Oncology

Oncology and Informatics – Review

Oncology 2021;99:483–490
DOI: 10.1159/000515698

Received: January 9, 2021
Accepted: March 5, 2021
Published online: May 21, 2021

A Review of Applications of Machine Learning in Mammography and Future Challenges

Sai Batchu[a] Fan Liu[b] Ahmad Amireh[c] Joseph Waller[d] Muhammad Umair[e]

[a]Cooper Medical School of Rowan University, Camden, NJ, USA; [b]Stanford University School of Medicine, Stanford, CA, USA; [c]Duke University Medical Center, Durham, NC, USA; [d]Drexel University College of Medicine, Philadelphia, PA, USA; [e]Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Keywords
Machine learning · Mammography · Artificial intelligence

Abstract
Background: The aim of this study is to systematically review the literature to summarize the evidence surrounding the clinical utility of artificial intelligence (AI) in the field of mammography. Databases from PubMed, IEEE Xplore, and Scopus were searched for relevant literature. Studies evaluating AI models in the context of prediction and diagnosis of breast malignancies that also reported conventional performance metrics were deemed suitable for inclusion. From 90 unique citations, 21 studies were considered suitable for our examination. Data was not pooled due to heterogeneity in study evaluation methods. ***Summary:*** Three studies showed the applicability of AI in reducing workload. Six studies demonstrated that AI can aid in diagnosis, with up to 69% reduction in false positives and an increase in sensitivity ranging from 84 to 91%. Five studies show how AI models can independently mark and classify suspicious findings on conventional scans, with abilities comparable with radiologists. Seven studies examined AI predictive potential for breast cancer and risk score calculation. ***Key Messages:*** Despite limitations in the current evidence base and technical obstacles, this review suggests AI has marked potential for extensive use in mammography. Additional works, including large-scale prospective studies, are warranted to elucidate the clinical utility of AI.

Introduction

Mammography remains a critical tool for screening and diagnosing breast cancers. Advocates for mammography screening refer to its widely documented contribution in reducing breast cancer mortality rates [1–4]. While mammographic screening has established reduction in mortality, like any examination, there is a false-positive rate associated with screening mammography. While only 7–12% of women are falsely recalled after only one mammogram, over 50% of women who have undergone annual mammography screening for 10 years will be recalled incorrectly [5, 6]. These false positives translate into increased benign biopsies, increased spending, and negative psychological effects for patients involved [7–9]. Likewise, potentially malignant neoplasms are at risk of

karger@karger.com
www.karger.com/ocl

Correspondence to:
Sai Batchu, batchus1@tcnj.edu

Karger

Table 1. Literature sources and search terms

Literature sources	Search terms
PubMed	"Artificial Intelligence [Mesh]" AND "mammography"
IEEE Xplore	"Mammography" AND "Artificial Intelligence", filter applied: Journals only
Scopus	TITLE-ABS-KEY (mammography AND "artificial intelligence") AND (LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015)) AND (LIMIT-TO (DOCTYPE, "ar"))

being missed due to their small size or surrounding dense fibroglandular tissue [10]. False-negative mammogram results have a higher incidence in women aged 50–89 with previous benign biopsies. However, the rate of false-negative results is still relatively low, reported in 1.0 to 1.5 per 1,000 women [11]. Although its accuracy continues to improve with technical improvements, diagnostic mammography is the gold standard for evaluation of breast cancer [12]. In order to further increase the accuracy and to further reduce the rates of false positives and false negatives, recent advances in artificial intelligence (AI) have been exploited to develop software capable of aiding radiologists in clinical practice.

AI is a field incorporating computer science, engineering, and statistics to produce intelligent computer programs capable of performing tasks that would otherwise require human intelligence [13]. AI has been widely implemented in healthcare, ranging from one of its earliest uses in blood disease diagnosis to current implementations in genomic medicine to characterize mutations [14, 15]. Although most AI in medicine is limited to research, certain commercial algorithms have already been approved for use by radiologists in clinical practice to help decrease erroneous readings [16–18]. Currently, many of these AI-based tools designed for aiding radiologists and interpreting mammograms are developed with machine learning. Machine learning is a specific domain of AI and is concerned with constructing algorithms used by computers to perform certain tasks without using explicit instructions, but instead relying on inference and patterns and are able to improve their performance with experience [19].

In regards to mammography, the program usually trains on clean labelled mammograms with abnormal diagnosis and normal anatomy serving as labels. As the program trains with greater and more diverse images, it adjusts the algorithms used with the ultimate goal of becoming more accurate in reading images and predicting patterns. As recent advancements allow for the increased scalability of AI software into clinical practice, it is imperative to examine the current environment surrounding AI in mammography.

In this literature review, we critically evaluate published literature describing machine-learning algorithms currently used in mammography and their efficacy in reducing erroneous readings. We conclude with future speculations of the field with a focus on the effects of AI.

Methods and Materials

The primary aim was to evaluate existing literature that used AI in mammography. Several databases were searched including PubMed, IEEE Xplore, and Scopus, with the search terms described in Table 1. Studies were limited to the interval January 2015 to March 2020. Exclusion criteria included (1) lack of conventional performance metrics such as sensitivity, specificity, area under the receiver operating characteristic curve (AUROC), etc., (2) the study investigated the accuracy of image segmentation rather than disease classification, (3) the study only published guidelines, review articles, and abstracts, (4) animal studies, (5) non-English sources, (6) sample size smaller than 10. After removing duplicate studies and those with redundant or non-novel information, 60 unique studies were ultimately included in this review.

Results

Reducing Diagnostic Workload

One of the main potentials of AI in mammography is to reduce workload by helping expedite the interpretation of more obvious cases and thereby allowing radiologists to focus on more challenging cases. Most of the models that classify images use deep learning (DL). DL is a subset of machine learning where algorithms are created and operate similarly to those in machine learning, but there are several "layers" of these algorithms – each providing a unique interpretation of the image feature (e.g., shape, morphology, texture) it analyzes [19]. DL involves training a complex neural network, in which data from the input layer feeds into the first hidden layer, which processes the data further and is sent to the next layer of algorithms (each layer consists of nodes and is loosely modelled from

neurons), until the output layer is reached. Before the DL system is implemented, however, it is important to examine whether the reduced caseload will cause any detrimental effects to radiologists' abilities to classify.

A study by Rodríguez-Ruiz et al. [17] used a commercially available DL system (Transpara 1.4.0, Screenpoint Medical BV) to pre-select digital mammogram images based on the machine learning interpreted likelihood of breast cancer. The system assigned a numerical likelihood score to each image from 1 to 10 (10 implying high likelihood cancer was present in the exam). Pre-selection scenarios included exams-to-be-evaluated as those with scores greater than a certain likelihood score. For example, one scenario would only include images scored 1 or higher to be evaluated by radiologists. This was done for all likelihood scores 1–9. The radiologists' average AUROC, a metric used to evaluate ability to classify, was compared before versus after pre-selection for each scenario. An AUROC of 1 represents perfect classification ability, while 0.5 represents incorrect random classification. Although the absolute area under the curve was not reported, the average AUROC of the radiologists was non-inferior at a margin of 0.05 for 8 out of the 9 scenarios tested ($p < 0.05$) [17]. Thus, pre-selecting cases did not affect the ability of radiologists to classify in a majority of the scenarios tested.

Another recent study by Yala et al. [20] developed an in-house DL model to triage out true-negative cases. After comparing the performance of the radiologists during their original screening assessment to a retrospective scenario where the radiologists did not read any of the DL-triaged cases, the model showed a workload reduction of 19.3%. The study did not specify how accurate the machine learning algorithm was at identifying true negatives. The radiologists in the study showed a significant improvement in specificity (93.5–94.2%; $p = 0.002$) and a non-inferior sensitivity (90.6 vs. 90.1%; $p < 0.001$) at a margin of 0.05 [20].

Another method by which these models can reduce workload is through consensus. Some clinical settings utilize a double-reading process where two radiologists examine the same image. The study by McKinney et al. [21] replaced the second reader with their AI model. The conclusion of the first reader was deemed final when the model and first reader agreed. Any disagreement invoked the second reader's opinions as usual. The study concluded that this process can alleviate workload of the second reader up to 88% [21]. These studies illustrated the ability of AI to reduce caseload while not affecting the overall detection performance of radiologists, and may in fact aid their detection performance. Of note, two readers are not a standard of care clinical practice yet in the US due to cost; however, this study points towards future possibility towards reduction in cost for a second reader if aided by machine learning if clinical standard of care moves in that direction in future.

Table 2. Comparison of area under receiver operator curve of radiologists aided versus unaided by AI in breast cancer diagnosis on digital mammograms

Study	Unaided	Aided	Significance
Rodríguez-Ruiz et al. [22], 2019	0.87	0.89	$p < 0.002$
Watanabe et al. [18], 2019	0.76	0.815	$p < 0.01$

Diagnostic Support

In addition to reducing caseload, models have been developed to aid radiologists examining mammograms in real-time. Rodríguez-Ruiz et al. [22] compared breast cancer detection performance of radiologists aided versus unaided by a commercial AI system. The system provided radiologists with interactive decision support and cancer likelihood scores. On average, AUROC increased with AI support compared to unaided (0.89 vs. 0.87, respectively; $p = 0.002$). Sensitivity increased (83–86%; $p = 0.046$) and reading time per case was similar (146 s unaided vs. 149 s aided; $p = 0.15$) [22]. An additional study showed similar results where radiologists who used an AI-based computer-aided detection (AI-CAD) software showed an increase in their classification performance; AUROC increased with AI-CAD support (0.76 to 0.815; $p < 0.01$) (Table 2) [18].

A drawback of currently available CAD systems is the high rate of false-positive markings, which may hinder performance of radiologists. A recent retrospective study compared a commercial AI-CAD (cmAssist, CureMetrix) to a conventional CAD system (ImageChecker, Hologic) with respect to the false-positive findings marked. The findings showed an overall 69% reduction in false positives per image (FPPI) using AI-CAD compared to CAD. Specifically, AI-CAD exhibited 83% reduction in FPPI for calcifications and 56% reduction for masses. There was no significant decrease in sensitivity [23]. Further studies have also created their own in-house models to be able to mark suspicious masses for radiologists to scrutinize more carefully. The mass segmentation model from Hmida et al. [24] exhibits 91.1% sensitivity, while another DL model from Sapate et al. [25] also shows 91% specificity and a sensitivity of 84%.

To improve digital mammogram readings through a different approach, Zeng et al. [26] incorporated radiologists' subjective thresholds from Breast Imaging Reporting and Data System (BI-RADS). Using the BI-RADS descriptors and decision classification categories assigned to mammograms by radiologists, a probabilistic Bayesian network was modelled. A Bayesian network represents the probability distribution of random variables with possible causal relationships [27]. Zeng et al. [26] hypothesized that each BI-RADS assessment category was thought to have a probability associated with it leading the radiologist to classify the mammogram as a positive or negative finding. The probability thresholds used by each radiologist are subjective. Therefore, if a Bayesian network incorporated the unique threshold observed from a radiologist, it may aid to reduce erroneous readings. Indeed, the study found that this method resulted in a 28.9% reduction of false positives [26]. This novel technique exemplifies how AI can gauge the conservativeness of a radiologist and use the data to support its classification decisions.

Independent Detection and Classification

Programs that can interpret and identify abnormalities in mammograms without radiologist intervention have also been evaluated. When comparing commercial AI software (Transpara 1.4.0, Screenpoint Medical BV) to radiologists in the detection of breast cancer in digital mammograms, the performance of the AI system was found to be statistically non-inferior to the average performance of the radiologists at a non-inferiority margin of 0.05 [28]. Agnes et al. [29] recently proposed a DL model that can categorize mammograms into normal, malignant, and benign categories with an overall sensitivity of 96%. Another group evaluated three different DL models for classifying a breast lesion as benign or malignant. The overall accuracies ranged from 88 to 95% [30].

While it is useful to distinguish between negative and positive cases, it is also clinically important to discriminate recalled-benign from malignant cases. Aboutalib et al. [31] developed a DL model that boasts an AUROC of up to 0.96 for correctly identifying recalled but biopsy-benign cases from malignant and negative cases.

DL tools are also being developed for novel breast-imaging techniques. Digital breast tomosynthesis (DBT) is an emerging tool to overcome limitations of conventional full-field digital mammography (FFDM) and allows volumetric reconstruction of the whole breast. DL models have already been tested on DBT datasets, with future plans to incorporate DBT and FFDM images for more accurate diagnosis [32]. Likewise, images from contrast-enhanced digital mammography are being used to train DL models with the goal of extracting more image characteristics for better diagnostic ability [33].

Breast Cancer Prediction

Previous studies have explored breast cancer risk factors related to hormonal and genetic information [34, 35]. Mammographic breast density, which corresponds to the amount of fibroglandular tissue, has received substantial attention as a major risk factor [36]. Indeed, commonly used breast cancer risk assessments, such as the Tyrer-Cuzick model, are based on mammographic density estimations and questionnaires [37]. However, the use of breast density as a substitute to mammograms for risk prediction is limited because density estimates vary across radiologists and compressing the detailed information of digital scans into a single number loses some of the inherent data [38]. Studies have shown the possibility of unique image features extractable by DL, beyond breast density, which can be used to create more accurate breast cancer risk models.

In a retrospective study, Debrower et al. [39] developed a DL model that calculated breast cancer risk scores (DL risk score) from mammograms and compared the prediction ability of these scores to conventional dense area and percentage density values. The AUROC for using the age-adjusted DL risk scores were significantly higher than using the dense area and percentage density values (DL risk score: 0.65, dense area: 0.60, percentage density: 0.57; $p < 0.001$) [39]. A similar study communicated a DL model that showed greater predictive potential (odds ratio = 4.42 [95% CI: 3.4–5.7]) compared to using breast density alone (odds ratio = 1.67 [95% CI: 1.4–1.9]) [40]. Arefan et al. [41] compared two DL models against a basic linear regression model using breast density, showing the models' two different prior normal image types (mediolateral oblique, craniocaudal) to assess risk prediction efficacy. They found that both DL models consistently exhibited superiority in predicting the short-term breast cancer risk than the density-based model [41].

Breast density is also an important risk factor for interval breast cancers, which are cancers detected within 12 months after normal mammographic screening and comprise up to 28% of all breast cancers in the United States [42]. Hinton et al. [43] implemented a model to predict whether a pre-cancer mammogram will result in a screen-detected or interval cancer within 12 months from imaging. When using the BI-RADS density alone,

Table 3. Comparison of area under receiver operator curve of hybrid-deep learning models and traditional models used for breast cancer prediction on digital mammograms

Study	Traditional model	Hybrid-deep learning model	Significance
Akselrod et al. [44], 2019	Gail model: 0.54	0.78	$p < 0.004$
Yala et al. [45], 2019	Tyrer-Cuzick model: 0.62	0.7	$p < 0.05$

the capacity to correctly distinguish interval cancers from screen-detected cancers was moderate (AUROC = 0.65). Yet, when optimized with the mammograms, the model achieved an AUROC of 0.82 (Table 2) [43]. These studies have revealed how AI can use subtle imaging features, not otherwise detectable by humans, to develop better prediction tools than common density-based models.

Hybrid-DL models that incorporate both patient history and radiologic images have also been tested. When used to predict biopsy malignancy, a hybrid-DL algorithm attained 87% sensitivity, 77.3% specificity, and an overall AUROC of 0.91. Even when trained on clinical data alone, the model performed significantly better than the Gail model, another commonly used breast cancer risk tool (AUROC, 0.78 vs. 0.54, respectively; $p < 0.004$) [44]. These findings are similar to those from Yala et al. [45], whose hybrid-DL model was compared to a risk factor-based regression model (RF-LR) that used traditional risk factor information only and the established Tyrer-Cuzick model (TC). The findings showed that the hybrid-DL model better predicted the development of breast cancer within 5 years based from the initial FFDM exam when compared to the RF-LR and TC (AUROC, hybrid-DL: 0.70, RF-LR: 0.67, TC: 0.62, $p < 0.05$) (Table 3) [45].

Discussion

This review derived its findings from retrospective studies and small reader studies. While there have been preliminary discussions for implementing standards for integration of AI into clinical practice, benchmarks are controlled by multiple stakeholders with a current lag in guidelines [46]. Despite these shortcomings, the potential for AI in mammography is encouraging as evidenced by recent research. Independent groups, plus commercial organizations, have demonstrated the ability of machines to construct more efficient workloads for radiologists, all the while assisting in diagnosis and predicting cancer risk. As technological advancement continues, we expect AI will play a critical role to assist radiologists practicing mammography.

One of the largest obstacles facing the progression of AI in clinical medicine is the limitation to development and reliability of complex algorithms with the capability of handling multivariable data that factors into physician decision-making [47]. This also brings in the question of responsibility when evaluating complex medical cases. Furthermore, the usage of AI requires the implementation of an ethics code promoting transparency and reliability. Geis et al. [48] call attention to the fact that various AI models are relatively straightforward to produce and develop, which emphasizes the importance of streamlining existing perspectives on the subject to focus on radiology specifically. Few existing studies satisfactorily evaluate how AI can be utilized in a way that maximizes impact in a responsible way, and further evaluation of the accuracy and reliability of AI is required in order for this technology to be confidently used in clinical application. Despite these limitations, our general expectation is that AI is capable of playing a significant role in mammography.

Future Challenges

AI is a field that does not yield to one approach or solution. There are numerous algorithms that can be deployed, not to mention countless ways of training, testing, and validating the models. This makes choosing the right framework still more difficult, but also for the same reason, allows great flexibility for the integration of AI into mammography. Nevertheless, as AI is poised to enter into the field at an increasing rate, there are certain issues that must be anticipated.

After training a model, validation is required. The validation stage refers to evaluating the model on how well it has been trained and tuning its hyperparameters. The steps of training and validation are extremely important when developing a clinically relevant model with optimum predictive potential [19]. An increasing worry revolves around creating fair AI practices that promote equity in diagnosis and treatment [49]. Accordingly, devel-

opers will need to prove their model's applicability for large and diverse populations, including minority ethnic/racial groups and those with less common risk factors. This will require cooperation with large hospital networks and other organizations with heterogeneous imaging data to be able to properly train and validate the AI programs [50]. Indeed, AI systems that are generalizable across large populations are already being tested [21].

While FFDM is the most common imaging modality currently used, emerging screening technologies such as DBT are rapidly gaining popularity. Future AI programs must be able to handle the transition from 2-dimensional to 3-dimensional data, the latter of which will require greater storage and computing needs. As with FFDM, these models will need to train on large and diverse DBT datasets to remain clinically relevant [51]. Moving forward, AI designers must adapt to any other novel mammographic screening techniques or risk lagging behind the clinical setting.

Given the esoteric nature of both AI algorithms and mammography, it is crucial for AI programmers and radiologists to work together to alleviate knowledge gaps from both areas. These groups must also be transparent with regulatory agencies, such as the United States Food and Drug Administration (FDA), who ultimately approve whether an AI model can be used as a legitimate medical tool. Most studies evaluating AI are retrospective and small-scale reader studies aimed to demonstrate noninferiority compared to mammography experts; it is uncertain if prospective or other types of trials are required to convince stakeholders of a model's clinical utility [52].

There have also been growing concerns over cybersecurity as the amount of data stored on medical databases, including imaging data, increases [53]. In order to determine the true performance of a model, only mammograms linked to clinically confirmed outcomes are used. Thus, developers not only require a diverse array of mammogram registries to use, but also detailed patient history – especially if the model is designed to use both imaging and clinical data. Novel technologies for disseminating medical imaging data are being implemented, such as blockchain [54]. Nonetheless, imaging data organizations may be reluctant to share millions of mammograms and patient history to research groups without complex data use agreements. Even once the AI tool has been approved and implemented, AI algorithms need to be continuously monitored for possible improvements and security risks to mitigate probability of imaging data sabotage [55].

Conclusion

Programs using AI may be useful with increased efficiency, assistance in diagnosis, and risk prediction in mammography. Future work will need to consider appropriate standards when evaluating the suitability of a model to be used regularly in clinical practice.

Statement of Ethics

This article does not contain any studies with human participants performed by the author. Our research complies with the guidelines for human studies and was conducted ethically in accordance with the World Medical Association Declaration of Helsinki (includes also research on identifiable human material and data).

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

No funding was received for this study.

Author Contributions

J.W. formulated the experimental design. S.B., F.L., and A.A. gathered data and interpreted results. M.U. provided guidance. The manuscript was written and edited by all authors. All authors read and approved the final manuscript.

References

1 Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomized trials. Lancet. 2002;359:909–19.

2 Duffy SW, Tabar L, Vitak B, Day NE, Smith RA, Chen HH, et al. The relative contributions of screen-detected in situ and invasive breast carcinomas in reducing mortality from the disease. Eur J Cancer. 2003;39(12):1755–60.

3 Duffy SW, Tabár L, Chen HH, Holmqvist M, Yen MF, Abdsalah S, et al. The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties. Cancer. 2002 Aug 1;95(3):458–69.

4 Tabar L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. Cancer. 2001;91:1724–31.
5 Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med. 2016;164(4):244–55.
6 Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med. 2011;155(8):481–92.
7 Braithwaite D, Zhu W, Hubbard RA, O'Meara ES, Miglioretti DL, Geller B, et al. Breast Cancer Surveillance Consortium. Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? J Natl Cancer Inst. 2013;105:334–41.
8 Ong MS, Mandl KD. National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year. Health Aff (Millwood). 2015;34(4):576–83.
9 Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med. 2016;164(4):256–67.
10 Huynh PT, Jarolimek AM, Daye S. The false-negative mammogram. Radiographics. 1998;18(5):1137–4.
11 Nelson HD, O'Meara ES, Kerlikowske K, Balch S, Miglioretti D. Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data. Ann Intern Med. 2016;164(4):226.
12 Linden OE, Hayward JH, Price ER, Kelil T, Joe BN, Lee AY. Utility of Diagnostic Mammography as the Primary Imaging Modality for Palpable Lumps in Women With Almost Entirely Fatty Breasts. AJR Am J Roentgenol. 2020;214(4):938–44.
13 Charniak E, McDermott D. Introduction to Artificial Intelligence. Reading, MA: Addison-Wesley; 1985.
14 Buchanan BG, Shortliffe EH. Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Reading, MA: Addison-Wesley; 1984.
15 Rowlands CF, Baralle D, Ellingford JM. Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. Cells. 2019;8(12):1513.
16 Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Breast Cancer Surveillance Consortium. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. JAMA Intern Med. 2015;175:1828–37.
17 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(9):4825–32.
18 Watanabe AT, Lim V, Vu HX, Chim R, Weise E, Liu J, et al. Improved Cancer Detection Using Artificial Intelligence: a Retrospective Evaluation of Missed Cancers on Mammography. J Digit Imaging18. 2019;32(4):625–37.
19 Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006.
20 Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A Deep Learning Model to Triage Screening Mammograms: A Simulation Study. Radiology. 2019;293(1):38–46.
21 McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89–94.
22 Rodríguez-Ruiz A, et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. Radiology. 2019 Feb;290(2):305–14.
23 Mayo RC, Kent D, Sen LC, Kapoor M, Leung JWT, Watanabe AT. Reduction of False-Positive Markings on Mammograms: a Retrospective Comparison Study Using an Artificial Intelligence-Based CAD. J Digit Imaging. 2019;32(4):618–24.
24 Hmida M, Hamrouni K, Solaiman B, Boussetta S. Mammographic mass segmentation using fuzzy contours. Comput. Methods Programs Biomed. 2018;164:131–42.
25 Sapate SG, Mahajan A, Talbar SN, Sable N, Desai S, Thakur M. Radiomics based detection and characterization of suspicious lesions on full field digital mammograms. Comput Methods Programs Biomed. 2016;163:1–20.
26 Zeng J, Gimenez F, Burnside ES, Rubin DL, Shachter R. A Probabilistic Model to Support Radiologists' Classification Decisions in Mammography Practice. Med Decis Making. 2019 Apr;39(3):208–16.
27 Cowell RG, Dawid P, Lauritzen SL, Spiegelhalter DJ. Probabilistic Networks and Expert Systems. New York: Springer; 1999.
28 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 Sep 1;111(9):916–22.
29 Agnes SA, Anitha J, Pandian SIA, Peter JD. Classification of Mammogram Images Using Multiscale all Convolutional Neural Network (MA-CNN). J Med Syst. 2020;44(1):30.
30 Al-Antari MA, Al-Masni MA, Kim TS. Deep Learning Computer-Aided Diagnosis for Breast Lesion in Digital Mammogram. Adv Exp Med Biol. 2020;1213:59–72.
31 Aboutalib SS, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S. Deep Learning to Distinguish Recalled but Benign Mammography Images in Breast Cancer Screening. Clin Cancer Res. 2018;24(23):5902–9.
32 Zhang X, Zhang Y, Han EY, Jacobs N, Han Q, Wang X, et al. Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. IEEE Trans Nanobioscience. 2018;17(3):237–42.
33 Gao F, Wu T, Li J, Zheng B, Ruan L, Shang D, et al. SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. Comput Med Imaging Graph. 2018;70:53–62.
34 Claus EB, Risch N, Thompson WD. The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. Breast Cancer Res Treat. 1993;28(2):115–20.
35 Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004;23(7):1111–30.
36 Kerlikowske K, Zhu W, Tosteson AN, Sprague BL, Tice JA, Lehman CD, et al. Identifying women with dense breasts at high risk for interval Cancer: a cohort study. Ann Intern Med. 2015;162(10):673–81.
37 Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004;23(7):1111–30.
38 Sprague BL, Conant EF, Onega T, Garcia MP, Beaber EF, Herschorn SD, et al. Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study. Ann Intern Med. 2016;165(7):457–64.
39 Debrower K, Liu Y, Azizpour H, Eklund M, Smith K, Lindholm P, et al. Comparison of a Deep Learning Risk Score and Standard Mammographic Density Score for Breast Cancer Risk Prediction. Radiology. 2020 Feb;294(2):265–72.
40 Ha R, Chang P, Karcich J, Mutasa S, Pascual Van Sant E, Liu MZ, et al. Convolutional neural network based breast cancer risk stratification using a mammographic dataset. Acad Radiol. 2019;26(4):544–9.
41 Arefan D, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S. Deep learning modeling using normal mammograms for predicting breast cancer risk. Med Phys. 2020;47(1):110–8.
42 Carney PA, Steiner E, Goodrich ME, Dietrich AJ, Kasales CJ, Weiss JE, et al. Discovery of breast cancers within 1 year of a normal screening mammogram: how are they found?. Ann Fam Med. 2006 Nov-Dec;4(6):512–8.
43 Hinton B, Ma L, Mahmoudzadeh AP, Malkov S, Fan B, Greenwood H, et al. Deep learning networks find unique mammographic differences in previous negative mammograms between interval and screen-detected cancers: a case-case study. Cancer Imaging. 2019;19(1):41.

44 Akselrod BA, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. Radiology. 2019;292:331–42.
45 Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology. 2019;292(1):60–6.
46 Allen B, Jr, Seltzer SE, Langlotz CP, Dreyer KP, Summers RM, Petrick N, et al. A road map for translational research on artificial intelligence in medical imaging: from the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop. J Am Coll Radiol. 2019 Sep;16(9 Pt A):1179–89.
47 Mendelson EB. Artificial Intelligence in Breast Imaging: Potentials and Limitations. AJR Am J Roentgenol. 2019;212(2):293–9.
48 Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. Can Assoc Radiol J. 2019;70(4):329–34.
49 Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866.
50 Lee CI, Houssami N, Elmore JG, Buist DSM. Pathways to breast cancer screening artificial intelligence algorithm validation. Breast. 2020 Aug;52:146–9.
51 Geras KJ, Mann RM, Moy L. Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Concepts and Future Perspectives. Radiology. 2019 Nov;293(2):246–59
52 Redberg RF, Dhruva SS. Moving from substantial equivalence to substantial improvement for 510(k) devices. J Am Med Assoc. 2019 Sep 10;322(10):927–8.
53 Pangalos G, Gritzalis D, Khair M, Bozios L. Improving the security of medical database systems. In: Eloff JHP, von Solms SH, editors. Information Security – the Next Decade. IFIP Advances in Information and Communication Technology. Boston, MA: Springer; 1995.
54 Patel V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health Informatics J. 2019 Dec;25(4):1398–411.
55 Becker AS, Jendele L, Skopek O, Berger N, Ghafoor S, Marcon M, et al. Injecting and removing suspicious features in breast imaging with CycleGAN: A pilot study of automated adversarial attacks using neural networks on small images. Eur J Radiol. 2019;120:108649.