Mammography in 2022, from Computer-Aided Detection to Artificial Intelligence Applications

Filippo Pesapane¹,*, Chiara Trentin¹, Marta Montesano¹, Federica Ferrari¹, Luca Nicosia¹, Anna Rotili¹, Silvia Penco¹, Mariagiorgia Farina¹, Irene Marinucci¹, Francesca Abbate¹, Lorenza Meneghetti¹, Anna Bozzi¹, Antuono Latronico¹, Alessandro Liguori², Giuseppe Carrafiello², Enrico Cassano¹

¹Breast Imaging Division, IEO European Institute of Oncology IRCCS, 20141 Milan, Italy
²Department of Radiology and Department of Health Sciences, Fondazione IRCCS Ca Granda Ospedale Maggiore Policlinico and University of Milano, 20122 Milan, Italy
*Correspondence: filippo.pesapane@ieo.it (Filippo Pesapane)

Abstract

Objective: During the last decades, advances in computing power, structured data, and algorithm development, developed a technology based on Artificial intelligence (AI) which is currently applied in medicine. Nowadays, the main use of AI in breast imaging is in decision support in mammography, where it facilitates human decision-making as opposed to replacing radiologists. In this paper, we analyze how AI is currently involved in radiological decision-making and how will change both interpretation efficacy and workflow efficiency in breast imaging. Mechanism: We performed a non-systematic review on Pubmed and Scopus and Web of Science electronic databases from January 2001 to January 2022, using the following keywords: artificial intelligence, machine and deep learning, breast imaging and mammography. Findings in Brief: Many retrospective studies showed that AI can match or even enhance performances of radiologists in mammography interpretation. However, to assess the real role of AI in clinical practice compelling evidence from accurate perspective studies in large cohorts is needed. Breast imaging must face with the exponential growth in imaging requests (and consequently higher costs) and a predicted reduced number of trained radiologists to read imaging and provide reports. To mitigate these urges, solution is being sought with increasing investments in the application of AI to improve the radiology workflow efficiency as well as patient outcomes. Conclusions: This paper show the background on the evolution and the application of AI in breast imaging in 2022, in addition to exploring advantages and limitations of this innovative technology, as well as ethical and legal issues that have been identified so far.

Keywords: breast imaging; artificial intelligence; mammography

1. Introduction

Detection and localization of suspicious findings is a main task for breast radiologists when reporting mammography. Suspicious findings include abnormalities like clustered microcalcifications, nodules and mass lesions, and architectural distortions of the breast.

Screening programs have contributed to a reduction in breast cancer–related mortality [1,2] but such workflow is money- and time-consuming and it has some drawbacks. The ability to appropriately find lesions by radiologists may be limited by some features indeed, like the suboptimal image quality of the exams, the extremely dense tissue of the breast, the inaccurate assessment of subtle or complex patterns [2,3]. Moreover, the fatigue or other radiologists’ limitations may influence the detection of abnormalities [4].

Accordingly, computer-aided detection (CAD) systems was used (and developed) for decades to support radiologists in the detection of suspicious findings in mammography [5]. More recently, the inclusion of standard digital imaging among the sources of big data for precision medicine has been one of the innovative frontiers of research [6,7]. Nowadays, big data can be analyzed through Artificial Intelligence (AI), which refers to a technology dedicated to the creation of algorithms and models performing tasks that traditionally require human intelligence [8]. In breast cancer, AI has been recently applied for the detection and characterization of suspicious finding, especially in mammography [6].

The aim of this paper is to review the advancement of AI in mammography, the current state of the art and to analyze the advantages and limitations of this new changing technology.

Primary publications concerning AI and breast imaging were identified using the PubMed, Scopus, and Web of Science electronic databases in the last 20 years (from January 2001 to January 2022). The search strategy was developed without any language or other restrictions, and used a combination of the free text words and...

2. Why do We need AI in Mammography?

Breast cancer is the most common non–skin cancer and a leading cause of cancer death in North American and European women [9]. Although mammography is currently the only screening test which has shown to reduce breast cancer–related mortality [2], it has limitations that AI could overcome [3]. These limitations include significant variation in human interpretation, a high rate of both false-positive and false-negative results, a restricted global access to mammography programs due to shortages of specialized radiologists capable of interpreting such exams, and overall costs leading to inequalities between low- and high-income countries [10,11].

AI-based algorithms may save radiologist’s time scrutinizing mammography screenings by detecting and characterizing abnormalities on mammograms allowing radiologists to read faster cases labelled as abnormalities-free and, on the contrary, pay more attention to the exams with abnormalities, while making screening cheaper and more accessible to population [12].

Moreover, while performance of radiologists decreases after 70–80 minutes of reading [4], AI has consistent performance, never getting tired. Mammography screening supported with AI can help to reduce the radiologists’ overload of work, decreasing the increasing burnout rate of physicians that was reported in the last few years, probably due also by COVID-19 pandemics [13–15]. Radiologists consistently work under pressure indeed as they must read a large volume of mammograms in a relatively short time to avoid delay in diagnosis [8]. Accordingly, the introduction of AI in mammography may assist radiologists in the triage of negative exams and helping them with the detection of suspicious findings, reducing the stress during their work. On the other hand, AI may generate in radiologists an over-confidence and a dependence of AI and, at the same time, an under-confidence in themselves and their abilities as physicians [16].

When performance of AI in mammography will be assessed in real practice medicine (most of the studies published in literature are retrospective and they may not represent the clinical practice in real-world) [3], this technology may be the solution for accessing reliable breast cancer screening in low-/middle-income countries where cancer screening is limited due to equipment cost and the expert skill required for interpretation of mammography, and it may help reduce existing health inequalities [11,13]. The introduction of AI in breast cancer screening mammogram interpretation could be essential even in high-income countries to face the current (and the expected) shortage of radiologists [17].

Accordingly, AI is increasingly considered a possible solution to these issues and several studies are evaluating how and when it will be successfully used in clinical practice.

3. From CAD to AI

Radiologists are already familiar with CAD systems which were first introduced in the 1960s in mammography [18]. In 1998 the US Food and Drug Administration approved the use as a second-reader in mammography [19], namely the radiologist first performs his/her own reading of mammography and would only view the CAD system output afterward. Although CAD systems were proclaimed as a technology that significantly improved the performance of mammography, some large-scale prospective studies demonstrated no real benefit of CAD technology in improving breast radiologist performance in mammography reading [16].

During the last decades, advances in algorithm development, combined with the easier access to computational resources, allows AI to be applied in medical imaging at a higher functional level by analyzing large volumes of quantitative information from data (derived by images), supporting radiologists in image interpretation as a concurrent/secondary/autonomous reader at various steps of the radiology workflow [20].

The main feature that distinguishes AI-based image classification algorithms from previous conventional CAD is that the assessment of which images features are suggestive of abnormalities being present is achieved by the algorithm itself during its training, and not input by the (human) programmer [21]. Then, the AI model is not taught what a breast cancer looks like (i.e., shape, size, texture patterns) but it learns what it looks like. This is accomplished during the training process, by providing the algorithm many examples of data/images (portions or complete images) with and without abnormality present, each of them labeled with its actual status (presence of abnormality or not). During the training, for each input example image the algorithm adjusts its internal variables to minimize the difference between its predicted status of the image to its reference. Therefore, the network recognizes what the image features are that point to a cancer being present.

As opposed to many other pathologies, the classification of suspicious finding in mammography follows a well-accepted standard [21]. Accordingly, screening mammography is a precious domain for AI because it has essential characteristics that make it suitable for AI [3] as the screening test itself has a binary outcome: the patient undergoing screening mammography is either cleared (if no suspicious findings are detected) or recalled for additional examinations. Also, the diagnosis is binary: namely, the patient
is classified as disease negative or positive, facilitating the development and the application of an AI algorithm.

AI can extend human skills in ways that CAD cannot indeed, and its strongest potential role could be in new applications beyond assisting the radiologist to detect early-phase breast tumours [8].

However, the development of AI has not yet been followed by its integration in routine radiological practice. Radiology can learn from the previous experiences of CAD applied to mammography and leverage that knowledge to translate our discoveries more quickly in AI to improved patient care.

4. The State of the Art of AI in Mammography

Usually, studies about performance of AI applications in radiology evaluate sensitivity, specificity, area under the curve (AUC) and computation time (namely, the time taken to the process to provide an outcome). In mammography such AI systems achieved a sensitivity from 0.56 to 0.82 with a specificity of 0.84–0.97 [22,23], showing a cancer detection accuracy comparable to a radiologist specialized in breast imaging [24].

Researchers from Imperial College London and Google Health showed that DeepMind’s medical AI system may outperform radiologists on identifying breast cancer from mammography [22], paving the way for clinical trials to improve the accuracy and efficiency of breast cancer screening by AI.

In a simulation study, a deep learning model performed triage of screening mammograms, demonstrating an improvement in radiologist efficiency and specificity, without decreasing of sensitivity [25].

In 2019, Rodríguez-Ruiz et al. [24] retrospectively compared the performance of an AI as a stand-alone model to the radiologists’ performance in the detection of breast tumors, yielding a total of 2652 mammograms (653 malignant) and interpretations by 101 radiologists (28296 independent interpretations), showing a non-inferior accuracy to breast radiologists. Particularly, when analysing the performance of the AI model and to the radiologists’ performances, authors reported that the former performed better than 61% of the radiologists while the performances of the AI system was similar to that of the radiologists.

In another study analyzing the use of a AI software by 24 radiologists who retrospectively analyzed an 260 mammography exams, Conant et al. [12] reported an average specificity and sensitivity of the readers were 62.7% and 77.0%, respectively, for human radiologist and 41% and 91%, respectively, for the AI system.

In an international retrospective study of an AI algorithm for breast cancer screening by McKinney et al. [22], the AI system performed non-inferiorly when compared to the performance of radiologists, with better sensitivity (+9.40%), specificity (+5.70%) and AUC (0.740 for AI and 0.625 for radiologists). Specifically, The AI system had a 0.840 (95% CI (Confidence interval): 0.820–0.860) area under the ROC curve (AUC) and the average AUC of the radiologists was 0.814 (95% CI: 0.787–0.841), resulting in a difference 95% CI of ~0.003–0.055.

Finally, in a retrospective multi-readers study, Kim et al. [26] analyzed an AI software using 3 separate international datasets comparing the performance of the AI algorithm in the detection of breast cancer in 320 mammography exams to that of 14 radiologists, reporting that AI was more accurate than the average radiologist (AUC = 0.810) and the all radiologists (AUC = 0.940).

Although in literature there are a lot of other recent studies about AI in mammography [21], there are only few prospective studies on the use of AI in radiology [27], with a recent systematic review only reporting one randomized trial registration and two prospective studies (non-randomized) in radiology [28]. Prospective studies are essential to fully understand the influence of AI on human performance and the interaction between radiologists and computer.

Finally, we need to consider that as amount of collected data and AI applications in healthcare (not only in breast imaging scenario) can only grow in the next years, actions regarding regulation of data and cybersecurity will face continuous challenges. Before using government over-regulation, we need to face the cybersecurity implications technologically because data protection can no longer rely on current technologies that allow spreading of personal data at a large and uncontrolled scale [29]. A possible solution could come from blockchain technology (BCT), namely an open-source software, that allows the creation of a large, decentralized, and secure public databases, containing ordered records arranged in a block structure [30].

5. Advantages

If used as a second reader, AI may provide a more confident diagnosis, supporting radiologists in making their clinical evaluation and radiological assessments.

Particularly, for unexperienced radiologists the interpretation of mammography might be challenging, especially in breast images of dense breasts as parenchymal and breast density are strong variables in breast cancer risk estimation [31]. Radiologists assess breast density by the four-category Breast Imaging Reporting and Data System (BI-RADS) density ratings [32]. In some breast cancer centers, automated assessment of breast density on mammograms can be currently performed by commercial AI software [33].

This may help radiologists to increase their productivity. Breast radiologists spend most of their time scrutinizing mammography screenings indeed (it is estimated that around 5 screening mammograms out of 1000 shows breast cancer) [12]. By detecting and characterizing suspicious findings on a mammogram, or indicating their ab-
sence, AI may triage a negative mammography (47–60%), which would imply that it would not need to be re-assessed by two (or even one) radiologist [34].

Moreover, AI could help breast radiologists to focus on significant findings only, lowering the rate of false-negative and false-positive results [34]. Particularly, AI may decrease false-positive findings compared with the number associated with currently available CAD systems. Accordingly, if a certain (even small) percentage of cases can be identified as negative with high accuracy, the same amount of radiologists’ time may be used at more complicated cases and other activities requiring natural/human intelligence. The false-positive mammography examinations are estimated to cost the US health care system approximately $4 billion each year, then a decrease in false-positive CAD flags has strong potential to lower the overall cost of breast screening programs [35].

Finally, we already mentioned that radiologists consistently operate under pressure and fatigue. The same holds true for the perceived risk of malpractice lawsuits. On the contrary, AI has consistent performance (and, obviously, never gets tired).

6. Limitations

Although AI is gradually being accepted as a useful tool in radiological and clinical practice, the attention is gradually moving from excitement and fear for jobs to worries about the effect on patient care. Particularly, there is an increasing awareness about the necessity to obtain more evidence about the usefulness and value of AI applications, requiring thorough and continuous evaluation and monitoring of outcomes, even on a multicenter level.

Although the above-mentioned studies announced exceptional improvement in AI performance over the results of radiologists, there are based on in-silico data or retrospective studies that may not be representative of real-world radiological practice.

In a recent systematic review by Freeman et al. [36] assessed the accuracy of AI algorithms, alone or in combination with radiologists, to detect cancer in digital mammograms: they reported that 34 of 36 AI systems evaluated were less accurate than a single radiologist, and all were less accurate than the consensus of two or more radiologists. Moreover, authors reported inadequate methodological quality because design of the study was retrospective and showed high risk of bias due to the tumor-enriched populations, reader study laboratory effect, differential verification of outcomes, and insufficient follow-up. Such biases may have resulted in a sensitivity over-estimation and a specificity underestimation, as it already happened previously in CAD studies, showing an interesting overlap in the methods of evaluations of those technologies [16,19,37,38].

Although mammography screening with AI could be cost-effective as they improve the breast cancer detection in early phase, the great investments in AI development require careful monitoring regarding the cost-effectiveness evaluation [27]. Such evaluations demand continuous real-time studies which considers the effect on nationally and internationally reported standards (e.g., detection rate of cancers, recall rates, tumor size and lymph node status). However, the warnings of testing such as how to access datasets and a consensus about the clinically relevant thresholds still need to be reached. The current heterogeneous digital infrastructure is a limit indeed: ingestion of new solutions needs to be more streamlined as most hospital systems are still functioning separately and the infrastructure of most hospitals currently is not ready for seamless integration of AI-based solutions [39].

Another limit is the urge of huge data storage for the curation of datasets and the storage of additional image analysis created by AI systems. This comes with other issues: when using special data like personal health information of patients. AI algorithms need to comply with and legal rules and regulatory frameworks [8,40]. Therefore, health data would need to be anonymized or—at least—depersonalized with an informed consent processes which include the possibility of worldwide distribution [5].

The ownership of data, particularly the personal data concerning health information, is another part of the discussion on the application of different ownership rules to original, de-identified, anonymized, and processed data. However, the current anonymization or de-identification techniques are still substandard and there are no available certifications for tools and methods for anonymization, as far as we know [41].

Only the collaboration between physicians, data scientists, healthcare operators and providers, patients, and policy makers will be able to prevent the risks of inappropriate use of sensitive datasets, inaccurate disclosures, and limitations in deidentification techniques [8]. At the same time, healthcare providers need to develop and train a new multidisciplinary team of computer scientists and data scientists who will collaborate with clinicians to incorporate AI analysis into clinical decisions. Additionally, university needs to update the training of the new generation of radiologists including AI into their curriculum (currently there is no standardized curriculum for AI education nor are their relevant accreditation requirements within most medical training programs) [20]. Accordingly, more funding and grants for support of AI-based research will be needed, and radiological societies will continue to adapt in terms of both education and research, as well as provision of new training opportunities for radiologists and others.

The financing issues also include the lack of optimal business models to invest in AI solutions. A balance needs to be found between investing in quality/value vs. productivity. In a recent paper, Chen et al. [42] explain that as payment systems in healthcare progressively evolve towards more mature value-based payment models where measuring improvement in quality becomes in-
creasingly important at decreased costs, AI is also likely to become a valuable and indispensable tool for radiologists and healthcare systems. In many European countries, the reimbursement system in healthcare it tends to make it more economically advantageous to provide a greater number of medical exams/procedures—whatever the quality of the same—rather than delivering fewer exams/procedures of guaranteed high quality [41–45]. Accordingly, a shift towards value-based healthcare systems and quality-based reimbursement, might increase the perceived importance of quality and thus the value of using AI solutions.

Finally, AI is currently deeply influenced by interpretability, that is a crucial issue for scientists who try to understand how some models come to conclusions, and, therefore, how to interpret potential failures. Such challenge has a practical drawback when reporting the outcome to the radiologists and to the clinicians, who may not understand all the processes behind the AI proposed clinical response [7].

7. Ethical and Legal Constraints

When introducing AI into pre-existing clinical workflows, it is important to not underestimate the potential ethical and legal issues that appear.

In some AI models, particularly if built on unbalanced datasets underrepresenting some populations (with a lack of diversity in ethnicity, age, and socioeconomic conditions) there is an intrinsic latent and dangerous bias that may made generalization impossible to provide [27].

The introduction of AI in mammography is not only a medicine matter, but it comes with also juridical issues as they alter the standard criteria that regulate the evaluation of medical liability. New legal implications will impact not only healthcare providers, but also the industries and companies involved in clinical tools based on AI, the governments and the regulators of such AI technology, and the patients whose treatment plan might be supplemented by an “opinion” expressed by AI [40]. If the AI outcome aligns to the opinion of the human clinician, there is no particular issue, as the AI device merely provide a confirm of a previously assessed opinion. However, the medical professional might feel comforted in an inaccurate opinion/decision and be therefore less inclined to explore any doubt by consulting colleagues. On the contrary, if the opinion expressed by the AI device differs from the opinion of the clinician, the scenario becomes more complicated. For instance, the human professional may trust the AI device over his own judgment. This is a delicate and difficult scenario and any clinicians should be left to face such choices alone; it is essential that both healthcare providers and associations of clinicians take an active step towards their employees and members, respectively, by proposing the appropriate instruments including protocols, guidelines, and training programs, that can support clinicians to truly understand of the AI algorithms they are using [41].

Moreover, when AI will start making autonomous decisions concerning diagnosis of diseases or management of patients, stopping to be only a support tool, new issues will arise as whether its developer can be held accountable for the decision. Preliminary studies showed that errors in AI mainly happen when confounding variables are correlated with relative pathologic entities in the training datasets rather than actual symptoms [27]. When AI devices decide, such decision is based on the combination of the collected data and the algorithms they are based on and what they learnt.

8. Patients’ Perceptions

Patient acceptability about introduction of AI in mammography and its effect on uptake of screening should be included among measures for clinical evaluation. International collaborations resulted in sort of guidelines for the development of AI technology, highlighting the demand for patient involvement to guide the implementation of patient-center AI, which is a crucial point to gain the trust of the patient population [27].

Sechopoulos et al. [46] wondered if it would be ethical to automate the interpretation of mammography. Is it acceptable, or will the screened population accept, that some of their images will not be reviewed by any human? In a recent survey study of 922 Dutch women [10], 77.8% of respondents opposed standalone AI interpretations. However, using AI triaging for a second read had more supporters: 31.5% agreed with this method while 41.7% disagreed. Overall, authors showed that respondents did not agree with a total independent use of autonomous systems (namely, without the involvement of radiologists). On the other hand, the combination of a radiologist as a first reader and an AI system as a second reader in a breast cancer screening program found most support.

Mc Bride et al. [47] reported that both clinicians and patients consistently agree with a scenario where AI healthcare innovations are fully integrated within healthcare systems and support the work of clinicians instead of substitute them.

Overall, the most acceptable approach for the interviewed women is currently the combination of a radiologist as a first reader and an AI system as a second reader. Specifically, women need to be fully informed about the use of AI in healthcare, and they want to retain human interaction in the diagnostic process [17].

On the other hand, Jonmarker et al. [48] reported that their respondents with higher education were more likely to prefer a standalone AI interpretation. Moreover, in a survey among physicians/general practitioners, the majority of them (76%) would accept the use of AI as a tool for triage, letting it to filter out likely negatives examinations (i.e., mammography) without radiologist confirmation, accepting the use of AI to make decisions about likely negative exams without radiologist second evaluation [43].
Studies in this area are still relatively limited, particularly regarding patients’ perspectives. There is a pressing need for the development of comprehensive, large-scale studies to understand patients’ needs, expectations, and concerns when it comes to AI applications.

However, important questions remain about the rate of missed findings that would be acceptable for AI when used in routine screening and if this can be accepted by women undergoing breast cancer screening who need to be informed of both the possible benefits and the possible risks.

9. Conclusions

There are many steps to be taken before AI will become a worldwide application in breast imaging workflow.

Future research is needed, and better designed studies have to investigate the clinical application intended for the AI models. Particularly, AI should help radiologists to read mammography and detect suspicious finding with a higher degree of accuracy. To obtain valuable results and to really introduce this technology in the radiology workflow, studies should not focus on the replacement of humans/radiologists by machines, but on the fusion of machines and human vision: this is the combination that needs to be emphasized. Moreover, the different ethical and legal issues about introduction of AI in clinical practice must be discussed among regulators, companies, clinicians, and patients, to provide updated guidelines for healthcare professionals to follow [40].

Finally, the legal accountability should be clearly stated for companies and healthcare professionals when using AI systems.

Author Contributions

(1) Conception and design—FP, ALa, ALi, GC, AB; (2) Administrative support—EC, FA, LM; (3) Provision of study materials—CT, MM, FF, LN, AR, SP, MF; (4) Collection and assembly of data—IM, FA, EC, CT; (5) Data analysis and interpretation—CT, LM, LN, AR; (6) Manuscript writing—FP, MM, AR, SP, FA, AB; (7) Final approval of manuscript—All authors.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest. EC and LN had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to MHD.

References

[17] de Vries CF, Morrissey BE, Duggan D, Staff RT, Lip G. Screening participants’ attitudes to the introduction of artificial intelli-


