Deep Learning-Based Artificial Intelligence for Mammography

Jung Hyun Yoon[1] and Eun-Kyung Kim✉2

Abstract

During the past decade, researchers have investigated the use of computer-aided mammography interpretation. With the application of deep learning technology, artificial intelligence (AI)-based algorithms for mammography have shown promising results in the quantitative assessment of parenchymal density, detection and diagnosis of breast cancer, and prediction of breast cancer risk, enabling more precise patient management. AI-based algorithms may also enhance the efficiency of the interpretation workflow by reducing both the workload and interpretation time. However, more in-depth investigation is required to conclusively prove the effectiveness of AI-based algorithms. This review article discusses how AI algorithms can be applied to mammography interpretation as well as the current challenges in its implementation in real-world practice.

Keywords: Breast cancer, Mammography, Computer-aided diagnosis, Artificial intelligence, Deep learning

MAMMOGRAPHY FOR BREAST CANCER SCREENING

Currently, breast cancer is the most commonly diagnosed cancer in women worldwide as well as the leading cause of their mortality [1]. The annual incidence of breast cancers was an estimated 1.68 million in 2012 and has continuously risen with a 30% increase estimated in 2025 [1]. Screening programs using mammography have been implemented in several countries for the early detection and treatment of breast cancer, with the aim to mitigate mortality and other serious consequences. Mammography screening seems to have an impact on mortality; randomized controlled trials have shown an approximately 20% reduction in breast cancer-related mortality after mammography was included in breast cancer screening [2]. Based on these results, several countries formulated national recommendations or guidelines that included mammography in breast cancer screening. Although mammography screening has proven effective, the potential drawbacks of this technique have also been acknowledged: 1) false-positive recalls leading to additional imaging studies or biopsies, which in turn increase medical expenses and emotional stress for the patient; 2) false-negatives when breast cancers are either not detectable on mammography or if interpretation errors occur, ultimately delaying diagnosis; 3) radiation exposure; and 4) overdiagnosis of cancers that may not be life-threatening such as low risk ductal carcinoma *in situ* [3].

In the last decade, considerable improvements have been made to overcome these pitfalls of mammographic screening by adding readers, increasing screening frequencies, or adding supplementary imaging modalities to conventional mammography. For instance, the European guidelines for quality assurance in breast cancer screening and diagnosis recommend 'double reading', i.e., mammograms are read independently by two radiologists to enhance sensitivity and reduce unnecessary recalls [4]. Other imaging modalities such as digital breast tomosynthesis (DBT), ultrasonography (US), or magnetic resonance imaging (MRI) have been added to conventional four-view mammography to enhance breast cancer screening outcomes. Although intensifying screening practices and using supplementary imaging modalities may improve breast cancer detection, ensuring sufficient resources may be problematic because the burden of mammography interpretation for radiologists will increase in a double reading setting, along with an inevitable increase in medical expenses as more sensitive and advanced equipment is used [3,5]. Among additional imaging modalities, DBT has been associated with issues such as increased radiation exposure, while there is insufficient evidence to show that it can actually reduce mortality [1,3]. Lastly, as population-based breast cancer screening programs have become commonplace, daily demands for breast cancer screening and the subsequent volume of related tests have continued to rise. However, medical resources remain limited. Therefore, it is critical that current screening workflows are optimized and streamlined [6].

INTRODUCTION OF AUTOMATED DECISION SUPPORT FOR MAMMOGRAPHY INTERPRETATION

Advances in technology and computer programming as well as an urgent need for improved efficiency and accuracy of imaging interpretation workflows have piqued interest in computer-automated analyses of medical images. The main objectives of using computer programs to assist image interpretation are: 1) the automated detection of lesions focusing on the localization of suspicious abnormalities in an image, which is known as computer-aided detection (CADe); and/or 2) characterization of abnormalities detected by either the radiologist or the computer, which is known as computer-aided diagnosis (CADx). Based on the CADe/CADx analysis, the interpreting radiologist determines the clinical significance of the detected abnormality and whether it warrants further investigation. Although the term 'CAD' refers to 'diagnosis' using computers, CAD can be used for purposes other than diagnosis depending on the need of the radiologist. Computer programs can also provide quantitative imaging parameters such as breast parenchymal density that traditionally have been subjectively assessed by the human eye. There are certain time points in the overall workflow of mammography interpretation at which computer assistance is considered most beneficial (Fig. 1); these areas have been discussed in depth in this review.

[Fig. 1](#)

Workflow of mammography interpretation with AI-CAD.

Mammography interpretation relies on visual analysis by radiologists, commonly using the American College of Radiology Breast Imaging Reporting and Data System categories. Computer assistance may be beneficial for mammography interpretation in parenchymal density assessment, lesion detection and characterization, and risk prediction. AI = artificial intelligence, CAD = computer-aided detection/diagnosis

MAMMOGRAPHY INTERPRETATION USING CONVENTIONAL CAD

Initial CAD programs, termed as 'conventional CAD' throughout this review, were based on mathematical models that identified patterns associated with breast cancer and displayed areas with these specific patterns as 'marks' on mammograms [7]. Briefly, these marked areas indicated the spots that the radiologist needed to investigate after screening [8]. When used as a 'spell check' type of system, initial studies showed promising results with CAD for the accurate marking of abnormalities that were later proven to be cancerous [9]. After approval by the U.S. Food and Drug Administration in June 1998 [7,10], the Centers for Medicare and Medicaid Sevices increased reimbursement for CAD in mammography interpretation, which subsequently led to increased CAD usage starting from 5% in 2003, then rapidly increasing to 74% in 2008 and 83% in 2012 [11].

With the exponential increase in the usage of conventional CAD, several studies have evaluated the outcomes of implementing CAD in actual clinical practice; the screening settings used in these studies have been summarized in [Table 1](#). Although there were variations among screening environments studied in previous studies, the majority of these were able to show that using conventional CAD either resulted in higher or similar cancer detection rates at the expense of consistently higher recall rates compared to interpretations without CAD. Conventional CAD programs are biased towards high sensitivity as they are constructed to detect potential malignancies. Sensitivity was reported to be 90.0% for overall cancers, 98.2% for microcalcifications, and 88.7% for suspicious masses [12,13], which may explain the high recall rates. In addition, higher sensitivity resulted in a trade-off with decreased specificity (87.2% from 90.2%) and increased biopsy rates (up by 19.7%) after implementation of conventional CAD [14]. None of the screening performance metrics were improved with CAD in digital mammography (DM), using the data from the Breast Cancer Surveillance Consortium

[11] or the Digital Mammography Imaging Screening Trial (DMIST) [15]. Both the high sensitivity and low specificity seen with conventional CAD result in a loss of reliability because the investigation of excessive marks on images may be tiresome for the radiologist. This was apparent from a previous study, wherein radiologists dismissed approximately 97.4% of the marks drawn by conventional CAD (Fig. 2) [8].
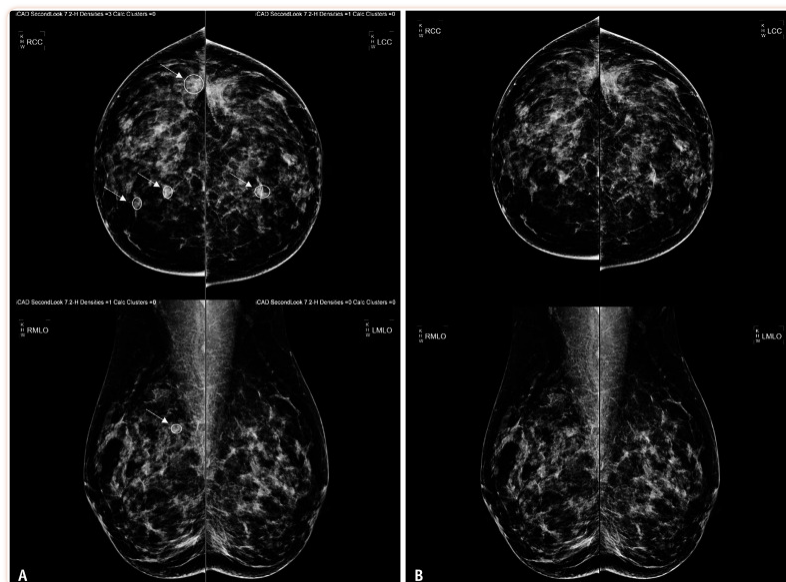

Fig. 2

Representative case of a 46-year-old female who underwent mammography screening.

A, B. Conventional CAD (**A**, SecondLook v7.2, iCAD) shows multiple areas marked in both breasts (arrows at CAD marks) that warrant the attention of the radiologist. In comparison, AI-CAD (**B**, Lunit INSIGHT for Mammography, Lunit Inc.) did not mark any areas on the same image, with low abnormality scores for both breasts. The mammography of this female was interpreted as negative, and this result was confirmed during a follow-up of approximately 46 months during which no suspicious features were detected. AI = artificial intelligence, CAD = computer-aided detection/diagnosis

Table 1

Summary of Published Studies on Diagnostic Performances Using Conventional CAD

	USA				cancers)	
Gilbert et al. 2006 [83]	Screening, 10267 mammograms, 2 centers, 1996, UK	Single reading + CAD vs. double reading	Film mammography	ImageChecker M1000 V5.0; R2 Technology	6.5% more cancers in single reading + CAD than double reading	-
Morton et al. 2006 [84]	Screening, 21349 mammograms, 2001–2002, USA	Pre-CAD vs. CAD interpretations	Film mammography	ImageChecker M1000 V2.2; R2 Technology	7.62% increase in breast cancer detection	-
Gilbert et al. 2008 [85]	Screening, 28204 women in 3 centers, 2006–2007, UK, CADET II trial	Single reading + CAD vs. double reading	Film mammography	ImageChecker DMax V8.1; R2 Technology	7.02 in single reading + CAD, similar to 7.06 per 1000 in double reading	87.2% in single reading + CAD, similar to 87.7% in double reading
Fenton et al. 2011 [10]	Screening, > 1.6 million mammograms at 90 facilities in the BCSC, 1998–2006, USA	Facilities using CAD, 'CAD group' vs. facilities without CAD, 'non-CAD group'	Film mammography	Not specified	3.2, significantly decreased from 3.6 per 1000 before CAD	81.1%, no significant difference from 79.7% before CAD
Lehman et al. 2015 [11]	Screening, 323973 women in 66 facilites of the BCSC, 2003–2009, USA	Mammography interpreted with CAD vs. interpreted without CAD	Film + digital mammography	Not specified	4.1 per 1000, no significant differences to 4.1 in no CAD	85.3%, no significant difference to 87.3% in no CAD

BCSC = Breast Cancer Surveillance Consortium, CAD = computer-aided detection/diagnosis, CADET II = Computer-Aided Detection Evaluation Trial II, UK = United Kingdom, USA = United States

ARTIFICIAL INTELLIGENCE IN MEDICAL IMAGING

The past decade can be marked as an era of 'artificial intelligence (AI)' owing to massive technological advances that have enabled easy access, processing, and storage of huge amounts of data. AI is a branch of computer science dedicated to developing algorithms that accomplish tasks that are traditionally associated with human intelligence [16], and AI is already being applied to simpler technical tasks such as speech and text recognition, language processing, object detection, and classification [17,18,19]. AI has brought about both excitement and concern because it is expected to increase the value and efficiency of medical imaging [20].

Computer-extracted features can serve as input to 'machine learning' algorithms, a subset of AI that use complicated statistical techniques to enable machines to improve at certain tasks by learning data patterns [18]. Further, 'deep learning' (DL) is a sub-classification of machine learning, wherein multiple layered neural networks are used to assess complex patterns within the input data. After the introduction of the 2012 ImageNet Large Scale Visual Recognition Challenge [21], deep convolutional neural networks (dCNNs) are now the technique of choice for computer visualizations and are used in various fields of image classification including breast imaging. When provided with raw data, dCNNs discover features or combinations of features that are associated with or predictive of a specific outcome (in this case, 'breast cancer detection' or 'breast cancer diagnosis') instead of requiring them to be delicately crafted by humans, which is referred to asrepresentation learning; thereby the software trains itself to perform the task as long as sufficient quality and quantity of data are provided [18,22].

APPLICATION OF AI-CAD TO MAMMOGRAPHY INTERPRETATION

Several AI-based algorithms have been developed for mammography, and the results clearly show that their use has considerably narrowed the gap between the diagnostic performances of computers and humans (Table 2). Majority of the in-house AI-algorithms show high performance in breast cancer detection; in some studies, the performance of AI-CAD is similar to or even superior to those of radiologists. Moreover, even the stand-alone performance of AI-CAD has been significantly higher than the average performance of radiologists [23,24,25]. In a study comparing the predictions of AI algorithms and readers using representative screening data from the United Kingdom (UK, double reading) and the United States (USA, single reading), use of AI resulted in significant improvement with a sensitivity of 2.7% and 9.4%, and a specificity of 1.2% and 5.7%, respectively [24]. Further, non-inferiority was seen for both sensitivity and specificity for the second reader when using AI to analyze the UK data, reducing the workload for the second reader by 88%.

Table 2

Diagnostic Performances of AI-CAD When Applied to Digital Mammography Interpretation

References	Purpose	Cancer Proportion	AI Category	External Validation*	AUC	Sensitivity
Kooi et al. 2017 [28]	Compare between mammography CAD vs. CNN CNN vs. radiologists for test set	1.5% (271 annotated cancers in 18182 images)	Deep CNN	No, 18453 images from 2188 cases for test set	CAD 0.910 vs. CNN 0.929 Test set: CNN 0.878, radiologists 0.911	-
Becker et al. 2017 [86]	Evaluate diagnostic accuracy of deep learning-based software	7.7% (18 of 233 cases)	dANN	No, 30% saved for validation	0.840 (experienced readers: 0.890, inexperienced readers: 0.790)	84.2% (84.2%, 84.2%)
Al-Masni et al. 2018 [87]	Detection and classification of masses on DM	50.0% (300 of 600 cases)	ROI-based CNN	No	0.877	93.2%
Bandeira Diniz et al. 2018 [88]	Detection of mass/non-mass regions in non-dense and dense breast	- (2482 images from 1241 women)	Deep CNN	No, 20% saved as test set	-	91.5% in non-dense, 90.4% in dense breast
Ribli et al. 2018 [89]	Propose a CAD system that detects and classifies malignant or benign lesions	- (2949 cases)	Faster R-CNN	Yes, DM DREAM challenge (AUC 0.85)	0.950	-
Chougrad et al. 2018 [90]	Deep learning CAD to aid radiologists to	51.0%	Deep CNN	Yes, MIAS database	DDSM 0.98, INIbreast 0.97, BCDR	

*With independent test set. AI = artificial intelligence, AUC = area under the receiving operator characteristics curve, BCDR = Breast Cancer Digital Repository, CAD = computer-aided detection/diagnosis, CNN = convolutional neural network, dANN = deep artificial neural networks, DDSM = Digital Database of Screening

Mammography, DM = digital mammography, MIAS = Mammographic Image Analysis Society, ROI = region-of-interest, UK = United Kingdom, USA = United States

In addition to the stand-alone results, two studies using commercially available deep-learning CAD algorithms for mammography interpretation showed that interpretation with AI-based CAD improved the diagnostic performance of radiologists compared to interpretation without AI-CAD; the area under the receiving operator characteristics curve (AUC) of radiologists alone improved after using AI-CAD from 0.810 to 0.881 ($p < 0.001$) [23] and from 0.87 to 0.89 ($p = 0.002$) [26]. AI-CAD showed superior performance in detecting cancers that presented as masses, distortions, asymmetries, were early stage, node-negative invasive cancers, or cancers in mammographically dense breasts [23], indicating that AI-CAD could overcome major difficulties in breast cancer detection using mammography.

Such promising results have increased expectations regarding the role of AI-CAD in screening [27]. Most recent studies on AI algorithms are from cancer-enriched populations (Table 2), with the exception of one study (1.1%, conventional CAD vs. CNN) [28], wherein the cancer proportion ranges from 7.7–50%. The results of the DM Dialog on Reverse Assessment and Methods (DM DREAM) Challenge may give us an idea of how AI-CAD will perform in screening settings: 144231 screening mammograms including 952 (0.7%) cancers were used for algorithm training/validation and tested on a second independent validation cohort of 166578 examinations including 780 (0.5%) cancers [29]. This study showed that while no single AI algorithm outperformed the radiologists, the combination of AI with radiologists resulted in a higher AUC of 0.942, significantly improving specificity and overall accuracy [29]. In another study that conducted an external evaluation of three commercially available AI-CAD algorithms as independent mammography readers, high AUC values ranging from 0.920–0.956 with high sensitivities of 67.0–81.9% (77.4% for the first reader radiologist and 80.1% for the second reader radiologist) were achieved at the same specificity [30]. These findings support that AI-CAD can indeed contribute to improving performance in a real-world screening environment. However, these results need to be validated in future prospective studies.

DEEP LEARNING APPLIED TO DIGITAL BREAST TOMOSYNTHESIS

DBT provides multiple low-dose projection images of the breast that can be used to reconstruct a three-dimensional dataset of mammography images [31]. This in turn reduces the negative effect of overlapping breast tissues. Studies have proven the efficacy of DBT over full-field DM when used for breast cancer screening, with several reports of increased cancer detection rates and decreased recall rates [32,33,34]. Although DBT may show superior performance, its acquisition time is longer and its interpretation time is reported to be almost twice the time needed for DM [35,36,37], a factor that may critically impact the workload of radiologists.

Even with the superior performance of DBT over DM, perception or interpretation errors still occur [38]. Compared to the single images used to interpret each plane in DM, for DBT interpretation, radiologists have to scroll through stacked images for each mammographic projection, where the number of images per stack is proportional to the breast thickness under compression. More images mean a heavier workload, and this is the main reason for longer interpretation time and radiologists' fatigue with DBT. Automated detection of

abnormalities among multiple projection images could help clinicians localize and assess the clinical significance of a detected abnormality. Commercial or in-house softwares have been developed to assist DBT interpretation, for which initial studies commonly report a reduction in reading time while maintaining reader performance (Table 3) [39,40,41]. The results were more promising when an AI-based algorithm was applied to DBT interpretation [35], with the reading time being reduced by 52.7% (64.1 to 30.4 seconds) along with improvements in all diagnostic metrics such as increased sensitivity (77.0% to 85.0%), specificity (62.7% to 69.6%), AUC (0.795 to 0.852), and decreased recall rates (38.0% to 30.9%).

Table 3

Summary of Studies Using AI-CAD for DBT Interpretation

References	No. of Readers	No. of DBT Exams	Software Used	Reading Time	AUC		Sensitivity (%)		Specifici
					Without CAD	With CAD	Without CAD	With CAD	Without CAD
Balleyguier et al. 2017 [39]	6	80 (23 cancers)	PowerLook Tomo Detection, iCAD	↓ 48.2 sec† ↓ 23.5%	0.854	0.850	86.5	86.5	57.6
Benedikt et al. 2018 [40]	20	240 (60 cancers)	PowerLook Tomo Detection, iCAD*	↓ 19.6 sec† ↓ 29%	0.841	0.850	84.7	87.1	52.7
Chae et al. 2019 [41]	4	100 (70 cancers)	In-house DBT CAD system	↓ 10.04 sec† ↓ 14%	0.778	0.776	77.5	78.6	66.7
Conant et al. 2019 [35]	24	260 (65 cancers)	PowerLook Tomo Detection, iCAD	↓ 34.7 sec† ↓ 52.7%	0.795	0.852	77.0	85.0†	62.7

*CAD system focuses on detecting soft tissue lesions and does not detect calcifications, †With statistical significance. AI = artificial intelligence, AUC = area under the receiving operator characteristics curve, CAD = computer-aided detection/diagnosis, DBT = digital breast tomosynthesis

IMPROVING MAMMOGRAPHY INTERPRETATION WORKFLOW WITH AI-CAD

In addition to improving breast cancer detection, studies have evaluated AI-CAD for triaging mammography examinations, which is critical for screening because the majority of screening exams are negative. If AI-CAD can accurately identify cases that require less time and fewer resources without endangering the patient, we can reduce the workload of radiologists, and more time can be spent on images with suspicious features and on any subsequent diagnostic

workup. Three recent studies have used AI-CAD to triage mammography examinations in this way [42,43,44]. In an initial study, the probability of malignancy score (scale of 0–10) generated by a commercially available AI-CAD was used to predesignate cancer-free examinations as 'normal' so that they were not listed as cases that needed further interpretation by radiologists. In this study, setting the threshold at an AI score of 5 resulted in approximately 50% workload reduction with 7% of cancers being interpreted as false-negatives, whereas setting the threshold at an AI score of 2 resulted in a 17% workload reduction with 1% of cancers being interpreted as false-negatives [43]. Preselection of examinations according to the AI score did not change the average AUC of radiologists, except for when the AI score was 9. This finding indicated that triaging with AI-CAD maintains radiologist performance even as it reduces the workflow. In the other two studies, DL algorithms were constructed to triage mammography examinations as cancer-free using either imaging features [44] or imaging features combined with non-imaging features and pathologic outcomes [42]. These recent results corroborate previous studies that reported a workload reduction of approximately 20–34% to 91% (in a screening setting) albeit with non-inferior levels of sensitivity, negative predictive values, and improved specificity.

Triaging mammography examinations offers two advantages: First, triaging negative examinations can spare the radiologists' time and effort, thereby reducing their overall workload, as previously mentioned. Second, by identifying cancers that have been missed by radiologists, AI can act as a final consultant (Fig. 3). A simulation study performed with this in mind reported that AI can reduce the radiologists' workload by more than half [45]. Further, by using AI scores for women with negative double readings it also increases early detection of interval cancers by 12–50% and of next-round screen-detected cancers by 14–59%, depending on which AI score is used as the cutoff. In this study, AI scores showed higher accuracy for predicting future interval cancers and next-round screen-detected cancers than mammographic density, and the researchers speculated that this difference was due to AI algorithms detecting subtle unidentified tumor features, whereas mammographic density was associated with tumor masking. These results suggest that AI may have another potential role in supplementary screening for women after negative findings on the mammograph. The role of AI in breast cancer risk prediction has been discussed in depth in the next section.
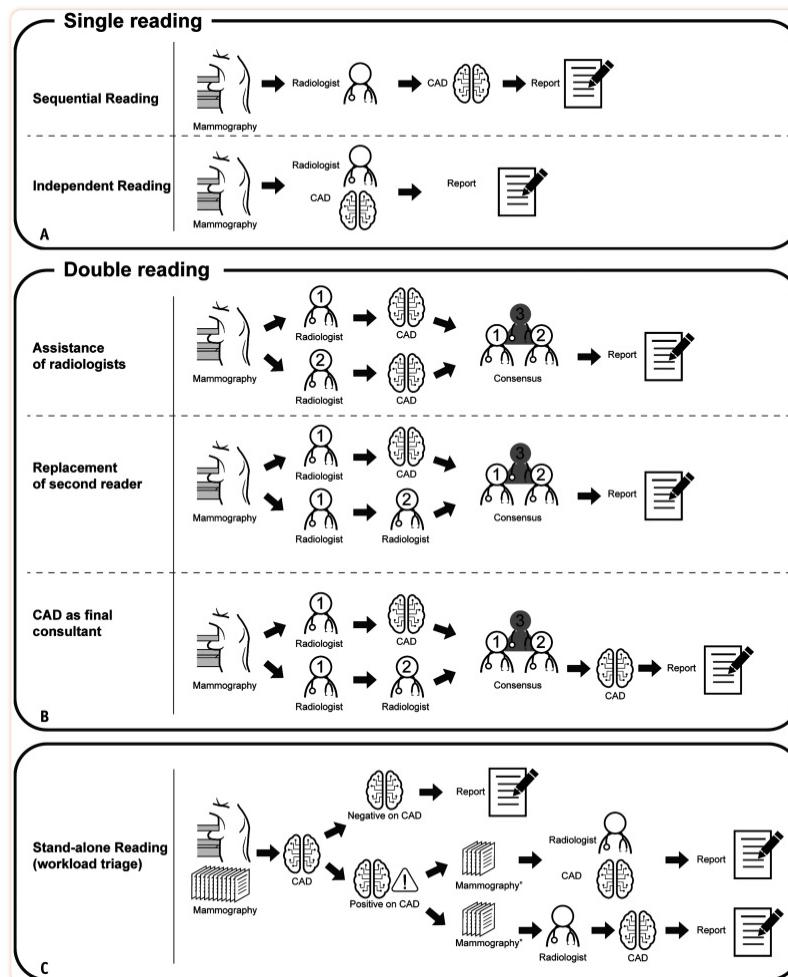
[Fig. 3](#)

Diagram of various workflows using AI-CAD for mammography interpretation.

Before applying AI-CAD to mammography interpretation, the following aspects should be considered: **A.** The impact of AI-CAD differs depending on whether the radiologists are exposed to the analytic results, i.e., sequential reading, wherein the AI-CAD result is provided after the radiologist has reached a conclusion vs. independent reading, wherein the AI-CAD result is provided with the mammography images before the radiologist has reached a conclusive interpretation. **B.** The impact of AI-CAD also differs according to the reading environment, i.e., single reading vs. double reading, wherein AI-CAD may not only assist radiologists, but even has the potential to replace a second reader or serve as a final consultant in cases where a consensus among multiple readers is required. **C.** If AI-CAD can be considered for stand-alone reading, its analytic results can be used for workload triage such that examinations with 'negative' results on AI-CAD will not be interpreted by radiologists. AI = artificial intelligence, CAD = computer-aided detection/diagnosis

ASSESSMENT OF MAMMOGRAPHIC PARENCHYMAL DENSITY USING DEEP LEARNING

Breast parenchymal density is important in two aspects: 1) an increased proportion of fibroglandular tissues has been associated with a four-to six-fold increased risk for breast cancer [46,47,48]; and 2) the detection sensitivity of mammography can be severely affected by increased parechymal density or 'dense breast' as breast masses can be potentially masked, leading to increased rates of interval cancers [47,49,50], which is why additional screening

modalities such as US or MRI are used. Legislation on breast density notification was first passed in Connecticut, USA in 2009, and then in more than 30 states of the USA. Radiologists are now required to notify women on their breast density after they are screened with mammography and they also need to discuss the possibility of missed cancers in dense breasts [51,52]. Based on this requirment, we need to report parenchymal density on mammography using quantitative and objective analysis, but the qualitative four-tiered density categories of the American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS) [53] depend solely on the radiologist's subjective interpretation and vary widely, with a κ = 0.40–0.87 reported in literature [54,55,56]. Currently, several commercially-available automated volumetric density measurement programs (Volpara, Volpara Solutions; Quantra, Hologic) enable the quantification of parenchymal data by calculating the ratio of fibroglandular tissue to the total breast volume in percentages [57].

Even with automation, inconsistencies between measurements made by radiologists and current softwares remains an issue. Studies have reported differences of 14–22.3% and fair to moderate agreement (κ = 0.32–0.61) between radiologists and commercially-available softwares in density classification [57,58]. To narrow the gap between computers and radiologists in density assessment, DL algorithms have been constructed and applied in several recent studies. In experimental settings, several state-of-the-art DL models showed strong similarity or agreement with the BI-RADS density assessments made by radiologists [59,60,61,62,63]. When implemented in routine clinical practice, a DL algorithm showed excellent agreement with radiologists with κ = 0.85. Further, 94% of the assessments made by the DL algorithm were accepted by the radiologists in binary categorization of non-dense or dense breasts [64]. Although DL models are in the early stages of development, they show potential for providing consistent and reliable data for breast density, which is both useful and important, especially against the backdrop of breast density legislation. Consistent and objective data are required to predict breast cancer risk and to discuss the need for supplementary studies and future management plans with the patient, which is why we expect DL algorithms to have a greater role in the assessment of parenchymal density in the future.

ASSESSMENT OF BREAST CANCER RISK USING DEEP LEARNING

Starting with the publication of the Gail model [65], various risk models have been developed during the past decade using multiple risk factors related to hormonal and genetic information to predict breast cancer risk [66,67,68]. Until recently, image-related information was not incorporated into the risk prediction models, but now mammographic breast density is a well-acknowledged risk factor [47]. Previous studies have shown that the AUC for predicting breast cancer risk has significantly improved by adding subjective mammographic parenchymal density as a risk factor [68,69]. As mentioned in our discussion on mammographic breast density, DL models have been introduced to enable more objective and quantitative density assessment. Not only do DL algorithms provide consistent assessments of density, they are also thought to provide more accurate breast cancer risk predictions based on pixel-based information embedded in mammographic images that are not perceptible to the human eye. This hypothesis has been supported in several recent publications: DL CNN using mammography for a pixel-based prediction of breast cancer risk had greater predictive potential than using breast density assessments by radiologists (odds ratio 4.42 vs. 1.67) with

an overall accuracy of 72% [70]. Similarly, risk scores generated by a DL network allowed more accurate predictions of future breast cancer risk with lower false-negative rates for more aggressive cancers when compared to density-based models [71].

As the amount of readily available data has increased with corresponding advances in processing abilities, attempts have been made to use computer-analyzed imaging data as input in the DL models and combine it with traditional risk factors obtained from medical records to predict breast cancer risk with promising results. By using computerized image analysis to extract features of parenchymal texture, radiomic phenotypes resulted in significantly higher discrimination performance when added to a breast cancer risk model that included breast density and body mass index as risk factors (AUC 0.84 vs. 0.80) [72]. Similarly, a hybrid DL model using both traditional risk factors and mammograms showed the highest diagnostic performance (AUC 0.70) compared to a clinical risk-factor-based model (AUC 0.62–0.67) or image-only DL model (AUC 0.68) [73]. By using a DL model with a broader range of input data extracted from electronic health records linked to mammographic data, the algorithm showed the potential to assess breast cancers at levels comparable to radiologists (AUC 0.91, sensitivity 87%, specificity 77.3%), and detected 48% of false-negative mammography interpretations [74]. These recent studies indicate that image-based DL models show promise for more accurate breast cancer risk prediction and that we can expect more from these models in the future including personalized management for women. This is thought to be especially relevant to clinical practice as the passing of the breast density notification legislation in the USA brought about increased medical costs and workload burdens for breast cancer screening [75,76]. The applications of DL algorithms for breast cancer risk prediction are still in the early stages of development, and we anticipate studies evaluating the effect of DL algorithms when selecting women for intense breast cancer screening in the near future.

CHALLENGES TO OVERCOME BEFORE CLINICAL APPLICATION

Even with rapid progress, more investigation is required to prove the utility of AI in aiding radiologists with mammography in real-world settings. First, proper external validation is required [17]; while various machine learning techniques have been used to develop CAD algorithms for breast imaging modalities during the last few years [77], very few have been commercially utilized for clinical use, mostly due to the lack of clinical validation. We need studies that demonstrate how AI will work in the real world while considering generalizability, efficiency, user variability, and ways to optimize algorithms for consistent outcomes [78]. A recent study by Salim et al. [30] performed an external evaluation of currently released, commercially-available AI-CAD algorithms as independent readers of mammography and also in combination with radiologists. They found that commercially available AI-CAD algorithms can assess screening mammograms with sufficient diagnostic performance to be evaluated as an independent reader in future prospective studies, and more positive cases were detected when combining AI with radiologists than double reading with only radiologists. This study focused on the external validation of AI, and we hope to see more similar studies that will evaluate AI algorithms based on the features they are supposed to analyze.

Second, feasibility testing should be conducted while considering certain clinical aspects such as the incorporation of AI-CAD in clinical practice with comparisons of screening vs. diagnostic settings, single reading vs. double reading, or for reading sequences etc (Fig. 3). In addition, the methodology for estimating the true performance of an AI algorithm needs to be refined.

As its name implies, the role of conventional CAD is to 'assist' the radiologist who can either choose to follow the CAD prompts or to neglect them when reaching a final diagnosis. Typically, final diagnoses are made by radiologists with the CAD results also taken into consideration. When assessing the performance of CAD under these circumstances— circumstances reflecting real-world practice since radiologists are the ones legally responsible for their interpretation, not CAD— the gain achieved by using a good AI-CAD may be underestimated [79]. In addition, although it is difficult to estimate related data, we need to consider how to measure the time and resources invested into interpretation to evaluate whether integrating the AI algorithm improves the efficiency of interpretation workflows.

Third, there are more technical issues to consider when developing AI algorithms for DBT than DM. Since most of the current DL algorithms have been built using two-dimensional mammography, transfer learning has been adapted through the application of a pre-trained CNN to build algorithms for DBT interpretation [80]. Furthermore, poor spatial resolution per image limits the detection accuracy of tomosynthesis, and consequently we would expect lower performance for DL algorithms with DBT compared to DM. In addition, DBTs acquired from different vendors have different angular ranges, acquisition techniques, pixel binning, and reconstruction techniques that affect mammography images [81].

Finally, there are ethical or legal issues to consider before incorporating AI into our interpretation workflow. Should AI be considered as an independent reader? Should the analytic data provided by AI algorithms be reported in medical records? As of now, the interpreting radiologist is legally responsible for his or her image interpretation regardless of whether they choose to use the AI marks, but the legal ramifications of using AI data and the degree of its use should be discussed in depth. If AI-CAD truly becomes part of everyday practice, an ethical or legal framework for the application of AI algorithms will be required, and this framework must reflect a consensus of all participants in a real-world mammography setting, from patients to radiologists. In addition, we need to be prepared for the unintended consequences of incorporating AI algorithms such as the detection of many *in situ* cancers rather than invasive cancers and the regression of interpretive skills due to radiologists' overdependence on AI [5,19].

CONCLUSION

Recent advances in technology have enabled the application of AI to mammography, with stand-alone diagnostic performances comparable to those of radiologists, improvements in sensitivity or specificity in breast cancer diagnosis, and the potential reduction in the workflow or interpretation time. Although the current results for several AI algorithms in mammography seem quite positive, clinical validation is required to guarantee generalizability, efficiency, and consistency. Social consensus is also required for the role AI algorithms will play in mammography interpretation, along with ethical and legal considerations. Although AI is still in the preliminary stages of validation, there is increasing demand for its application in the medical field, and more effort is being put into implementing AI technology in actual clinical settings. Moreover, with the initiative towards prospective clinical validation studies, we need to be prepared to accept AI in clinical practice and to be aware of the impact it may have on the future of breast imaging.

Acknowledgments

The authors thank Medical Illustration & Design, part of the Medical Research Support Services of Yonsei University College of Medicine, for artistic support related to this work.

Footnotes

References

1. World Health Organization. *IARC handbooks. Breast cancer screening. Volume 15.* Lyon: International Agency for Research on Cancer; 2015. [Google Scholar]

2. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghate S, et al. Benefits and harms of breast cancer screening: a systematic review. *JAMA.* 2015;314:1615–1634. [PubMed] [Google Scholar]

3. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening-- viewpoint of the IARC Working Group. *N Engl J Med.* 2015;372:2353–2358. [PubMed] [Google Scholar]

4. Taylor-Phillips S, Stinton C. Double reading in breast cancer screening: considerations for policy-making. *Br J Radiol.* 2020;93:20190610. [PMC free article] [PubMed] [Google Scholar]

5. Houssami N, Lee CI, Buist DSM, Tao D. Artificial intelligence for breast cancer screening: opportunity or hype? *Breast.* 2017;36:31–33. [PubMed] [Google Scholar]

6. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin.* 2019;69:127–157. [PMC free article] [PubMed] [Google Scholar]

7. Abbasi J. Artificial intelligence improves breast cancer screening in study. *JAMA.* 2020;323:499 [PubMed] [Google Scholar]

8. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology.* 2001;220:781–786. [PubMed] [Google Scholar]

9. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. *Radiology.* 2005;236:451–457. [PubMed] [Google Scholar]

10. Fenton JJ, Abraham L, Taplin SH, Geller BM, Carney PA, D'Orsi C, et al. Effectiveness of computer-aided detection in community mammography practice. *J Natl Cancer Inst.* 2011;103:1152–1161. [PMC free article] [PubMed] [Google Scholar]

11. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med.* 2015;175:1828–1837. [PMC free article] [PubMed] [Google Scholar]

12. Khoo LA, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. *Radiology.* 2005;237:444–449. [PubMed] [Google Scholar]

13. Malich A, Marx C, Facius M, Boehm T, Fleck M, Kaiser WA. Tumour detection rate of a new commercially available computer-aided detection system. *Eur Radiol.* 2001;11:2454–2459. [PubMed] [Google Scholar]

14. 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;356:1399–1409. [PMC free article] [PubMed] [Google Scholar]

15. Cole EB, Zhang Z, Marques HS, Edward Hendrick R, Yaffe MJ, Pisano ED. Impact of computer-aided detection systems on radiologist accuracy with digital mammography. *AJR Am J Roentgenol.* 2014;203:909–916. [PMC free article] [PubMed] [Google Scholar]

16. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism.* 2017;69S:S36–S40. [PubMed] [Google Scholar]

17. Park SH, Kressel HY. Connecting technological innovation in artificial intelligence to real-world medical practice through rigorous clinical validation: what peer-reviewed medical journals could do. *J Korean Med Sci.* 2018;33:e152. [PMC free article] [PubMed] [Google Scholar]

18. Giger ML. Machine learning in medical imaging. *J Am Coll Radiol.* 2018;15:512–520. [PubMed] [Google Scholar]

19. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA.* 2017;318:517–518. [PubMed] [Google Scholar]

20. Chang PJ. Moving artificial intelligence from feasible to real: time to drill for gas and build roads. *Radiology.* 2020;294:432–433. [PubMed] [Google Scholar]

21. Litjens G, Kooi 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. [PubMed] [Google Scholar]

22. Mendelson EB. Artificial intelligence in breast imaging: potentials and limitations. *AJR Am J Roentgenol.* 2019;212:293–299. [PubMed] [Google Scholar]

23. Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health.* 2020;2:e138–e148. [PubMed] [Google Scholar]

24. 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;577:89–94. [PubMed] [Google Scholar]

25. 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;111:916–922. [PMC free article] [PubMed] [Google Scholar]

26. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology.* 2019;290:305–314. [PubMed] [Google Scholar]

27. Trister AD, Buist DSM, Lee CI. Will machine learning tip the balance in breast cancer screening? *JAMA Oncol.* 2017;3:1463–1464. [PMC free article] [PubMed] [Google Scholar]

28. Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal.* 2017;35:303–312. [PubMed] [Google Scholar]

29. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw Open.* 2020;3:e200265. [PMC free article] [PubMed] [Google Scholar]

30. Salim M, Wåhlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncol.* 2020;6:1581–1588. [PMC free article] [PubMed] [Google Scholar]

31. Vedantham S, Karellas A, Vijayaraghavan GR, Kopans DB. Digital breast tomosynthesis: state of the art. *Radiology.* 2015;277:663–684. [PMC free article] [PubMed] [Google Scholar]

32. Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol.* 2013;14:583–589. [PubMed] [Google Scholar]

33. Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA.* 2014;311:2499–2507. [PubMed] [Google Scholar]

34. McCarthy AM, Kontos D, Synnestvedt M, Tan KS, Heitjan DF, Schnall M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *J Natl Cancer Inst.* 2014;106:dju316. [PMC free article] [PubMed] [Google Scholar]

35. Conant EF, Toledano AY, Periaswamy S, Fotin SV, Go J, Boatsman JE, et al. Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis. *Radiol Artif Intell.* 2019;1:e180096. [PMC free article] [PubMed] [Google Scholar]

36. Gilbert FJ, Tucker L, Gillan MG, Willsher P, Cooke J, Duncan KA, et al. Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY Trial) *Radiology.* 2015;277:697–706. [PubMed] [Google Scholar]

37. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology.* 2013;267:47–56. [PubMed] [Google Scholar]

38. Korhonen KE, Weinstein SP, McDonald ES, Conant EF. Strategies to increase cancer detection: review of true-positive and false-negative results at digital breast tomosynthesis screening. *Radiographics.* 2016;36:1954–1965. [PMC free article] [PubMed] [Google Scholar]

39. 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;97:83–89. [PubMed] [Google Scholar]

40. 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. *AJR Am J Roentgenol.* 2018;210:685–694. [PubMed] [Google Scholar]

41. Chae EY, Kim HH, Jeong JW, Chae SH, Lee S, Choi YW. Decrease in interpretation time for both novice and experienced readers using a concurrent computer-aided detection system for digital breast tomosynthesis. *Eur Radiol.* 2019;29:2518–2525. [PubMed] [Google Scholar]

42. Kyono T, Gilbert FJ, van der Schaar M. Improving workflow efficiency for mammography using machine learning. *J Am Coll Radiol.* 2020;17:56–63. [PubMed] [Google Scholar]

43. 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;29:4825–4832. [PMC free article] [PubMed] [Google Scholar]

44. Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A deep learning model to triage screening mammograms: a simulation study. *Radiology.* 2019;293:38–46. [PubMed] [Google Scholar]

45. Dembrower K, Wåhlin E, Liu Y, Salim M, Smith K, Lindholm P, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *Lancet Digit Health.* 2020;2:e468–e474. [PubMed] [Google Scholar]

46. Harvey JA, Bovbjerg VE. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. *Radiology.* 2004;230:29–41. [PubMed] [Google Scholar]

47. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med.* 2007;356:227–236. [PubMed] [Google Scholar]

48. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2006;15:1159–1169. [PubMed] [Google Scholar]

49. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst.* 2000;92:1081–1087. [PubMed] [Google Scholar]

50. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA.* 1996;276:33–38. [PubMed] [Google Scholar]

51. Bahl M, Baker JA, Bhargavan-Chatfield M, Brandt EK, Ghate SV. Impact of breast density notification legislation on radiologists' practices of reporting breast density: a multi-state study. *Radiology.* 2016;280:701–706. [PubMed] [Google Scholar]

52. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. *Radiology.* 2012;265:59–69. [PubMed] [Google Scholar]

53. American College of Radiology. *Breast imaging reporting and data system.* 5th ed. Reston, VA: American College of Radiology; 2013. [Google Scholar]

54. Spayne MC, Gard CC, Skelly J, Miglioretti DL, Vacek PM, Geller BM. Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study. *Breast J.* 2012;18:326–333. [PMC free article] [PubMed] [Google Scholar]

55. Gard CC, Aiello Bowles EJ, Miglioretti DL, Taplin SH, Rutter CM. Misclassification of breast imaging reporting and data system (BI-RADS) mammographic density and implications for breast density reporting legislation. *Breast J.* 2015;21:481–489. [PMC free article] [PubMed] [Google Scholar]

56. 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;165:457–464. [PMC free article] [PubMed] [Google Scholar]

57. Youk JH, Gweon HM, Son EJ, Kim JA. Automated volumetric breast density measurements in the era of the BI-RADS fifth edition: a comparison with visual assessment. *AJR Am J Roentgenol.* 2016;206:1056–1062. [PubMed] [Google Scholar]

58. Brandt KR, Scott CG, Ma L, Mahmoudzadeh AP, Jensen MR, Whaley DH, et al. Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. *Radiology.* 2016;279:710–719. [PMC free article] [PubMed] [Google Scholar]

59. 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;35:1322–1331. [PubMed] [Google Scholar]

60. Lee J, Nishikawa RM. Automated mammographic breast density estimation using a fully convolutional network. *Med Phys.* 2018;45:1178–1190. [PMC free article] [PubMed] [Google Scholar]

61. Mohamed AA, Luo Y, Peng H, Jankowitz RC, Wu S. Understanding clinical mammographic breast density assessment: a deep learning perspective. *J Digit Imaging.* 2018;31:387–392. [PMC free article] [PubMed] [Google Scholar]

62. Ciritsis A, Rossi C, Vittoria De Martini I, Eberhard M, Marcon M, Becker AS, et al. Determination of mammographic breast density using a deep convolutional neural network. *Br J Radiol.* 2019;92:20180691. [PMC free article] [PubMed] [Google Scholar]

63. Mohamed AA, Berg WA, Peng H, Luo Y, Jankowitz RC, Wu S. A deep learning method for classifying mammographic breast density categories. *Med Phys.* 2018;45:314–321. [PMC free article] [PubMed] [Google Scholar]

64. 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;290:52–58. [PubMed] [Google Scholar]

65. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989;81:1879–1886. [PubMed] [Google Scholar]

66. 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;28:115–120. [PubMed] [Google Scholar]

67. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med.* 2004;23:1111–1130. [PubMed] [Google Scholar]

68. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat.* 2005;94:115–122. [PubMed] [Google Scholar]

69. Brentnall AR, Harkness EF, Astley SM, Donnelly LS, Stavrinos P, Sampson S, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res.* 2015;17:147. [PMC free article] [PubMed] [Google Scholar]

70. 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;26:544–549. [PMC free article] [PubMed] [Google Scholar]

71. Dembrower 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;294:265–272. [PubMed] [Google Scholar]

72. Kontos D, Winham SJ, Oustimov A, Pantalone L, Hsieh MK, Gastounioti A, et al. Radiomic phenotypes of mammographic parenchymal complexity: toward augmenting breast density in breast cancer risk assessment. *Radiology.* 2019;290:41–49. [PMC free article] [PubMed] [Google Scholar]

73. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology.* 2019;292:60–66. [PubMed] [Google Scholar]

74. Akselrod-Ballin A, 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;292:331–342. [PubMed] [Google Scholar]

75. Houssami N, Lee CI. The impact of legislation mandating breast density notification - Review of the evidence. *Breast.* 2018;42:102–112. [PMC free article] [PubMed] [Google Scholar]

76. Saulsberry L, Pace LE, Keating NL. The impact of breast density notification laws on supplemental breast imaging and breast biopsy. *J Gen Intern Med.* 2019;34:1441–1451. [PMC free article] [PubMed] [Google Scholar]

77. Yassin NIR, Omran S, El Houby EMF, Allam H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. *Comput Methods Programs Biomed.* 2018;156:25–45. [PubMed] [Google Scholar]

78. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J Radiol.* 2019;20:405–410. [PMC free article] [PubMed] [Google Scholar]

79. Sechopoulos I, Mann RM. Stand-alone artificial intelligence-The future of breast cancer screening. *Breast.* 2020;49:254–260. [PMC free article] [PubMed] [Google Scholar]

80. Mendel K, Li H, Sheth D, Giger M. Transfer learning from convolutional neural networks for computer-aided diagnosis: a comparison of digital breast tomosynthesis and full-field digital mammography. *Acad Radiol.* 2019;26:735–743. [PMC free article] [PubMed] [Google Scholar]

81. Geras KJ, Mann RM, Moy L. Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. *Radiology.* 2019;293:246–259. [PMC free article] [PubMed] [Google Scholar]

82. Gur D, Sumkin JH, Rockette HE, Ganott M, Hakim C, Hardesty L, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *J Natl Cancer Inst.* 2004;96:185–190. [PubMed] [Google Scholar]

83. Gilbert FJ, Astley SM, McGee MA, Gillan MG, Boggis CR, Griffiths PM, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. *Radiology.* 2006;241:47–53. [PubMed] [Google Scholar]

84. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: interpretation with computer-aided detection--prospective evaluation. *Radiology.* 2006;239:375–383. [PubMed] [Google Scholar]

85. Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med.* 2008;359:1675–1684. [PubMed] [Google Scholar]

86. 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;52:434–440. [PubMed] [Google Scholar]

87. Al-Masni MA, Al-Antari MA, Park JM, Gi G, Kim TY, Rivera P, et al. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Comput Methods Programs Biomed.* 2018;157:85–94. [PubMed] [Google Scholar]

88. Bandeira Diniz JO, Bandeira Diniz PH, Azevedo Valente TL, Corrêa Silva A, de Paiva AC, Gattass M. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. *Comput Methods Programs Biomed.* 2018;156:191–207. [PubMed] [Google Scholar]

89. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with deep learning. *Sci Rep.* 2018;8:4165. [PMC free article] [PubMed] [Google Scholar]

90. Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening. *Comput Methods Programs Biomed.* 2018;157:19–30. [PubMed] [Google Scholar]

90. Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening. *Comput Methods Programs Biomed.* 2018;157:19–30. [PubMed] [Google Scholar]