

COMMENTARY



AI in breast cancer screening: a critical overview of what we know

José Luis Raya-Povedano^{1,2,3*} 

Population-based breast cancer screening programs with mammography have proven to be the most effective method for reducing breast cancer mortality (by up to 30–40% of participating women) and are implemented in most European countries.

Despite their undoubted benefits, they are not free of problems. The main ones are false negatives (cancers not detected in screening readings) and false positives (recalls due to benign findings). The reading of screening mammograms is a very heavy workload for examinations that, in most cases, will be normal. Double reading and tomosynthesis, used to reduce false negatives, multiply this workload. This problem is even greater in a setting where there is a lack of expert breast radiologists.

Artificial intelligence (AI) applied to breast imaging, with new algorithms based on deep learning, has undergone enormous development in recent years. These algorithms are capable not only of identifying lesions in mammography and tomosynthesis studies, but also of assigning a degree of suspicion to each finding and to the overall study. This capability allows sorting and classifying the studies according to the likelihood of cancer being present.

The main utility of these AI algorithms in breast cancer screening lies in their ability to reduce the reading workload by fully or partially replacing human reading. But to be safely applied to the screening workload, it should not lead to a decrease in cancer detection or an increase in false-positive recalls.

In the last four years, multiple retrospective simulated studies have been published studying this ability to safely reduce workload in long screening series. These studies involve different devices, different AI commercial software, and different countries. The first conclusion that can be drawn from them is that AI adequately classifies studies according to cancer risk. In a series of 122,012 mammograms and 752 cancers diagnosed in screening, Larsen et al [1] found that 86.8% of the cancers were in the 10% of studies classified by the AI as the highest risk and only 4.4% in the 70% of those with the lowest suspicion. Other publications obtained similar results. Knowing this distribution of cancers, these studies simulated different ways of applying the AI and compared the results with the original readings.

Some of these studies assess the ability to consider negative and dismiss low-risk studies for human readings. The retrospective study published by our team in 2021 [2] concluded that excluding for reading 70% of these studies considered by the AI as low risk did not result in a loss of sensitivity in the detection of cancers or an increase in false-positive recalls.

Other retrospective studies simulated various combinations in which AI was used to replace human readers fully or partially [3–5]. In general, the conclusion of these publications is that the best application of AI is in combination with human readings, either by replacing one reader [3, 4] or by including AI in a decision algorithm involving the reader [5].

Some of these studies also included interval cancers. Lang et al [6] demonstrated that most of the interval cancers considered false negative were retrospectively classified by the AI in the highest risk scores and could potentially have been detected prospectively with the help of the AI.

Evidence about the performance of AI in digital breast tomosynthesis (DBT) is less robust than in digital mammography (DM), and the results are, to date, worse. Our retrospective publication about stand-alone use of AI in DM

*Correspondence:

José Luis Raya-Povedano
joseluisrayapovedano@gmail.com

¹ Maimónides Biomedical Research Institute of Córdoba (IMIBIC), Córdoba, Spain

² Breast Cancer Unit, Department of Diagnostic Radiology, Reina Sofía University Hospital, Menéndez Pidal Avenue s/n, 14004 Córdoba, Spain

³ University of Córdoba, Córdoba, Spain

and DBT [7] concluded that in DBT, the performance was inferior to that of the original readings. Several reasons have been mentioned for the inferiority of AI in DBT. First, there are fewer sets of tomosynthesis studies to train the algorithms. Second, technically, the analysis of tomosynthesis is more complex, and third, there are many differences in the characteristics of the images obtained by the different manufacturers, which limits the extrapolation of the results.

Although the studies published to date report better performance of AI algorithms in DM than in DBT, the conclusions of two retrospective simulated studies in paired DM-DBT screening series [2, 8] suggest that AI-assisted DBT screening, excluding low-risk studies from reading, could replace DM screening with a lower reading workload, higher cancer detection, and fewer false-positive recalls.

The main limitation of these retrospective studies is that, although they are based on unenriched series, decisions are simulated, and it is not possible to know what the reader's behavior would have been if they had known the result of the AI when reading.

Two recently published prospective studies support the results of the retrospective studies. In the preliminary results of their randomized study (MASAI), Lang et al [9] demonstrated that detection of cancers was non-inferior and recall rate of false positives was not higher when 90% of lower suspicion studies were single read and only 10% of higher suspicion studies were double read. In the second, Dembrower et al [10] demonstrated higher sensitivity in detecting cancers when the AI was used to replace the second reader as compared to human double reading. The number of studies sent to consensus was higher, but not the final recall rate.

Prospective studies provide strong support for the incorporation of AI in breast cancer screening, with some limitations. Their good results are the combination of the automatic behavior of AI with the final decision of radiologists trained in the use of AI and aware of its strengths and weaknesses and may not be generalizable to other centers.

The application of AI in screening opens great possibilities and poses some challenges. It will facilitate the extension of DBT screening and may help to extend screening to populations where there is a lack of breast radiologists. The use of AI will also demand expert radiologists and advanced diagnostic techniques to manage recalled woman because of increasingly subtle findings. Finally, ethical and legal issues cannot be overlooked, particularly when AI is used to exclude human reading from all or part of the studies.

Funding

The author states that this work has not received any funding.

Declarations

Guarantor

The scientific guarantor of this publication is José Luis Raya-Povedano.

Conflict of interest

The author of this manuscript declares no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry

No complex statistical methods were necessary for this paper.

Informed consent

Written informed consent was not required for this study because it is a commentary.

Ethical approval

Institutional Review Board approval was not required because it is a commentary.

Study subjects or cohorts overlap

Not applicable

Methodology

• commentary

Received: 8 November 2023 Revised: 8 November 2023

Accepted: 10 December 2023 Published: 21 December 2023

References

1. Larsen M, Aglen CF, Lee CI et al (2022) Artificial intelligence evaluation of 122 969 mammography examinations from a population-based screening program. *Radiology* 303(3):502–11
2. Raya-Povedano JL, Romero-Martín S, Elías-Cabot E, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M (2021) AI-Based strategies to reduce workload in breast cancer screening with mammography and tomosynthesis: a retrospective evaluation. *Radiology* 300:57–65
3. Salim M, Wählin E, Dembrower K et al (2020) External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncol* 6(10):1581–8
4. McKinney SM, Sieniek M, Godbole V et al (2020) International evaluation of an AI system for breast cancer screening. *Nature*. 577(7788):89–94
5. Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L (2022) Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *Lancet Digital Health* 4(7):e507–19
6. Lång K, Hofvind S, Rodríguez-Ruiz A, Andersson I (2021) Can artificial intelligence reduce the interval cancer rate in mammography screening? *Eur Radiol*. 31(8):5940–7
7. Romero-Martín S, Elías-Cabot E, Raya-Povedano JL, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M (2022) Stand-alone use of artificial intelligence for digital mammography and digital breast tomosynthesis screening: a retrospective evaluation. *Radiology*. 302(3):535–42
8. Dahlblom V, Dustler M, Tingberg A, Zackrisson S (2023) Breast cancer screening with digital breast tomosynthesis: comparison of different reading strategies implementing artificial intelligence. *Eur Radiol*. 33(5):3754–65
9. Lång K, Josefsson V, Larsson AM et al (2023) 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*. 24(8):936–44
10. Dembrower K, Crippa A, Colón E, Eklund M, Strand F, Consortium ST (2023) Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Heal*. [https://doi.org/10.1016/S2589-7500\(23\)00153-X](https://doi.org/10.1016/S2589-7500(23)00153-X)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.