Radiología 63 (2021) 236–244

RADIOLOGÍA

www.elsevier.es/rx

UPDATE IN RADIOLOGY

Are artificial intelligence systems useful in breast cancer screening programmes?☆

O. Díaz [a], A. Rodríguez-Ruiz [b], A. Gubern-Mérida [b], R. Martí [c], M. Chevalier [d,e,*]

[a] *Departamento de Matemáticas e Informática, Universidad de Barcelona, Barcelona, Spain*
[b] *ScreenPoint Medical BV, Nijmegen, Netherlands*
[c] *Instituto de Visión Artificial y Robótica (VICOROB), Universitat de Girona, Girona, Spain*
[d] *Física Médica, Departamento de Radiología, Rehabilitación y Fisioterapia, Universidad Complutense de Madrid, Madrid, Spain*
[e] *Instituto de Investigación Sanitaria, Hospital Clínico San Carlos, Madrid, Spain*

Received 16 August 2020; accepted 16 November 2020

KEYWORDS
Mammography;
Artificial intelligence;
Deep learning;
CAD;
Breast cancer screening

Abstract Population-based breast cancer screening programmes are efficacious in reducing the mortality due to breast cancer. These programmes use mammography to screen the women who are invited to participate. Digital mammography makes it possible to develop computer-assisted diagnosis (CAD) systems that promise to reduce the workload of radiologists participating in screening programmes. However, various studies have shown that CAD results in a high rate of false positive diagnoses. Systems based on artificial intelligence are being more widely implemented, and studies have shown that these systems have better diagnostic performance than traditional CAD systems.

This article explains the fundamentals of artificial intelligence systems and an overview of possible applications of these systems within the framework of breast cancer screening programmes.

PALABRAS CLAVE
Mamografía;
Inteligencia artificial;
Aprendizaje profundo;
CAD;
Cribado de cáncer de mama

¿Son los sistemas de inteligencia artificial una herramienta útil para los programas de cribado de cáncer de mama?

Resumen Los programas poblacionales de detección precoz del cáncer de mama (PDPCM) contribuyen eficazmente a la reducción de la mortalidad debida a esta patología. Estos programas usan la mamografía para el seguimiento de las mujeres invitadas a participar. La mamografía digital posibilitó el desarrollo de sistemas de diagnóstico asistido por ordenador (CAD) con importantes expectativas para reducir la carga de trabajo de los radiólogos participantes en

☆ Please cite this article as: Díaz O, Rodríguez-Ruiz A, Gubern-Mérida A, Martí R, Chevalier M. ¿Son los sistemas de inteligencia artificial una herramienta útil para los programas de cribado de cáncer de mama? Radiología. 2021;63:236–244.

* Corresponding author.
E-mail address: chevalie@ucm.es (M. Chevalier).

dichos programas. Distintas investigaciones evidenciaron la obtención de un número importante de falsos positivos con los CAD. La creciente implementación de sistemas basados en inteligencia artificial ha propiciado el desarrollo de investigaciones sobre su rendimiento en mamografía que muestran su superioridad frente a los CAD tradicionales. En este artículo se describen los fundamentos de los sistemas de inteligencia artificial y se aporta una panorámica de sus posibles aplicaciones en el marco de los programas de cribado de mama.

Introduction

Breast cancer is the most common malignant tumour among women with an annual incidence of more than 2 million cases worldwide and a mortality rate of 30%.[1] In Spain, it is estimated that there were 32,800 new cases in 2020, with a mortality rate of 19.5%. Early detection of breast cancer is vital for reducing mortality[2,3] and improving the prognosis and quality of life of women after treatment. Breast cancer screening or early detection programmes (BCEDPs) aimed at asymptomatic women have been organised in most countries for decades.[4,5] Mammography is the diagnostic test on which these programmes are based due to its high sensitivity for cancer detection, as well as its adequate cost-benefit balance. The target population for most BCEDPs are women aged between 50 and 69 years and this population is usually screened every 2 years, in most cases with two mammograms per breast. The effectiveness of BCEDPs is limited by various factors, with the main one being the use of a 2D technique, such as a mammography, that results in tissue overlap and reduces detection capacity.[6–8] Other significant factors are the high workload due to the high number of mammographies (the majority normal) and the double reading procedure.[9] This procedure, adopted to increase the cancer detection rate, has recently been hampered by the growing shortage of radiologists.[10,11] These limitations influence, among other things, the rate of interval cancers (cancers diagnosed symptomatically after a negative screening test and before the next round of screening). According to the studies, these cancers affect approximately 2 per 1000 screenings in European BCEDPs.[7] The use of breast tomosynthesis has led to a notable increase in the breast cancer detection rate.[12–14] However, its inclusion in BCEDPs has been questioned as it increases the workload and reading time.[15,16]Computer-aided detection (CAD) systems first appeared in the 1990s as a diagnostic support tool for mammography. These systems automatically process the images and highlight the areas considered suspicious (Fig. 1) to reduce the number of lesions that are not visualised. There are two categories of CAD algorithms: those used to detect the presence of a lesion and those that also indicate whether the lesion is benign or malignant. In both cases, the algorithms ''search'' the image for characteristics specific to the lesions previously defined by the radiologists. The main drawback of CAD systems is their moderate specificity, meaning they mark a lot of areas as suspicious leading to false positives or negatives from radiologists.[17–20] Recent developments in the field of artificial intelligence (AI) have opened up new horizons in the field of diagnostic support systems. The results of numerous research projects have shown that systems based on deep learning (DL) perform better than traditional CAD systems.

This paper aims to provide an overview of the applications of AI systems for breast cancer detection and their roles in the context of BCEDPs. First, we look at the most relevant basic concepts that the AI algorithms are based on.

Artificial intelligence: evolution of machine learning techniques

Concepts such as DL and ''machine learning'' (ML) are commonplace today in a wide range of fields, including radiology.[21] Their use in diagnostic imaging is genuinely revolutionary, as the significant increase in the number of publications and scientific sessions devoted to the topic demonstrates. AI studies the incorporation of human intelligence behaviours, for example learning, in machines. The field of AI includes, for example, robotics and ML (Fig. 2), which is the element that gives machines with the ability to learn a task from previous experience without needing specific programming. In contrast, DL[22,23] comprises machine learning techniques in which the algorithm learns by itself which features of an image are the most important for performing a particular task. The learning architectures that are currently most widely used in DL are neural networks, which are interconnected nodes that form multiple layers that mimic biological cerebral neural networks. Once trained, the networks assign a weight to each node (image feature) to highlight its importance to the network's final task (segmentation, detection, etc.).

Within the DL, convolutional neural networks (CNNs) are a subset of algorithms based on convolution operations that started to be used in the early 1990s.[24] The first scientific work in which a CNN applied to mammography was used was published in 1996.[25] It was not until this decade that this technique was used on a massive scale driven by the availability of a large number of digital medical images, improved computing power and the decrease in the price of graphics processing units.

CNNs automatically learn the combination of image features that make it possible to perform a given task without the need for prior knowledge or, therefore, human intervention. This is the fundamental difference between CNNs and

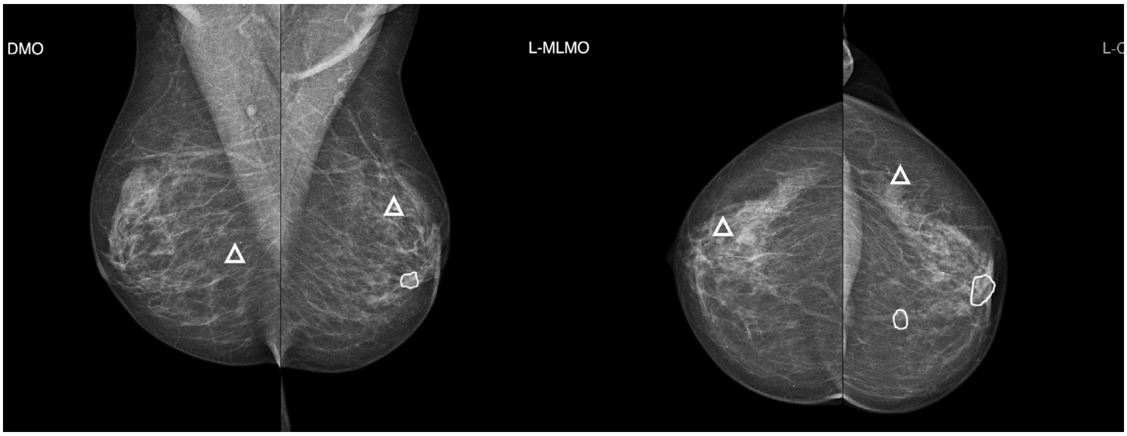

Figure 1 Example of a traditional computer-aided diagnosis system. It automatically processes the images and displays markings to the reader in several areas that the system considers suspicious.

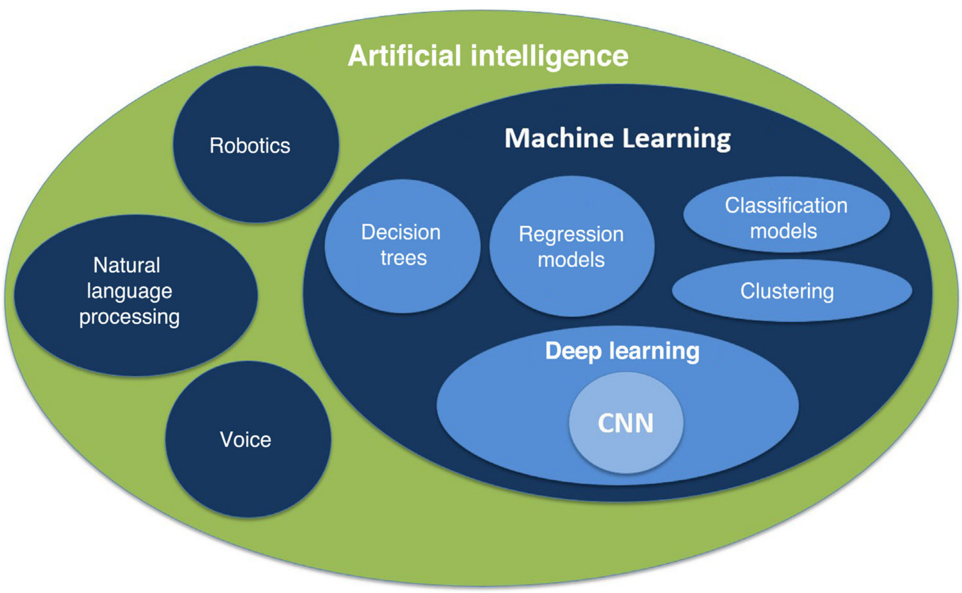

Figure 2 Diagram showing the areas covered by artificial intelligence. The acronym CNN stands for ''convolutional neural networks''.

traditional CAD systems. CNNs trained with a large number of cases perform much better than CAD systems and avoid the biases associated with the use of hand-picked features.[26]

Basic concepts in artificial intelligence

Within ML, two main families of algorithms are recognised: supervised learning and unsupervised learning. In the former, a task is learned from a set of cases with a known outcome (ground truth). In the field of radiology, these would be images with a known label depending on the problem being tackled: localisation of a lesion (detection), definition of the border of the lesion (segmentation) or diagnosis of the lesion (classification). Unsupervised learning models use unlabelled images that are grouped according to some common features. The lack of a large number of labelled images is one of the biggest current problems for supervised learning models. This has motivated increased research into unsupervised models and the proposal of algorithms that combine both paradigms. These days, there are public databases of labelled images, with the most common in digital mammography being INbreast[27] and CBIS-DDSM.[28] There are also private databases, such as OMI-DB.[29] A lot of DL architectures require an image as an input element, while others work from fragments or regions of interest, which enables image weights to be increased without having to have a large computing capacity.

Supervised algorithm learning is divided into three phases:

1. Training: the model's parameters are optimised based on a previously determined error that we are trying to minimise. The error is obtained by comparing the model's prediction with the actual label associated with the set of training images.

2. Validation: the model obtained after the training is used to predict the labels of a second set of images. The results obtained with this set are also used to perform an initial validation of the model and to adjust parameters and features to optimise the model.
3. Testing: the performance of the final model is evaluated using a set of new, previously unseen, images. The results of this analysis with this set of images determine the model's accuracy and generalisability.

The set of images used for the validation and evaluation should have a distribution of cases that is as similar as possible to the distribution of cases that the algorithm will evaluate in its final application (images from equipment from different manufacturers, proportion of cases with lesions observed in population screening, etc.).

The smaller size of the databases in radiology is a significant problem that we try to address by increasing the set of training images (data augmentation) by applying different techniques. The most popular technique is deforming the images, with rotation being the most common method. Other techniques, such as transfer learning, use everyday images to train the DL model by taking advantage of the existence of common features in images from different domains. Its use in mammography has led to a slight improvement in DL performance.[30] More recently, neural networks capable of generating synthetic images and thus enriching the training set (GANs,[31] generative adversarial networks) have been introduced.[32] These networks were originally proposed as unsupervised learning algorithms and their use has spread to semi-supervised learning models.

CNNs process images through multiple sequential stages known as layers, in which different mathematical procedures are applied. Firstly, the images are preprocessed, normalising their intensity to make the learning of the network independent of the origin of the images (different equipment, etc.). The next stages involve different procedures; the most common are convolutions, the reduction of dimensions (max pooling) and renormalisation (softmax), among others. There are CNNs with different architectures that differ in aspects such as number of layers, the activation functions, etc.

Artificial intelligence algorithms applied to medical imaging

The lack of large public databases has limited the clinical application of AI in medical imaging even though the current picture archiving and communication systems (PACSs) represent a great source of information. Nevertheless, a large number of very useful applications have been developed for diagnostics in general and, specifically, the diagnosis of breast pathologies.

The most popular task for AI applications is the automatic detection of lesions in different imaging modalities. This task involves locating regions in the image which have a high probability of being lesions and being able to detect different types of lesions, depending on the training that the algorithm has received. It has recently been demonstrated that CNNs can achieve the detection rates of an average radiologist in mammography.[33] Once the lesion has been detected, the CNN can classify it as malignant or benign or identify what type of lesion it is (nodule, microcalcification, etc.). The classification by the CNN has the advantage of eliminating the variability that may exist between observers. Another of the most common applications of CNNs is tissue or lesion segment (definition of the borders).[34,35] The most popular networks for this purpose are U-nets that enable the differentiation of tissues (glandular or adipose tissue or pectoral muscle in mammography images) and the calculation of the volume of a lesion.[35] Image registration is used to analyse the changes in images from the same patient taken at different times (longitudinal registration) or taken using different imaging modalities (multimodal registration). Trying to simulate the way humans perceive changes in an image is a complex task for CNNs, but the results obtained show an increase in the speed and accuracy when performing this task.[36]

Artificial intelligence as a tool to improve breast cancer screening

Recent advances in AI open up possibilities that go beyond those offered by traditional CAD systems when it comes to supporting radiologists in mammography image reading in BCEDPs. The objectives of AI systems are the same as those of CAD systems: improving the detection of malignant lesions, reducing interval cancers and, at the same time, reducing the reading load. Eventually, it may be the case that new AI systems and techniques even improve the cost-benefit ratio of BCEDPs.[37]

Performance of artificial intelligence systems in the detection of cancer in mammography

The first question to ask is whether AI systems are sufficiently accurate in detecting cancer in mammography. We would expect them to be, given that these systems are based on machine learning techniques and are trained with databases that include thousands of mammography images with a wide variety of clinical case material. Research has shown that these systems are more accurate than traditional CAD systems.[38] Kim et al.[39] studied the performance of an AI system using examinations from a screening population (1200 examinations, half with cancer), obtaining a sensitivity of 76% and a specificity of 89%, which is much higher than the values for traditional CAD systems (around 50%). Studies in which the performance of AI systems and that of radiologists is compared have shown a performance inferior or similar to that of the radiologists when using enriched samples[40] or screening examinations.[41] More recently, studies have been published comparing the same AI system against 14 radiologists using a sample of 240 cases (100 with cancer)[42] and against 101 radiologists using a sample of 2600 cases (approximately 650 with cancer).[33] In both cases, the AI system achieved a performance that was statistically similar to that of the average of all the radiologists in terms of the area under the curve (AUC) of the ROC (receiver operating characteristic) curve (Fig. 3).

The success of an BCEDP depends, to a significant extent, on the uniformity of the performance of the radiologists involved. AI systems would reduce the notable differences

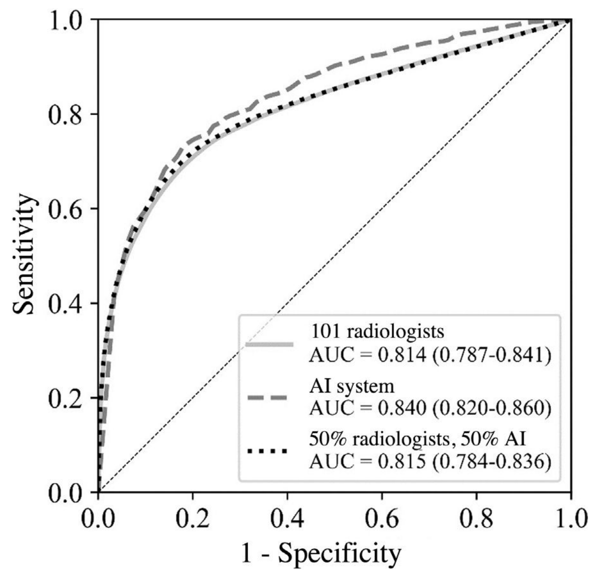

Figure 3 ROC (receiver operating characteristics) curves obtained in a study with 101 radiologists and 2600 cases, 650 of them with cancer. The mean ROC curve of the 101 radiologists (continuous line), the curve obtained with an artificial intelligence system (dashed curve) and the ROC curve obtained in a scenario where 50% of the cases were read by the artificial intelligence system and the other 50% by the 101 radiologists (dotted curve) are shown. The AUC (area under the curve) parameter represents the area under the ROC curve. Courtesy of A. Rodríguez-Ruiz.[33]

that exist between professionals in terms of sensitivity, which, according to a recent study, ranges from 53% to 89% and varies depending on the type of lesion.[43]

Although the results are promising, it would still be possible to improve the performance of the AI algorithms to reach, at least, the level of the best human reader with an accuracy that is independent of the image quality of the mammograms produced by equipment from different manufacturers.[44,45] This requires continuous improvement of algorithms to emulate the complex procedures that radiologists follow during routine reading of mammograms, such as comparisons with the contralateral breast and with examinations from the previous round of screening.

Artificial intelligence as a support tool for radiologists in BCEDPs

As already mentioned, traditional CAD systems perform poorly in terms of detecting cancer due to their low specificity, and to this we have to add the absence of interaction between the radiologist and the system, which only displays markings. Both aspects result in an inadequate tool,[17] which can also generate significant problems when not used correctly.[18] The new generation of AI-based CAD systems for mammography aims to prevent up to one third of malignant lesions from going undetected[46] during the reading of screening mammograms. To achieve this, in addition to showing markings in the suspicious regions (with fewer false positives than a traditional CAD system), some systems introduce new concepts such as interactive support (Fig. 4) or a score for the examination as a whole that indicates the level of cancer suspicion.[42,47]

The first study published on the impact of the use of an AI system with interactive support, CAD system markings and a score for the examination as a whole on the performance of radiologists shows an improvement in both sensitivity and in specificity of the 14 participating radiologists with a sample of 240 cases, of which 100 had cancer.[42] Although the increase in performance was moderate (3%), it should be noted that this was independent of the radiologists' experience. This study also showed that the reading time per case did not increase with the AI system, in contrast to what had been observed in previous studies using traditional CAD systems where increases of up to 20% per case had been observed.[48]

All these promising results have to be compared with prospective studies in BCEDP populations in order to investigate in depth the effect of AI systems on improving radiologist performance.[49] This is an essential step for measuring the efficiency of AI algorithms in a realistic environment.

Artificial intelligence as an autonomous reader in BCEDPs

In the previous sections, we have discussed the results of the application of AI systems when they are used in parallel with radiologists, in other words when they are being used as a CAD system. However, these systems can do more than that and be used as an autonomous reader since they have the capacity to categorise mammograms according to their probability of showing cancer. Those cases classified by the algorithm as more suspicious would be reviewed by radiologists first, preventing lesions from going undetected and enabling the woman to be recalled as soon as possible. Considering this capability and that more than 95% of women in the screening population do not have an abnormality of any kind, it is possible that one of the main uses of AI systems in the future will be as a first reader. Its role would be to automatically classify the examinations, reducing the workload by preventing two radiologists from having to read normal examinations, if double reading is being used.

The use of AI systems as autonomous readers acting as first, second or even third readers of mammography examinations would significantly reduce the workload involved in BCEDPs. In this context, the biggest challenge for an AI system would be to automatically provide a recall decision based, for example, on the score assigned to the examination once the system's high accuracy has been proven.[43,48] In a recent retrospective study,[50] in which 8000 examinations from a multicentre BCEDP from the United Kingdom were used, the AI system (trained with some of these data) was used as an autonomous reader of the examinations with a lower likelihood of cancer and the rest were read by a radiologist. The results permitted the conclusion that the use of AI could lead to a reduction of the workload of 42.8% without causing an increase in false positives or false positives in the programme. In another retrospective study, Rodríguez-Ruiz et al.[51] analysed the possibility of using an AI system trained with independent data to read the cases with a lower probability of cancer from a screening sample

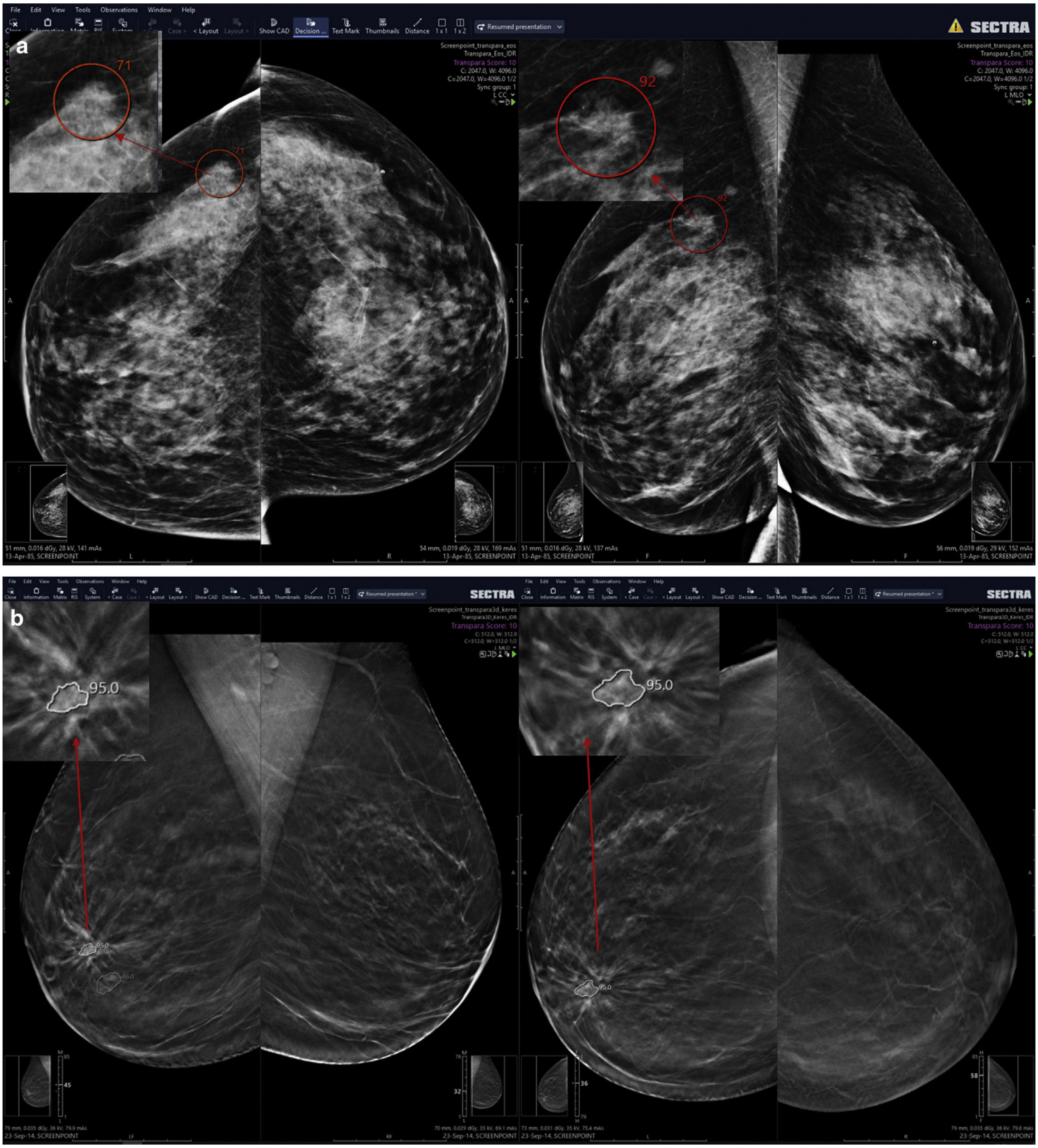

Figure 4 Example of an artificial intelligence system that provides interactive mammography support beyond the functionality of traditional computer-aided diagnosis systems. After indicating the area of interest, the system automatically displays the level of suspicion in the two views (craniocaudal and oblique mediolateral oblique) as can be seen in the inserts included in the cases (a) group of microcalcifications and (b) distortion. Courtesy of N. Karssemeijer. Radboud University Medical Centre, Nijmegen, The Netherlands.

of 2600 cases. The sample contained 650 cases with cancer and was enriched with images from different manufacturers. The researchers found that the AI system reduced the number of screening cases that had to be read by 47%, with a 27% decrease in false positives, although the false negatives increased by 7%. Expressed in a more general way, these results indicate that reducing the number of cases to be read in a screening programme by 20% leads to a 5% decrease in false positives and an increase of less than 1% in false negatives.

Artificial intelligence in tomosynthesis and future applications

We have already mentioned AI's great potential for reducing the reading time of tomosynthesis examinations.[15,16]

The task of the AI system in this scenario would be to search for and automatically show the radiologist the tomographic slice with the region in which the abnormality is located. In recently published studies, it has been shown that reductions of between 15% and 50% in reading time per examination can be achieved.[52–54] This reduction would favour the inclusion of tomosynthesis in BCEDPs. These studies also provided evidence of the possibility of using the AI system to generate a synthetic image in which the lesions detected in the slices are enhanced, thus significantly increasing the performance of these images.[54]

In addition to helping radiologists detect cancer or differentiate malignant lesions from benign lesions, AI systems could also differentiate invasive lesions from other lesions, such as ductal carcinomas in situ (DCIS), or determine which DCIS will eventually become invasive.[55,56] Despite the difficulty of this task, AI systems have the ability to analyse features that may be hidden from the human eye based on radiomics. The automatic quantification of the risk that parameters such as breast density[57,58] and other glandular tissue patterns[59] could represent is another potential advanced use of AI systems.

As demonstrated in the study by Wang et al.,[60] the use of AI also makes it possible to detect other pathologies in mammograms, for example, vascular lesions from an analysis of vascular calcifications in the breast, with an accuracy similar to that of expert radiologists.

Conclusions

AI systems perform better than traditional CAD systems with a similar sensitivity and a very significant reduction in the number of false positives. These systems also make it possible to categorise mammography examinations according to their likelihood of containing cancer. This capability is an important tool for reducing the workflow involved in BCEDPs and creates significant expectations for its use as a first reader.

AI systems also play an important role in reducing the variability in sensitivity between radiologists when they are reading a high volume of mammograms from BCEDPs. This aspect opens the door to the AI system being considered a reference for evaluating the effectiveness of BCEDPs once their capacity has been demonstrated.

To date, research into the autonomous performance of AI systems has been based on retrospective studies and, in many cases, has used enriched samples. The analysis of the independent use of these systems in more realistic scenarios and, specifically, in environments associated with BCEDPs is necessary to evaluate their real effectiveness.

The applications of AI algorithms in the context of medical imaging have to take into account ethical and medical-legal aspects, and these have to be discussed and be well-established. One of the most salient issues to be addressed is the question of liability in the event of system failure.

Authorship

1. Responsible for study integrity: OD, ARR, AGM, RM, MCH.
2. Study concept: OD, ARR, AGM, RM, MCH.
3. Study design: OD, ARR, AGM, RM, MCH.
4. Data collection: OD, ARR, AGM, RM, MCH.
5. Data analysis and interpretation: OD, ARR, AGM, RM, MCH.
6. Statistical processing: OD, ARR, AGM, RM, MCH.
7. Literature search: OD, ARR, AGM, RM, MCH.
8. Drafting of the article: OD, ARR, AGM, RM, MCH.
9. Critical review of the manuscript with intellectually significant contributions: OD, ARR, AGM, RM, MCH.
10. Approval of the final version: OD, ARR, AGM, RM, MCH.

Conflicts of interest

R. Martí, O. Diaz y M. Chevalier are part of the research team for the ICEBERG project (''Image Computing for Enhancing Breast Cancer Radiomics'' [Ref. RTI2018-096333-B-I00]) financed by the Spanish Ministry of Science, Innovation and Universities. A. Rodríguez Ruiz and A. Gubern Merida are employees of ScreenPoint Medical BV, Nijmegen, The Netherlands.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
2. Sociedad Española de Oncología Médica (SEOM); 2020. Available from: https://seom.org/seomcms/images/stories/recursos/Cifras_del_cancer_2020.pdf [accessed 31.07.20].
3. Sardanelli F, Aase HS, Álvarez M, Azavedo E, Baarslag HJ, Balleyguier C, et al. Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey. Eur Radiol. 2017;27:2737–43.
4. Tabár L, Vitak B, Hsiu-Hsi Chen T, Ming-Fang Yen A, Cohen A, Tot T, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology. 2011;260:658–63.
5. Mellado Rodríguez M, Osa Labrador AM. Cribado de cáncer de mama, Estado actual. Radiología. 2013;55:305–14.
6. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med. 2007;146:502–10.
7. 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.
8. Carreira Gómez MC, Estrada Blan MC. Mama densa ¿qué debemos saber? Implicaciones en el cribado. Radiología. 2016;58:421–6.
9. Karssemeijer N, Otten JD, Roelofs AA, van Woudenberg S, Hendriks JH. Effect of independent multiple reading of mammograms on detection performance. In: Chakraborty DP, Eckstein MP, editors. SPIE 5372, medical imaging 2004: image perception, observer performance, and technology assessment. Bellingham: Society of Photo-Optical Instrumentation Engineers (SPIE); 2004. p. 82–9.
10. Rimmer A. Radiologist shortage leaves patient care at risk, warns royal college. BMJ. 2017;359:j4683.

11. Wing P, Langelier MH. Workforce shortages in breast imaging: impact on mammography utilization. Am J Roentgenol. 2009;192:370–8.
12. Skaane P. Breast cancer screening with digital breast tomosynthesis. Breast Cancer. 2017;24:32–41.
13. Martín SR, Povedano JLR, García MC, Romero ALS, Garriguet MP, Benito MÁ. Prospective study aiming to compare 2D mammography and tomosynthesis+ synthesized mammography in terms of cancer detection and recall. From double reading of 2D mammography to single reading of tomosynthesis. Eur Radiol. 2018;28:2484–91.
14. Hofvind S, Hovda T, Holen ÅS, Lee CI, Albertsen J, Bjørndal H, et al. Digital breast tomosynthesis and synthetic 2D mammography versus digital mammography: evaluation in a population-based screening program. Radiology. 2018;287:787–94.
15. Rodriguez-Ruiz A, Gubern-Merida A, Imhof-Tas M, Lardenoije S, Wanders AJT, Andersson I, et al. One-view digital breast tomosynthesis as a stand-alone modality for breast cancer detection: do we need more? Eur Radiol. 2018;28:1938–48.
16. Dang PA, Freer PE, Humphrey KL, Halpern EF, Rafferty EA. Addition of tomosynthesis to conventional digital mammography: effect on image interpretation time of screening examinations. Radiology. 2014;270:49–56.
17. Katzen J, Dodelzon K. A review of computer aided detection in mammography. Clin Imaging. 2018;52:305–9.
18. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med. 2007;356:1399–409.
19. Nishikawa RM, Bae KT. Importance of better human-computer interaction in the era of deep learning: mammography computer-aided diagnosis as a use case. J Am Coll Radiol. 2018;15:49–52.
20. Nishikawa RM, Schmidt RA, Linver MN, Edwards AV, Papaioannou J, Stull MA. Clinically missed cancer: how effectively can radiologists use computer-aided detection? Am J Roentgenol. 2012;198:708–16.
21. Litjens G, Koo T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.
22. Dechter R. Learning while searching in constraint-satisfaction problems. In: AAAI'86: proceedings of the fifth AAAI national conference on artificial intelligence. Philadelphia: AAAI Press; 1986. p. 178–83.
23. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.
24. Vaillant R, Monrocq C, LeCun Y. Original approach for the localisation of objects in images. IEE Proc Vis Image Signal Process. 1994;4:245–50.
25. Sahiner B, Chan HP, Petrick N, Wei D, Helvie MA, Adler DD, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. IEEE Trans Med Imaging. 1996;5:598–610.
26. Kooi T, Litjens G, Ginneken B, Gubern-Mérida A, Sánchez C, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal. 2017;35:303–12.
27. Moreira I, Amaral I, Domingues I, Cardoso A, Cardoso M, Cardoso J. INbreast: toward a full-field digital mammographic database. Acad Radiol. 2012;19:236–48.
28. Heath M, Bowyer K, Kopans D, Kegelmeyer P Jr, Moore R, Chang K, et al. The digital database for screening mammography. Current status of the digital database for screening mammography. In: Karssemeijer N, Thijssen M, Hendriks J, van Erning L, editors. Digital mammography. Computational imaging and vision, 13. Dordrecht: Springer; 1998.
29. Halling-Brown M, Looney P, Patel M, Warren L, Mackenzie A, Young K. The oncology image database (OMI-DB). In: Law MY, CookProc Tessa S, editors. SPIE 9039, medical imaging 2014: PACS and imaging informatics: next generation and innovations, 903906. Bellingham: Society of Photo-Optical Instrumentation Engineers (SPIE); 2014.
30. Agarwal R, Diaz O, Lladó X, Martí R. Mass detection in mammograms using pre-trained deep learning models. J Med Imaging. 2019;6:031409.
31. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Ghahramani Z, Welling M, Cortés C, Lawrence ND, Weinberger KQ, editors. Proceedings of the 27th international conference on neural information processing systems. Cambridge (USA): MIT Press; 2014. p. 2672–80.
32. Korkinof D, Rijken T, O'Neill M, Yearsley J, Harvey H, Glocker B. High-resolution mammogram synthesis using progressive generative adversarial networks; 2019 arXiv:1807.03401
33. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. J Natl Cancer Inst. 2019;111:916–22.
34. Liu KL, Wu T, Chen PT, Tsai YM, Roth H, Wu MS, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. Lancet Digit Health. 2020;2:e303–13.
35. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical image computing and computer-assisted intervention – MICCAI 2015. Lecture notes in computer science. Munich: Springer; 2015. p. 234–41.
36. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. IEEE TMI: Trans Med Imaging. 2019;38: 1788–800.
37. Trister AD, Buist DS, Lee CI. Will machine learning tip the balance in breast cancer screening? JAMA Oncol. 2017;3:1463–4.
38. Cole EB, Zhang Z, Marques HS, Nishikawa RM, Hendrick RE, Yaffe MJ, et al. Assessing the stand-alone sensitivity of computer-aided detection with cancer cases from the digital mammographic imaging screening trial. Am J Roentgenol. 2012;199:W392–401.
39. Kim E-K, Kim H-E, Han K, Kang BJ, Sohn Y-M, Woo OH, et al. Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. Sci Rep. 2018;8:2762.
40. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Invest Radiol. 2017;52:434–40.
41. Geras KJ, Wolfson S, Shen Y, Kim S, Moy L, Cho K. High-resolution breast cancer screening with multi-view deep convolutional neural networks; 2017 arXiv:170307047
42. Rodríguez-Ruiz A, Krupinski E, Mordang J-J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology. 2019;290:305–14.
43. Salim M, Dembrower K, Eklund M, Lindholm P, Strand F. Range of radiologist performance in a population-based screening cohort of 1 million digital mammography examinations. Radiology. 2020;297:33–9.
44. Kooi T, Karssemeijer N. Classifying symmetrical differences and temporal change for the detection of malignant masses in mammography using deep neural networks. J Med Imaging. 2017;4:044501.
45. Mordang J-J, Janssen T, Bria A, Kooi T, Gubern-Mérida A, Karssemeijer N. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In: Tingberg A, Lång K, Timberg P, editors. Proc. of the international workshop on digital mammography. Berlin: Springer-Verlag; 2016. p. 35–42.

46. Weber RJ, van Bommel RM, Louwman MW, Nederend J, Voogd AC, Jansen FH, et al. Characteristics and prognosis of interval cancers after biennial screen-film or full-field digital screening mammography. Breast Cancer Res Treat. 2016;158:471–83.
47. iCAD. Artificial intelligence for digital breast tomosynthesis – reader study results (white paper); 2018. Available from: https://www.icadmed.com/assets/dmm253-reader-studies-results-rev-a.pdf [accessed 04.08.20].
48. Tchou PM, Haygood TM, Atkinson EN, Stephens TW, Davis PL, Arribas EM, et al. Interpretation time of computer-aided detection at screening mammography. Radiology. 2010;257:40–6.
49. Bahl M. Detecting breast cancers with mammography: will AI succeed where traditional CAD failed? Radiology. 2018;290:315–6.
50. Kyono T, Gilbert FJ, van der Schaar M. MAMMO: a deep learning solution for facilitating radiologist-machine collaboration in breast cancer diagnosis; 2018 arXiv:181102661
51. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Teuwen J, Broeders M, Gennaro G, 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;29:4825–32.
52. Chae EY, Kim HH, Jeong J-W, Chae S-H, Lee S, Choi Y-W. Decrease in interpretation time for both novice and experienced readers using a concurrent computer-aided detection system for digital breast tomosynthesis. Eur Radiol. 2019;5:2518–25.
53. Balleyguier C, Arfi-Rouche J, Levy L, Toubiana PR, Cohen-Scali F, Toledano AY, et al. Improving digital breast tomosynthesis reading time: a pilot multi-reader, multi-case study using concurrent Computer-Aided Detection (CAD). Eur J Radiol. 2017;97:83–9.
54. Benedikt RA, Boatsman JE, Swann CA, Kirkpatrick AD, Toledano AY. Concurrent computer-aided detection improves reading time of digital breast tomosynthesis and maintains interpretation performance in a multireader multicase study. Am J Roentgenol. 2018;3:685–94.
55. Brennan ME, Turner RM, Ciatto S, Marinovich ML, French JR, Macaskill P, et al. Ductal carcinoma in situ at core-needle biopsy: meta-analysis of underestimation and predictors of invasive breast cancer. Radiology. 2011;260:119–28.
56. Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, et al. Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features. J Am Coll Radiol. 2018;15:527–34.
57. Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, et al. Mammographic breast density assessment using deep learning: clinical implementation. Radiology. 2019;290:52–8.
58. Kallenberg M, Petersen K, Nielsen M, Ng AY, Pengfei D, Igel C, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans Med Imaging. 2016;35:1322–31.
59. Wanders JO, van Gils CH, Karssemeijer N, Holland K, Kallenberg M, Peeters PH, et al. The combined effect of mammographic texture and density on breast cancer risk: a cohort study. Breast Cancer Res. 2018;20:36.
60. Wang J, Ding H, Bidgoli FA, Zhou B, Iribarren C, Molloi S, et al. Detecting cardiovascular disease from mammograms with deep learning. IEEE Trans Med Imaging. 2017;36:1172–81.