Diagnostic performance of two versions of an artificial intelligence system in interval breast cancer detection

Article *in* Acta Radiologica · September 2023

DOI: 10.1177/02841851231200785

CITATIONS
4

READS
49

5 authors, including:

Davut Can Güner

5 PUBLICATIONS **34** CITATIONS

SEE PROFILE

Erkin Aribal

Acıbadem University

164 PUBLICATIONS **1,535** CITATIONS

SEE PROFILE

Acta Radiologica
2023, Vol. 64(11) 2891–2897

DOI: 10.1177/02841851231200785
journals.sagepub.com/home/acr

Diagnostic performance of two versions of an artificial intelligence system in interval breast cancer detection

Levent Çelik[1], Davut Can Güner[1], Ömer Özçağlayan[1], Rahmi Çubuk[1] and Mustafa Erkin Arıbal[2]

Abstract

Background: Various versions of artificial intelligence (AI) have been used as a diagnostic tool aid in the diagnosis of breast cancer. One of the most important problems in breast screening progmrams is interval breast cancer (IBC).
Purpose: To compare the diagnostic performance of Transpara v1.6 and v1.7 in the detection of IBC.
Material and Methods: Reports of screening mammograms of a total 2,248,665 of women were evaluated retrospectively. Of 2,129,486 mammograms reported as Breast Imaging Reporting and Data System (BIRADS) 1 and 2, the IBC group consisted of 323 cases who were diagnosed as having cancer on mammography and were correlated with pathology in second mammogram taken >30 days after first mammogram. Four hundred and forty-one were defined as the control group because they did not change over 2 years. Cancer risk scores of both groups were determined from 1 to 10 with Tranpara v1.6 and v1.7. Diagnostic performances of both versions were evaluated by the receiver operating characteristic curve.
Results: Cancer risk scores 1 and 10 in v1.7 increased compared to v1.6 ($P < 0.001$). In all cases, sensitivity for v1.6 was 56.6%, specificity was 90%, and, for v1.7, sensitivity was 65.9% and specificity was 90%, respectively. In all cases, area under the curve values were 0.812 for v1.6 and 0.856 for v1.7, which was higher in v1.7 ($P < 0.001$). Diagnostic performance of v1.7 was higher than v1.6 at the 7–12-month period ($P < 0.001$).
Conclusion: The present study showed that Tranpara v1.7 has a higher specificity, sensitivity and diagnostic performance in IBC determination than v1.6. AI systems can be used in breast screening as a secondary or third reader in screening programs.

Keywords
Breast, neoplasms, screening, mammography, neural networks

Date received: 27 May 2023; accepted: 21 August 2023

Introduction

Interval breast cancer (IBC) is defined as breast cancer that is clinically detected after a negative screening exam, before a scheduled screening mammography. IBC depicts more agressive phenotypes than screen detected breast cancers (BC) (1). Therefore, IBC is an important health problem for women. A negative mammogram may develop due to the presence of radiologically occult lesions, the or absence of detectable lesions or a failure to interpret (2). Because of difficulties for IBC determination, IBC is an important indicator for the quality of screening programs (3).

Artificial intelligence (AI) is a recent diagnostic tool for mammography that has frequently been used in the detection of breast cancer. AI systems are used in the diagnosis of breast cancer that provide diagnosis at the level of experienced breast radiologists and contribute to diagnosis

[1]Department of Radiology, Maltepe University, School of Medicine, Istanbul, Turkey
[2]Deparment of Radiology, Acıbadem University, School of Medicine, Istanbul, Turkey

Corresponding author:
Ömer Özçağlayan, Department of Radiology, Maltepe University School of Medicine Hospital, Bağlarbaşı Mah., Feyzullah Cd. No:39, 34844 Maltepe/İstanbul, Turkey.
Email: omer.ozcaglayan@gmail.com

as a second reader (4,5). There are also studies that have investigated the roles and contributions of AI systems in IBC detection (6). Transpara (ScreenPoint Medical, Nijmegen, The Netherlands) is an AI system dedicated to detecting breast cancer in mammography. Studies have determined that the Transpara system contributes to detecting breast cancer with high accuracy rates (7). Many versions of the Transpara AI system are used in breast cancer detection. However, to date, there are no studies comparing the diagnostic performance of AI systems for IBC.

The present study aimed to evaluate the diagnostic performance and compare accuracy rates of Transpara v1.7 in the detection of IBC compared to v1.6, comprising last two versions of the Transpara system.

Material and Methods

Study population and mammography evaluation

This retrospective study was accepted by the local ethics committee (04/21/2021 date and 2021-08/14 approval number). In total, 2,248,665 mammogram reports were scanned retrospectively for women aged 40–69 years who were enrolled in the Turkish National Breast Screening Program between March 2016 and December 2019 with a recall rate of 5.3%. Some 2,129,486 of mammograms were classified as Breast Imaging Reporting and Data System (BIRADS) 1 and 2.

The breast cancer study group consisted of 442 women who were diagnosed as having breast cancer with pathologic correlation on days 5 and 730 after the first screening mammogram. All of the breast cancers were cases with a pathologic diagnosis of breast cancer in the Turkish National Cancer Registry. From this group, women who were diagnosed as having cancer in the mammogram taken 30 days after the first mammogram with pathologic correlation constituted the IBC group. Thirty-six of these cases were excluded from the study because of computed radiography mammography. Eighty-three cancer cases that were detected after within the first month from the screening mammography were excluded from the study. Although mammographic screenings are performed by invitation in our country, screening centers also screen women who present with their own compliants. For that reason, these patients with signs and symptoms were not actually screening patients and were excluded from the study, and these lesions may be associated with cancers with misdiagnosed mammography examinations that refer to an early pathologic cancer diagnosis. As a result, 323 cases reported as normal by two readers in digital mammography, with no symptoms (such as palpable mass lesion or bloody nipple discharge) and diagnosed as breast cancer with pathologic correlation at day 30 day after screening mammogram and before day 730, were labeled as IBC. Four hundred and forty-one women whose mammograms were BIRADS 1 and 2 and were next round mammograms also reported as normal, labeled as "time proven normal", comprised the control group (Fig. 1).

Transpara system

All normal and cancer cases were evaluated separately with AI systems (Transpara v1.6 and v1.7; ScreenPoint Medical). In both AI systems, the intensities of the images were normalized primarily. Then, convolutional neural networks were run to detect calcification and soft tissue lesions in both AI versions. Calcification and soft tissue images were combined to label suspicious lesions. Cases were scored from 1 to 10 for malignancy risk by both AI systems. The accuracy of the data from both systems was compared statistically.

Statistical analysis

Data were analyzed using SPSS, version 25.0 (IBM Corp., Armonk, NY, USA). Sensitivity and specificity values were determined for the diagnostic performances of Transpara v1.6 and v1.7 according to pathologic diagnosis as a reference value in the IBC evaluation. Area under curve (AUC) values were determined by the receiver operating characteristic (ROC) curve at a 95% confidence interval (CI). Then, the cases were divided into 1–6 months, 7–12 months, 13–18 months and 19–24 months according to the IBC diagnosis times, and the diagnostic performance of the AI systems was evaluated separately with the ROC curve. $P < 0.05$ was considered stastistically significant.

Results

Lesions labeled as Transpara score 1 increased from 8% to 11% and Transpara score 10 increased from 26% to 29% in v1.7 compared to v1.6, respectively (Fig. 2). When the diagnostic performances of Transpara v1.6 were compared with v1.7 in all cases, the sensitivity of Transpara v1.6 was 56.6% and the specificity was 90.0%, and the sensitivity of Transpara v1.7 was 65.9% and the specificity was 90.0%, respectively (Fig. 3). According to the sensitivity and specifity, the ROC curve determined AUC as 0.812 for Transpara v1.6 and as 0.856 for Transpara v1.7, respectively. As a result, diagnostic performance of Transpara v1.7 was higher in all cases ($P < 0.001$) (Fig. 4).

In ROC analysis performed after dividing the cases into periods, AUC values of both versions in the 1–6-month period were higher compared to the others. In this period, the AUC value was higher for Transpara v1.7 compared to v1.6. However, no statistically significant difference was found in terms of diagnostic performance in this period in both modalities (AUC Transpara v1.6 = 0.801, AUC Transpara v1.7 = 0.819; $P = 0.221$). A statistical

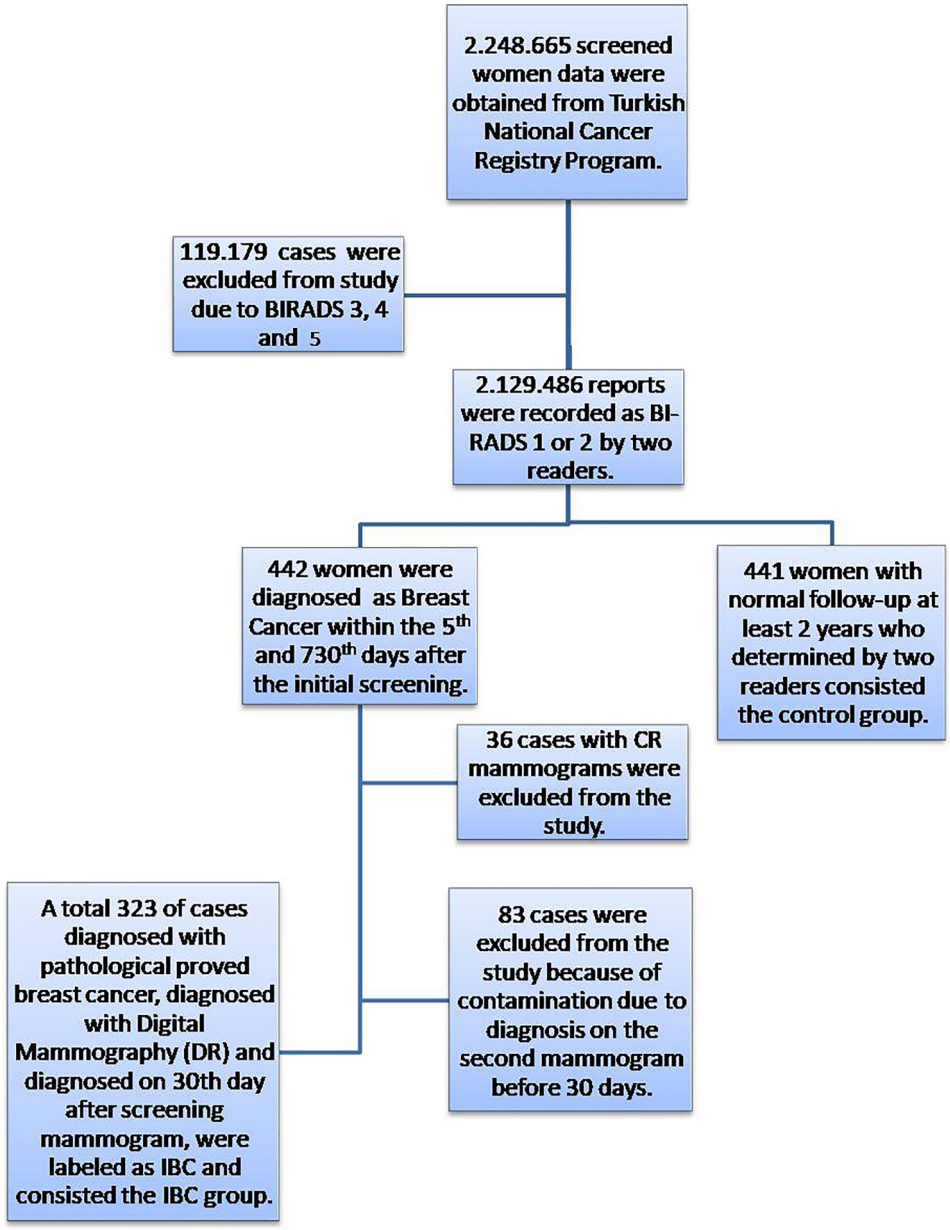

Fig. 1. Flow chart of study population. BIRADS, Breast Imaging Reporting and Data System; IBC, interval breast cancer.

difference between the diagnostic performance of the two modalities was observed at the 7–12-month period. The AUC values in this period were determined as 0.638 for Transpara v1.6 and 0.728 for Transpara v1.7, respectively, and were statistically higher in Transpara v1.7 ($P < 0.001$) (Table 1).

Discussion

The present study is the first multicenter, multireader study in the literature that has a large study population and evaluated the diagnostic performance of different versions of the same AI systems in IBC detection in a breast screening program. The study showed that Transpara v1.6 and v1.7 versions are important diagnostic tools with high specificity and sensitivity in IBC diagnosis, and the diagnostic performance of Transpara v1.7 is higher than v1.6 between the two versions.

Because the main purpose of breast screening programs is to detect the disease at an early and curable stage in the early diagnosis of breast cancer, it plays the most important role in cancer diagnosis (8,9). Studies have shown that screening programs reduce mortality of breast cancer by approximately 25% in women aged over 50 years (10)

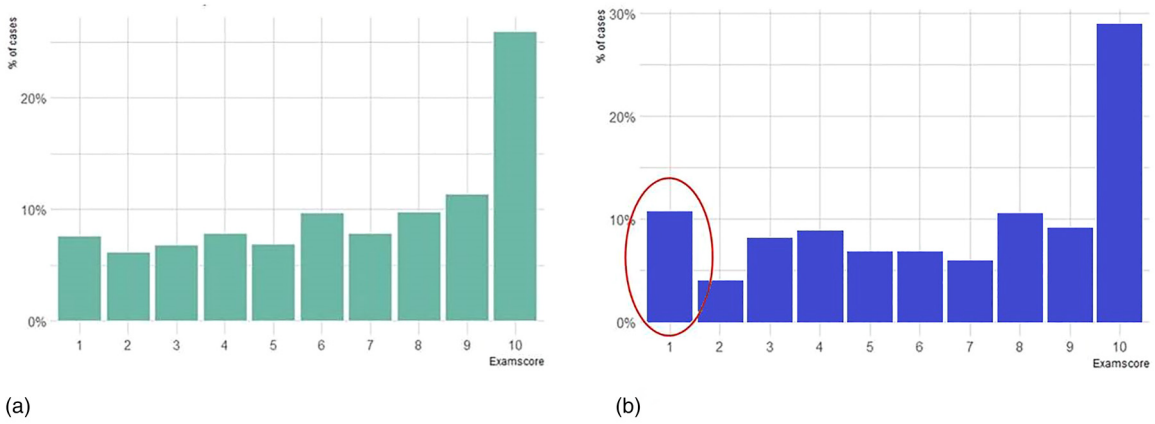

Fig. 2. (a) Distribution of Transpara score of v1.6 (green bars). Transpara 10 score of v1.6 was 26% of all cases. (b) Distribution of Transpara score v1.7 (blue bars). The red circle shows the increase in Transpara score 1 in v1.7 (11%) compared to v1.6 (8%). Transpara score 10 of v1.7 score was 29% of all cases and was also higher than v1.6.

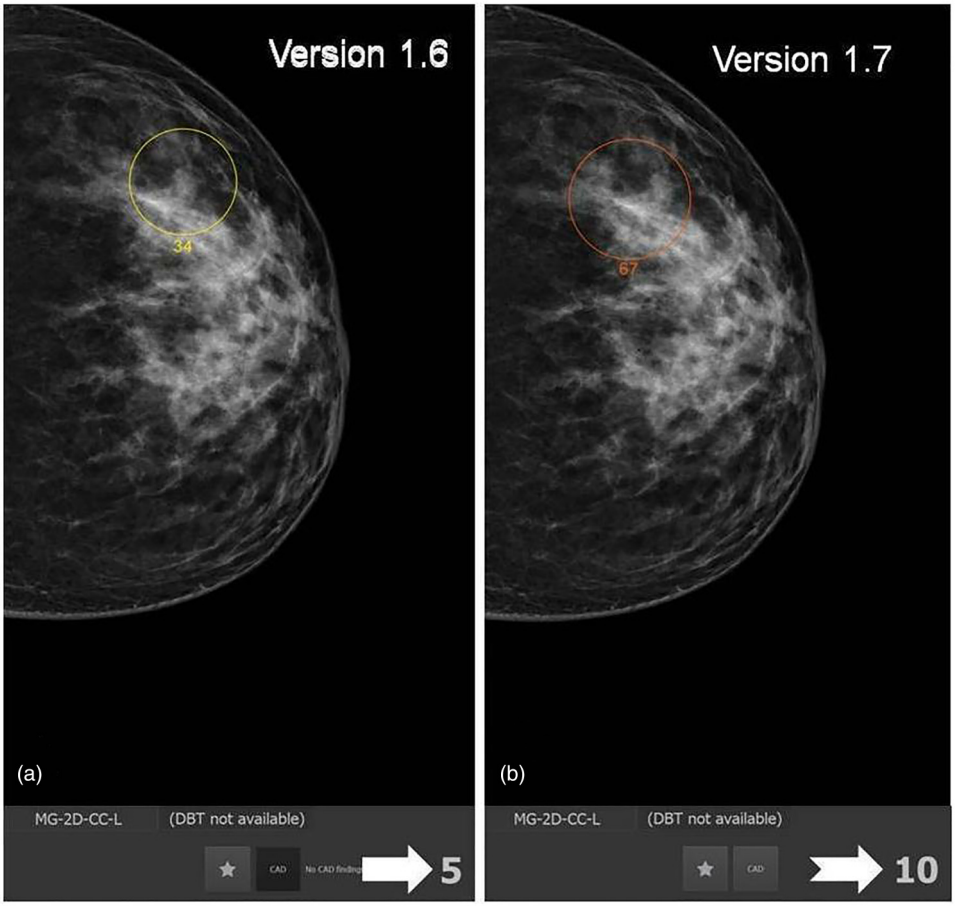

Fig. 3. Craniocaudal mammography of a 45-year-old woman shows pathology proven IBC in the left breast on the outer quadrant. (a) Transpara v1.6 could detected the lesion but did not indicate high risk (white arrow). (b) Transpara v1.7 correctly identified the lesion and labeled as high risk (white tailed arrow). IBC, interval breast cancer.

and by 15% in women aged 40–49 years (11). In the literature, the prevalence of IBC has been reported to be approximately 8.4–21.1 per 10,000 scans for a 2-year period (12). Risk factors for the development of IBC may be due to lesion-related and reader-related reasons. Reasons related to the lesion are high breast density, hormone replacement

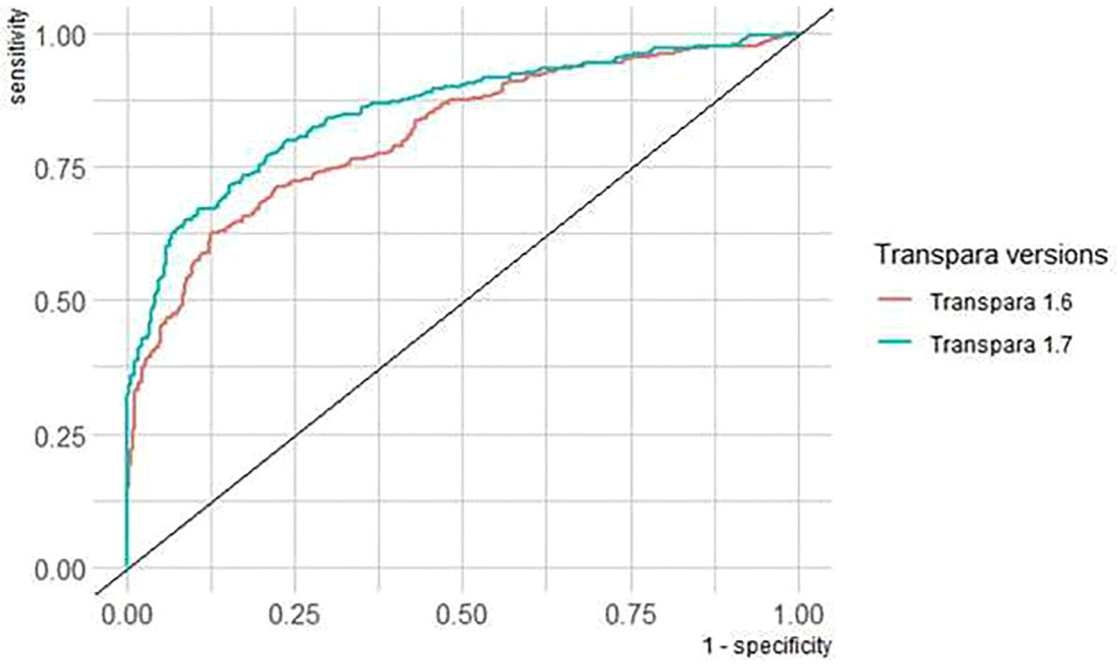

Fig. 4. ROC analysis shows increased diagnostic performance of Transpara v1.7 compared to v1.6 in all over women. AUC for v1.7 was 0.856 and AUC for v1.6 was 0.812, respectively. AUC, area under curve; ROC, receiver operating characteristic.

Table 1. Diagnostic performance of two versions of the Transpara AI system according to time periods.

		v1.6		v1.7		
Time after screening (months)	n	AUC	95% CI	AUC	95% CI	*P*
1–6	131	0.801	0.758–0.845	0.819	0.780–0.858	0.221
7–12	76	0.638	0.576–0.700	0.727	0.673–0.782	<0.001
13–18	60	0.698	0.638–0.758	0.720	0.660–0.780	0.271
19–24	56	0.596	0.523–0.668	0.576	0.502–0.650	0.429

AI, artificial intelligence; AUC, area under curve; CI, confidence interval.

therapy, family history of breast cancer and young age. Some cancers (e.g. lobular cancer) may have subtle findings in mammograms. Another reason for IBC is fast growing cancers (13). Reasons related to the reader can be due to perceptual and interpretive reasons. The most frequently missed lesions in screening mammograms can be listed as asymmetrical density, subtle distortion, microcalcification and poorly circumscribed mass lesion (14).

AI systems using multi-layered convolutional neural networks have been frequently used in breast cancer diagnosis in recent years (15). Studies in the literature showed that AI systems have sensitivity and specificity values as high as those of experienced radiologists in breast radiology in the evaluation of breast cancer. For this reason, studies agreed that AI systems are helpful as a second reader in the evaluation of breast lesions, especially with mammography (16–18). There are studies in the literature in which different versions of Transpara were used to evaluate breast cancer. In their study comparing Transpara v1.3 versus a human reader for detecting breast cancer with mammography, Sasaki et al. (7) showed that this version of Transpara has a similar sensitivity and specificity, but is lower than a human reader.

There are also studies in the literature comparing AI systems with human readers in IBC evaluation. The study by Graewingholt et al. (19), which investigated the effectiveness of AI in IBC assessment, found the sensitivity of AI for IBC detection to be approximately 48%. The present study achieved higher sensitivity and specificity values for IBC diagnosis in both versions of Transpara than the previous study. Lang et al. (6), in a study using Transpara v1.5, showed that the Transpara AI system achieved high accuracy in IBC detection and caused a decrease of approximately 19.3% in IBC rates in screening programmes and a decrease in recall rates. In their study, it was emphasized that AI is a diagnostic aid system in the detection of IBC without any additional examination (ultrasound or magnetic resonance imaging) (6).

Advances in AI systems are also depicted in the diagnostic performance of these systems. Indeed, the present study showed that the diagnostic performance of v1.7, which is the latest version of the Transpara system, has higher accuracy rates than v1.6 for all women. Both versions of Transpara had a similar high diagnostic performance for the 1–6-, 13–18- and 19–24-month periods. However, the 7–12-month period, v1.7, had a higher diagnostic performance compared to v1.6. In the 1–6-month period, and both versions had maximum diagnostic accuracy. This might be due to contamination of clinical breast cancer cases.

The present study also showed that a score of 10 for cancer risk detected by the Transpara system increased the detection of true positive cases in the latest version. This finding clearly showed the increase in sensitivity for the detection of IBC. It was also interesting to see that Transpara score 1 for cancer risk detected by v1.7 also increased compared to v1.6. This might also suggest that the latest version has an increased detection ability for true negative cases (increased specifity) of IBC. In addition to the present study, there are ongoing studies in the literature (Screen CAD-TRUST study) investigating the increase in true negative rates and hence the specificity of AI systems in breast screening programmes (20). We consider that both an increase in sensitivity and specifity will also increase the quality of detecting mammographies when using AI systems and might decrease the burden of routine radiology practice. This increase in accuracy will make the AI systems a second or third reader in breast screening programmes. The present study has some limitations. A lack of intra- and inter-observer realibility of the two readers was the main limitation. Another limitation is the risk of contamination of clinical cases within the first month after screening mammography.

In conclusion, both versions of Transpara v1.6 and v1.7 are diagnostic tools with high diagnostic performance in IBC detection. Our study has shown that, between two versions, v1.7 achieves higher accuracy rates compared to the other. The study also depicted that the more recent version of Transpara had higher sensitivity and specifity for detection of IBC. The increase in the rates of Transpara score 1 in the more recent version also allows high specifity and helps to detect true negative cases more accurately. This might provide higher accuracy in breast screening programs. We consider that developments in AI systems will increase the accuracy of detecting mammographies and help breast radiologists by reducing the burden of radiology practice in terms of screening and diagnosis in breast diseases, especially in IBC. We also consider that AI systems will be included in routine examinations as a second or third reader in the future.

Declaration of conflicting interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article

ORCID iDs

Davut Can Güner https://orcid.org/0000-0003-3141-6175
Ömer Özçağlayan https://orcid.org/0000-0001-5036-4919

References

1. Gilliland FD, Joste N, Stauber PM, et al. Biologic characteristics of interval and screen-detected breast cancers. J Natl Cancer Inst 2000;92:743–749.
2. Choi W, Cha JH, Kim H, et al. Analysis of prior mammography with negative result in women with interval breast cancer. Breast Cancer 2016;23:583–589.
3. Day N, McCann J, Camilleri-Ferrante C, et al. Monitoring interval cancers in breast screening programmes: the east anglian experience. Quality Assurance Management Group of the East Anglian Breast Screening Programme. J Med Screen 1995;2:180–185.
4. 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;111:916–922.
5. Jonmarker O, Strand F, Brandberg Y, et al. The future of breast cancer screening: what do participants in a breast cancer screening program think about automation using artificial intelligence? Acta Radiol Open 2019;8:2058460119880315.
6. Lång K, Hofvind S, Rodríguez-Ruiz A, et al. Can artificial intelligence reduce the interval cancer rate in mammography screening? Eur Radiol 2021;31:5940–5947.
7. Sasaki M, Tozaki M, Rodríguez-Ruiz A, et al. Artifcial intelligence for breast cancer detection in mammography: experience of use of the ScreenPoint Medical Transpara system in 310 Japanese women. Breast Cancer 2020;27:642–651.
8. Moss SM, Cuckle H, Evans A, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet 2006;368:2053–2060.
9. Løberg M, Lousdal ML, Bretthauer M, et al. Benefits and harms of mammography screening. Breast Cancer Res 2015;17:63.
10. Nystrom L, Andersson I, Bjurstam N, et al. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 2002;359:909–919.
11. Smith RA, Duffy SW, Gabe R, et al. The randomized trials of breast cancer screening: what have we learned? Radiol Clin North Am 2004;42:793–806.
12. Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. NPJ Breast Cancer 2017;13:3–12.
13. Meshkat B, Prichard RS, Al-Hilli Z, et al. A comparison of clinicalepathological characteristics between symptomatic and interval breast cancer. Breast 2015;24:278–282.
14. Cornford E, Sharma N. Interval cancers and duty of candour, a UK perspective. Curr Breast Cancer Rep 2019;11:89–93.
15. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast

tomosynthesis: state of the art. Semin Cancer Biol 2021; 72:214–225.
16. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology 2019;290:305–314.
17. Tagliafico AS, Piana M, Schenone D, et al. Overview of radiomics in breast cancer diagnosis and prognostication. Breast 2020;49:74–80.
18. Lång K, Josefsson V, Larsson AM, 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;24:936–944.
19. 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: 369–371.
20. Strand F, et al. Artificial intelligence in large-scale breast cancer screening (ScreenTrustCAD). ClinicalTrials.gov Identifier: NCT04778670.