, Nystrom L, Moss S, et al. Euro screen working group. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen*. 2012;19(Suppl 1):33-41. doi: 10.1258/jms.2012.012080
6. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet*. 2012;380(9855):1778-1786. doi: 10.1016/S0140-6736(12)61611-0
7. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast cancer tumor size, overdiagnosis, and mammography screening effectiveness. *N Engl J Med*. 2016;375(15):1438-1447. doi: 10.1056/NEJMoa1600249
8. Lehman CD, Arao RF, Sprague BL, et al. National performance benchmarks for modern screening digital mammography: update from the breast cancer surveillance consortium. *Radiology*. 2017;283(1):49-58. doi: 10.1148/radiol.2016161174
9. Rawashdeh MA, Lee WB, Bourne RM, et al. Markers of good performance in mammography depend on number of annual readings. *Radiology*. 2013;269(1):61-67. doi: 10.1148/radiol.13122581
10. Majid AS, de Paredes ES, Doherty RD, Sharma NR. Salvador × missed breast carcinoma: pitfalls and pearls. *Radiographics*. 2003;23(4):881-895. doi: 10.1148/rg.234025083
11. Posso M, Puig T, Carles M, Rué M, Canelo-Aybar C, Bonfill X. Effectiveness and cost-effectiveness of double reading in digital mammography screening: a systematic review and meta-analysis. *Eur J Radiol*. 2017;96:40-49. doi: 10.1016/j.ejrad.2017.09.013

12. Healy NA, O'Brien A, Knox M, et al. Consensus review of discordant imaging findings after the introduction of digital screening mammography: Irish national breast cancer screening program experience. *Radiology*. 2020;295(1):35-41. doi: 10.1148/radiol.2020181454
13. Giger ML, Chan HP, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. *Med Phys*. 2008;35(12):5799-5820. doi: 10.1118/1.3013555
14. Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology*. 2000;215(2):554-562. doi: 10.1148/radiology.215.2.r00ma15554.
15. Freer TW, Ulissey MJ screening mammography with computer aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*. 2001;220(3):781-786. doi: 10.1148/radiol.2203001282
16. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med*. 2007;356(14):1399-1409. doi: 10.1056/NEJMoa066099
17. Lehman CD, Wellman RD, Buist DS, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med*. 2015;175(11):1828-1837. doi: 10.1001/jamainternmed.2015.5231
18. Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian Association of radiologists (CAR) artificial intelligence working group. Canadian association of radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J*. 2018;69(2):120-135. doi: 10.1016/j.carj.2018.02.002
19. Akselrod-Ballin A, Chorev M, Shoshan Y, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*. 2019;292(2):331-342. doi: 10.1148/radiol.2019182622
20. Mayo RC, Kent D, Sen LC, Kapoor M, Leung JWT, Watanabe AT. Reduction of false-positive markings on mammograms: a retrospective comparison study using an artificial intelligence-based CAD. *J Digit Imaging*. 2019;32(4):618-624. doi: 10.1007/s10278-018-0168-6
21. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology*. 2019;290(2):305-314. doi: 10.1148/radiol.2018181371
22. Rodríguez-Ruiz A, Lång K, Gubern-Merida A, et al. Standalone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. *J Natl Cancer Inst*. 2019;111(9):916-922. doi: 10.1093/jnci/djy222
23. Watanabe AT, Lim V, Vu HX, et al. Improved cancer detection using artificial intelligence: a retrospective evaluation of missed cancers on mammography. *J Digit Imaging*. 2019;32(4):625-637. doi: 10.1007/s10278-019-00192-5
24. Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD high-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. *Radiology*. 2018;286(3):810-818. doi: 10.1148/radiol.2017170549
25. Shi B, Grimm LJ, Mazurowski MA, et al. Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features. *J Am Coll Radiol*. 2018;15 (3 Pt B):527-534. doi: 10.1016/j.jacr.2017.11.036
26. Conant EF, Toledano AY, Periaswamy S, et al. Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis. *Radiol Artif Intell*. 2019;1(4):e180096. doi: 10.1148/ryai.2019180096
27. Dembrower K, Liu Y, Azizpour H, et al. Comparison of a deep learning risk score and standard mammographic density score for breast cancer risk prediction. *Radiology*. 2019;294(2):265-272. doi: 10.1148/radiol.2019190872
28. Ha R, Chang P, Karcich J, et al. Convolutional neural network based breast cancer risk stratification using a mammographic dataset. *Acad Radiol*. 2019;26(4):544-549. doi: 10.1016/j.acra.2018.06.020
29. Kyono T, Gilbert FJ, Van der schaar MImproving workflow efficiency for mammography using machine learning. *J Am Coll Radiol*. 2020;17(1 Pt A):56-63. doi: 10.1016/j.jacr.2019.05.012
30. Gullo R L, Eskreis-Winkler S, Morris EA, Pinker K machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy. *Breast*. 2020;49:115-122. doi: 10.1016/j.breast.2019.11.009
31. Rodríguez-Ruiz A, Lång K, Gubern-Merida A, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. *Eur Radiol*. 2019;29(9):4825-4832. doi: 10.1007/s00330-019-06186-9
32. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*. 2019;292(1):60-66. doi: 10.1148/radiol.2019182716
33. Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A deep learning model to triage screening mammograms: a simulation study. *Radiology*. 2019(1);293:38-46. doi: 10.1148/radiol.2019182908
34. Zhou LQ, Wu XL, Huang SY, et al. Lymph node metastasis prediction from primary breast cancer US images using deep learning. *Radiology*. 2020;294(1):19-28. doi: 10.1148/radiol.2019190372
35. Mitchell TM. *Machine Learning*. McGraw-Hill; 1997.
36. LeCun YA, Bengio Y, Hinton GE. Deep learning. *Nature*. 2015;521(7553):436-444. doi: 10.1038/nature14539
37. Kim HE, Kim HH, Han BK, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health*. 2020;2(3):e138-e148. doi: 10.1016/S2589-7500(20)30003-0
38. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577.
39. Reston VA. *Illustrated Breast Imaging Reporting and Data System (BI-RADS)*. 4th ed. American College of Radiology (ACR); 2003.
40. Dembrower K, Wählin 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(9): e468-e474. doi:10.1016/S2589-7500(20)30185-0
41. Pacilè S, Lopez J, Chone P, Bertinotti T, Grouin JM, Fillard P. Improving breast cancer detection accuracy of mammography with the concurrent Use of an artificial intelligence tool. *Radiology: Artificial Intelligence*. 2020;2(6):e190208. doi: 10.1148/ryai.2020190208
42. Houssami N, Hunter K The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer*. 2017;3:12. doi: 10.1038/s41523-017-0014-x
43. Lång K, Hofvind S, Rodríguez-Ruiz A, Andersson I. Can artificial intelligence reduce the interval cancer rate in mammography screening? *Eur Radiol*. 2021;31(8):5940-5947. doi: 10.1007/s00330-021-07686-3.
44. Houssami N, Irwig L, Ciatto S radiological surveillance of interval breast cancers in screening programmes. *Lancet Oncol*. 2006;7(3):259-265. doi: 10.1016/S1470-2045(06)70617-9
45. Moberg K, Grundström H, Törnberg S, et al. Two models for radiological reviewing of interval cancers. *J Med Screen*. 1999;6(1):35-39. doi: 10.1136/jms.6.1.35
46. Harvey JA, Fajardo LL, Innis CA previous mammograms in patients with impalpable breast carcinoma: retrospective versus blinded interpretation. *AJR Am J Roentgenol*. 1993;161(6):1167-1172. doi: 10.2214/ajr.161.6.8249720
47. Ciatto S, Rosseli Del Turco M, Zappa M. The detectability of breast cancer by screening mammography. *Br J Cancer*. 1995;71(2):337-339. doi: 10.1038/bjc.1995.67
48. Bae MS, Moon WK, Chang JM, et al. Breast cancer detected with screening US: reasons for nondetection at mammography. *Radiology*. 2014;270(2):369-377. doi: 10.1148/radiol.13130724
49. 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;115:31-38. doi: 10.1016/j.ejrad.2019.03.020