Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Concepts and Future Perspectives

Krzysztof J. Geras, PhD • *Ritse M. Mann, MD, PhD* • *Linda Moy, MD*

From the Center for Biomedical Imaging (K.J.G., L.M.), Center for Data Science (K.J.G.), Center for Advanced Imaging Innovation and Research (L.M.), and Laura and Isaac Perlmutter Cancer Center (L.M.), New York University School of Medicine, 160 E 34th St, 3rd Floor, New York, NY 10016; Department of Radiology and Nuclear Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands (R.M.M.); and Department of Radiology, the Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands (R.M.M.). Received November 15, 2018; revision requested January 11, 2019; final revision received April 11; accepted May 13. **Address correspondence to** L.M. (e-mail: *linda.moy@nyulangone.org*).

Conflicts of interest are listed at the end of this article.

Although computer-aided diagnosis (CAD) is widely used in mammography, conventional CAD programs that use prompts to indicate potential cancers on the mammograms have not led to an improvement in diagnostic accuracy. Because of the advances in machine learning, especially with use of deep (multilayered) convolutional neural networks, artificial intelligence has undergone a transformation that has improved the quality of the predictions of the models. Recently, such deep learning algorithms have been applied to mammography and digital breast tomosynthesis (DBT). In this review, the authors explain how deep learning works in the context of mammography and DBT and define the important technical challenges. Subsequently, they discuss the current status and future perspectives of artificial intelligence–based clinical applications for mammography, DBT, and radiomics. Available algorithms are advanced and approach the performance of radiologists—especially for cancer detection and risk prediction at mammography. However, clinical validation is largely lacking, and it is not clear how the power of deep learning should be used to optimize practice. Further development of deep learning models is necessary for DBT, and this requires collection of larger databases. It is expected that deep learning will eventually have an important role in DBT, including the generation of synthetic images.

©RSNA, 2019

Online supplemental material is available for this article.

Multiple randomized clinical trials have demonstrated that screening mammography reduces the mortality from breast cancer by 20%–22% (1,2). As a result, mammography is the cornerstone of breast cancer screening (3,4). In addition, mammography is the initial examination for many women with breast symptoms (5,6). In 2015, 22.6 million mammograms were obtained in the United States alone (7). The evaluation of mammograms thus demands a large number of dedicated radiologists. Unfortunately, there is an increasing shortage of qualified readers in many countries (8). Even for women screened with mammography, as many as one in three cancers manifests as interval cancer; a large proportion of these cancers were, in retrospect, visible on the previous screening mammograms (9,10). Missed cancers at mammography are, therefore, one of the most common reasons for malpractice lawsuits in radiology (11,12). The recent introduction of digital breast tomosynthesis (DBT), in which multiple projections of the breast are obtained over a limited angular range to reconstruct a three-dimensional data set of mammography images (13,14), is only a partial solution. Although DBT depicts 30%–40% more cancers than full-field digital mammography (15,16), the reading time is approximately doubled (17,18) and cognitive and perception errors still occur (19). Consequently, there is a need for assistance with the evaluation of mammography and DBT, both to maximize the cancer detection rate and to address the workload issues.

Automated analysis (ie, artificial intelligence [AI]) of mammograms and DBT images may address these needs. Computer-aided diagnosis (CAD) for mammography has been under development since the late 1960s (20). Its primary aim is to assist radiologists in identifying subtle cancers that might otherwise be missed. CAD programs mark focal areas of increased density

Abbreviations

AI = artificial intelligence, CAD = computer-aided diagnosis, CNN = convolutional neural network, DBT = digital breast tomosynthesis

Summary

Because of the advances in deep learning, the quality of artificial intelligence is rapidly improving for breast imaging and it will likely play an important role for mammography and digital breast tomosynthesis in all steps—from image generation and denoising to risk prediction, cancer detection, and, ultimately, therapy selection and outcome prediction.

Essentials

- In clinical practice, the use of computer-aided diagnosis (CAD) does not improve diagnostic accuracy because the many false prompts lead to higher false-positive rates, recall rates, and biopsy rates.
- Neural networks are capable of learning intermediate, more abstract, representations of the data before classifying the entire image.
- The difference in the appearance of the normal breast parenchyma with digital breast tomosynthesis images obtained with machines from different vendors is much greater than that with full-field digital mammography; this is an important consideration when training deep learning models.
- Previous mammograms and images obtained with different imaging modalities can be exploited to improve the quality of prediction of neural networks.
- The performance of deep learning–based systems is better than that of classic CAD systems based on manually crafted features, approaching that of radiologists for specific tasks.

and microcalcifications. The first CAD software for screening mammography received U.S. Food and Drug Administration approval in 1998 (21). Early results were promising (22–25), and CAD has been widely adopted into clinical practice—with approximately 92% of all mammography facilities in the United States using this technology by 2016 (26). However, its clinical value is uncertain (27,28), mainly due to the large number of false-positive findings.

The success of deep convolutional neural networks (CNNs) in the 2012 ImageNet Large Scale Visual Recognition Challenge (29) triggered new interest in the development of better automated image analysis methods. In the past few years, similar deep neural networks were shown to be highly effective in tasks ranging from face recognition to self-driving cars (30–33). Recent studies have shown that CNNs can also be highly successful in various tasks in the health care industry, ranging from retina analysis to digital pathology (34–36), and in multiple applications in radiology (37–39). Several excellent reviews have been published on the general use of AI in these fields (40–42). Figure 1 illustrates the hierarchy of terms used in AI, as they are not completely interchangeable. It is foreseeable that deep learning will also lead to a major change in the automated analysis of images from mammography and DBT. In this review, we discuss the potential of deep learning techniques for mammography and DBT. In addition, we discuss the current technical approaches to improve on the available CAD systems for mammography and the potential use of these techniques in clinical practice.

Conventional AI

Most available CAD systems perform, in essence, two separate tasks. In the first task, potential lesions that stand out from the normal fibroglandular tissue are detected. The second task entails the reduction in the number of false-positive findings. In this step, the potential lesions are classified and obvious false-positive findings are removed from the list of potential lesions. To perform this task, classic CAD systems depend on human-designed features. For example, masses are detected by using their gray level (how white it is), gradient (if it stands apart from its surroundings), texture (how homogeneous it is), and shape (if it resembles a mass) (43); microcalcifications are detected by actively searching for rod-like high-intensity pixels within the mammogram (Fig 2) (44–46). To reduce the number of false-positive findings, candidate voxels are clustered into possible lesions and analyzed with use of additional features such as distribution, shape, margin, and texture (47–49). To reach a final classification about whether a finding should be flagged, the CAD systems combine the most discriminative features by using a classifier (eg, support vector machines, random forests), and lesions above a predefined threshold are subsequently marked (50).

CAD systems are classified into two groups: computer-aided detection systems and computer-aided diagnosis systems. Computer-aided detection systems focus on the localization task (ie, detection of a suspicious abnormality). They serve as a second reader to radiologists and leave subsequent patient management decisions to the radiologist (51,52). Computer-aided diagnosis systems, however, characterize an abnormality that is identified by a radiologist or a computer. The computer-aided diagnosis system estimates an abnormality's probability of disease and classifies it as benign or malignant. The radiologist then decides whether the abnormality warrants further evaluation and determines its clinical significance (51,52). The advances in the design of classifying features over the years has resulted in a substantial improvement in both the sensitivity and specificity of CAD (51–54). The performance of the top systems reported in the literature approached that of humans when using feature-based classification, albeit only on specific tasks (eg, mass detection) in selected data sets (55,56).

Most conventional CAD systems present their findings in the form of prompts on the mammogram, which requires the radiologist to decide whether the prompts represent an underlying malignancy. Because of the limited specificity of these systems, this is a challenging task. Ikeda et al (57) reported that, when using a feature-based CAD system, approximately 1000 prompts must be analyzed to detect one additional cancer. It is therefore not surprising that the use of CAD in mammography leads to a slightly higher detection rate (range, 1%–19%) when combined with single reading, but at the cost of a lower specificity (incremental recall rate, 6%–36%) and longer evaluation times. Double reading still seems to be significantly better than single reading with CAD (58–65). In clinical practice, the use of CAD does not improve diagnostic accuracy (27,28) because the many false prompts lead to higher false-positive rates, recall rates, and biopsy rates. In addition, the use of CAD does not appear to be
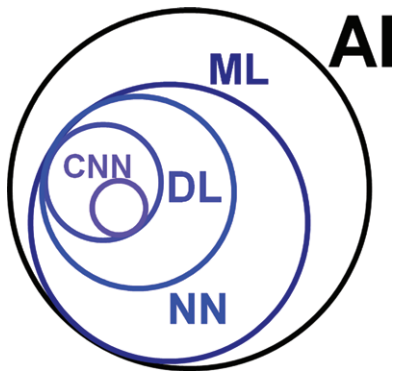
Figure 1: Diagram illustrates the relationship between artificial intelligence (AI), machine learning (ML), neural networks (NN), deep learning (DL), and convolutional neural networks (CNN). AI is the most general of these terms, as it includes systems that aim to mimic human intelligence by learning from data (machine learning) and by applying manually defined decision rules. Machine learning includes neural networks but also pertains to many other methods, such as kernel methods (eg, support vector machines) and decision tree–based methods. Among neural networks, deep learning, which involves study of neural networks consisting of many layers, is currently the most successful in practical applications and the subject of the most intense research. Finally, the type of deep neural networks most frequently applied in medical image analysis are the convolutional neural networks.

Figure 2: Left, mediolateral oblique view from screening mammogram in 54-year-old asymptomatic woman. A computer-aided diagnosis (CAD) prompt is present (arrowhead). Right, magnification view of area of interest. The CAD system identified a small group of calcifications (arrow), which were sampled for biopsy and yielded grade 2 ductal carcinoma in situ.

cost-effective (66). Several studies concluded that CAD applications require substantial improvement to really be beneficial for patient care. Table 1 summarizes most of the literature on the implementation of CAD into the clinical workflow.

Why Deep Learning?

AI, powered by the recent advances in machine learning, may make CAD for mammography more valuable in clinical practice. The most promising of these advances is deep learning—a family of machine learning methods focusing on developing multilayered neural networks (67,68). Like conventional CAD systems, neural networks are mostly trained by using supervised learning, in which every training example comes with an expected output. Logistic regression, decisions trees, and support vector machines, which are used for conventional CAD, are examples of supervised learning models not based on neural networks. However, what these methods have in common is that, although the decision process they use to arrive at the classification decision might be very complex, they do not learn any intermediate representations of the data. That implies that these methods can only work well if the input features they are presented with are very predictive to begin with. However, as apparent from the false-positive findings in conventional CAD, in mammographic evaluation it is very difficult to design features on the level of the input pixels that would allow the classifier to accurately predict the label for the entire image. Neural networks, on the other hand, are capable of learning intermediate, more abstract, representations of the data before
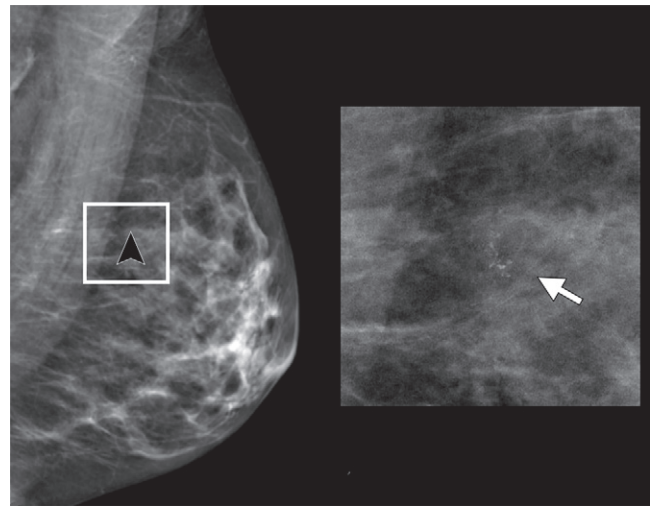
classifying the entire image (68). CNNs only combine information from voxels that are spatially close to each other and are therefore especially suited for image evaluation. This is key to understanding why neural networks work so well for image analysis in comparison to other methods. A more in-depth explanation of the functioning of neural networks and, in particular, CNNs can be found in Appendix E1 (online).

How Deep Learning Works in Mammography

Deep learning models appear to be successful in evaluating mammograms. In the Digital Mammography Dialogue for Reverse Engineering Assessments and Methods, or DREAM, Challenge, held between November 2016 and May 2017, many teams competed in developing machine learning models to classify screening mammograms according to whether cancer was present. Every team used the same data set, which consisted of 640 000 images from more than 86 000 women. The most successful teams used deep learning models (69,70), achieving a sensitivity of up to 87%. This is on par with the 88% sensitivity achieved by radiologists with the same data set. However, only the leaderboard teams achieved a specificity of 82%, a performance that approaches the specificity of the radiologists in the Breast Cancer Surveillance Consortium. The overwhelming majority of deep learning models developed in this challenge were based on relatively simple variations of the CNNs described earlier. A basic way of classifying the various models is according to whether they are trained by using only the examination-level labels (indicating whether the patient under examination has cancer) or both examination-level labels and pixel-level labels (annotations of malignant or benign lesions). The models trained with only examination-level labels are trainable end to end, whereas the models trained with both examination-level and pixel-level labels need a more complex training procedure.

Table 1: Summary of Landmark Decisions or Studies on CAD

Landmark and Year	Reference	Key Findings
First studies on automated analysis of mammograms, 1967	Winsberg et al (20)	Density within scanned mammograms could be automatically characterized
U.S. FDA approval for a CAD application (R2 Image Checker; Hologic, Marlborough, Mass) to detect cancers, 1998	U.S. FDA premarket approval database (21)	A computer system could identify and mark regions of interest on routine screening mammograms
Studies showing that CAD may detect cancers missed by radiologists		
2000	Warren Burhenne et al (22)	CAD could reduce the false-negative rate by 77% without increasing the recall rate
2001	Freer and Ulissey (23)	CAD led to a 19.5% increase in the number of cancers detected
2001	Birdwell et al (24)	CAD marked 77% of cancers missed at screening mammography
Studies showing potential for CAD as an independent second reader of screening mammograms		
2004	Destounis et al (25)	CAD could potentially decrease the false-negative rate at double reading by 39%
2008	Gilbert et al (65)	Single reading with CAD has equal sensitivity at slightly higher recall (3.9% vs 3.4%) compared with double reading
Studies showing that the large amount of false-positive findings generated by classic CAD may have a negative impact on radiologist performance		
2007	Fenton et al (27)	Use of CAD only led to nonsignificant increase in sensitivity and overall lower accuracy (AUC = 0.87 vs 0.92; $P < .01$)
2015	Lehman et al (28)	Use of CAD significantly reduces sensitivity and is not associated with any improved performance parameters

Note.—AUC = area under the receiver operating characteristic curve, CAD = computer-aided diagnosis, FDA = Food and Drug Administration.

Models trained with examination-level labels are usually the most similar to the standard deep CNNs (71–73). They are sometimes modified by taking into account multiple mammographic views simultaneously (71,72) or by adding a multiple-instance layer (73). On the other hand, models that also use pixel-level labels are trained as two separate models in different variations (70,74). Some of these models are also fine-tuned end to end after the two-stage training (74). Models that learn from both examination-level and pixel-level labels generally exhibit higher performance and/or require fewer cases because they learn from a more detailed supervision. However, the data collection is much more laborious, and performance is dependent on the quality of the annotations—which is a difficult problem as there is no real ground truth and interreader variability is substantial (75).

Technical Challenges Unique to Mammography

Not all deep learning algorithms have equal performance. Fine-tuning algorithms to specific tasks in mammography and DBT requires more effort than just the use of a very general CNN on a large data set. For one, it is difficult to train a CNN that is good in the detection of both masses and calcifications. Consequently, separate CNNs are often trained for both lesion types and the outcomes are only combined in the final output of the AI support system. Furthermore, algorithms must be consistent and reproducible over mammograms obtained by different technologists using mammography machines from various vendors. Validation of the deep learning algorithms across different vendors is substantial because all vendors use their own proprietary postprocessing data to make the mammograms ready for presentation, and the raw data are usually not stored. This has a large influence on image appearance and implies that a CNN trained on mammograms obtained with a machine from one vendor may not be applicable to mammograms obtained with a machine from another vendor (Fig 3) (76,77). Consequently, normalization of mammograms is an important task that must be carried out with machine learning techniques.

Most efforts in deep learning focused on applying existing techniques to mammography rather than proposing new ones specifically suited to the domain. Medical images have properties that make them very different from images from natural scenes (eg, images of a tree or dog) that are not synthetic or artificial (70). For example, although the objects of interest that
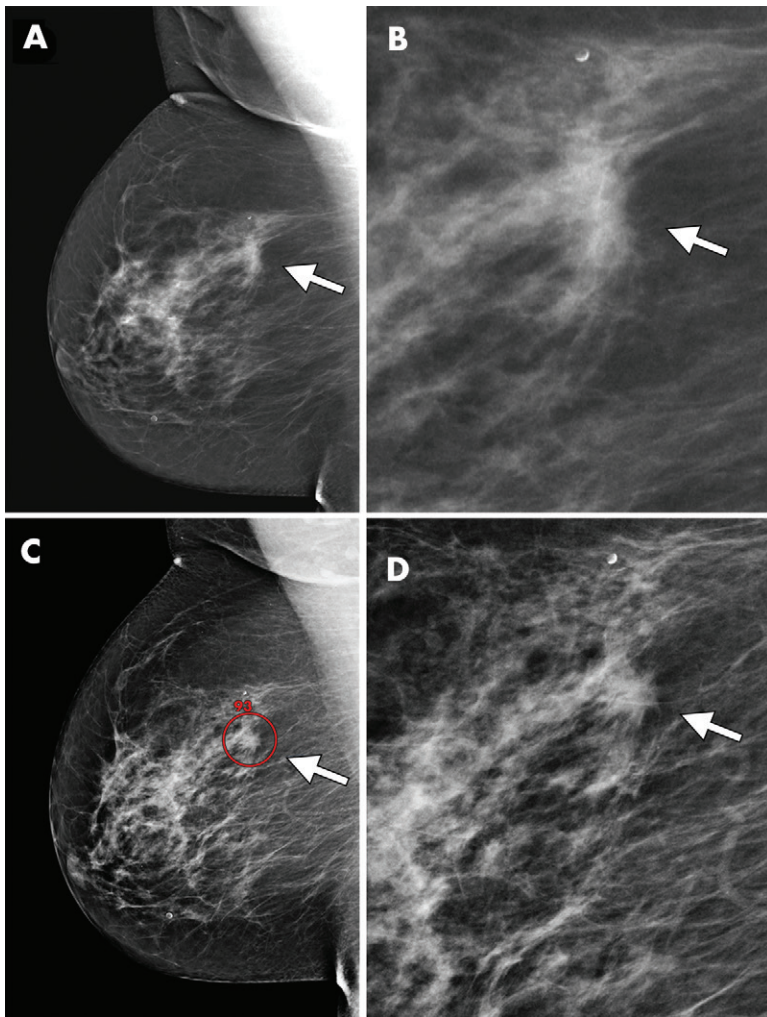
Figure 3: Images in 58-year-old asymptomatic woman who presented for screening mammography. Arrows indicate cancer. *A,* Right mediolateral oblique screening mammogram shows an asymmetry with associated architectural distortion (arrow). *B,* The asymmetry (arrow) is better seen on the magnification view. Subsequent biopsy yielded a grade 2 invasive ductal carcinoma. *C,* Mammogram obtained 2 years earlier, with the output of an artificial intelligence (AI)–based computer-aided diagnosis system (red circle; 93 represents a 93% likelihood of malignancy). Arrow shows same asymmetry seen in *A. D,* Magnification view of the cancer-containing area in *C.* Arrow indicates the same asymmetry seen in *B.* Because of the high likelihood of cancer predicted by the AI system, the lesion is prompted even when missed by the evaluating radiologist. Consequently, this cancer would likely have been detected earlier if an AI system had been used in the original reading. Note that the mammograms were obtained with machines from different vendors and look different mainly due to the applied postprocessing. (Image courtesy of Nico Karssemeijer, PhD.)

referred to as an attention or saliency map (Fig 4). Multiple powerful methods that draw attention to the locations in an image that contribute to the decision for a particular case have recently been proposed for natural images (78–81). Adaptation of these methods to mammography and DBT data will be technically challenging because the data are of much higher dimensionality than are data from typical natural images. However, the benefit that these methods could bring may extend beyond aiding in interpretation. A neural network can learn from millions of images in a few days; this is impossible for radiologists. Therefore, it is conceivable that neural networks may eventually be used as a knowledge discovery tool when their ability to explain predictions improves.

Clinical Applications

Increase the Cancer Detection Rate and Reduce the Recall Rate

The most important task of CAD systems so far has been the detection of cancer on digital mammograms, the quality of which has improved with the implementation of deep learning (69–72,82–87); a few AI systems are now performing at the level of radiologists (Fig 5) (82). An open question is how to use this strong performance to optimize the current diagnostic and screening processes. Recently, reimbursement for the use of CAD in the United States was bundled into the price of a mammogram; thus, it is no longer possible to charge directly for the use of CAD. Hence, CAD should improve the quality and/or efficiency of mammography reading to be profitable. As an initial step, deep learning–based systems may be used for cancer detection in a very similar way to the classic CAD systems, pointing out abnormalities. There are two ways to implement this: using prompts for all findings and, in an interactive setting, showing findings only when specific areas of the mammogram are queried. Although the second approach proved more effective when using conventional CAD (86), this poses ethical problems when using systems with human-like performance because evident cancers detected by the AI system might go unnoticed when the human reader does not click on the right spot. Consequently, hybrid systems have been proposed that function as a decision aid that provides interactive feedback and prompts for the most evident findings (Fig 3). The use of such an AI system increased reader performance significantly (82,85).

Because AI systems are much more specific than previous CAD systems (83,84), they may be used to reduce the recall rate—for example, by identifying specific mammographic features to differentiate recalled benign images from malignant and negative cases (87). Recent advances with the availability of large

determine the class usually occupy a large fraction of natural images, objects of interest in medical images are often relatively small. The standard well-known network architectures were designed for the natural images and do not take these peculiarities into account. Therefore, research is necessary to understand how these architectures can be optimized for medical images—mammograms and DBT in particular. If enough computational capacity was available, this problem could be largely solved by using an automated neural architecture search (78).

Furthermore, to integrate deep learning into clinical practice, it is necessary to explain their predictions in a form understandable to humans. The simplest form of such an explanation points to the input pixels that influenced its predictions. This is often
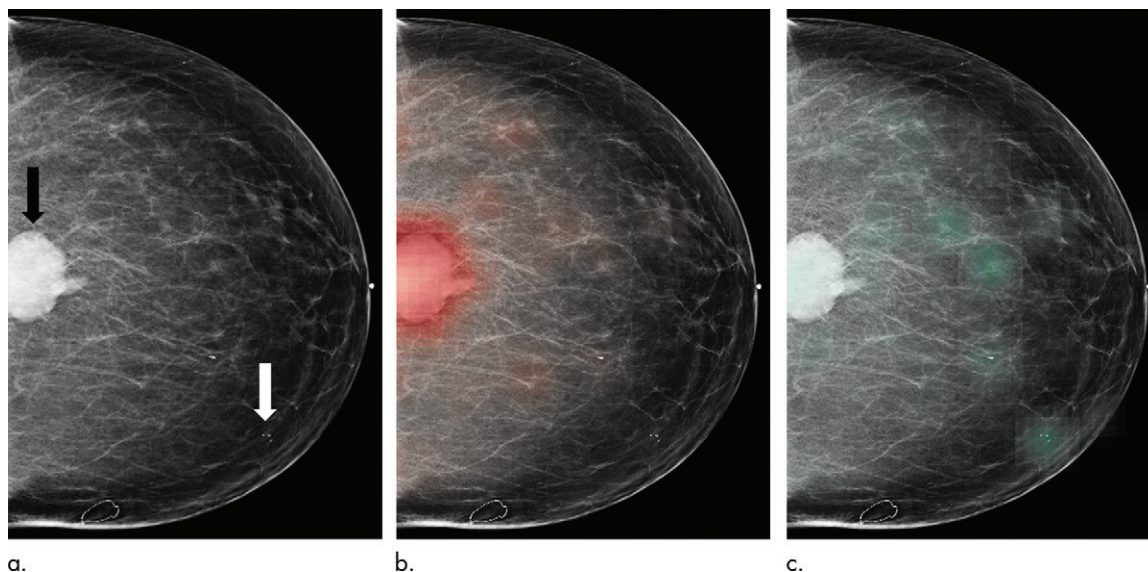
Figure 4: Examples of saliency maps for screening mammography examination classification in 67-year-old asymptomatic woman. Images are left craniocaudal mammograms without **(a)** and with **(b, c)** overlying heat maps. There is a 3.6-cm irregular round dense mass (black arrow in **a**) in upper central left breast and a 5-mm cluster of calcifications (white arrow in **a**) in medial inferior left breast. From a probability of malignancy score of 0–1, the maximum value of the benign green heat map is 0.71. The maximum value of the malignant red heat map is 0.881. Both values indicate that the classifier predicts with high certainly that the mass is malignant and the calcifications are benign. At pathologic examination, the mass was an invasive ductal carcinoma and the calcifications were benign fat necrosis. (Image courtesy of Nan Wu, PhD.)
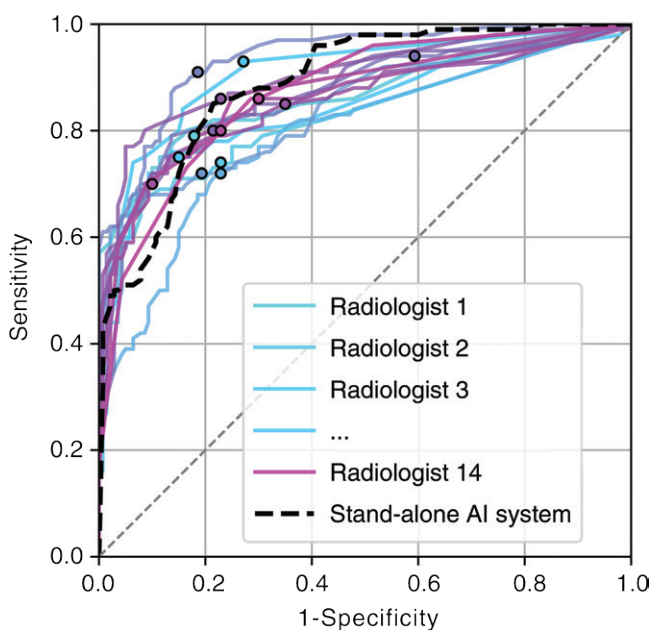

Figure 5: Receiver operating characteristic curves for radiologists reading mammograms unaided and stand-alone artificial intelligence (AI) computer system (Transpara; Screenpoint, Nijmegen, the Netherlands). Circles indicate the radiologists' operating points at Breast Imaging Reporting and Data System category 3 thresholds. (Reprinted, with permission, from reference 82.)

sets of annotated data used to train complex neural networks with many layers have shown a decrease in the number of false-positive prompts and a reduction in recall rates by 10%–20% (Fig 6). Another important feature of the AI systems is the feedback provided when, according to the system, the likelihood of cancer is very low. This may increase the confidence of the reporting radiologist and, hence, expedite the reading of normal cases while allowing more time for potential cancer-containing cases. Rodriguez-Ruiz et al (85) recently found that an AI case score, a metric ranging from 1 to 10 that describes the likelihood of malignancy, was significantly associated with a reduction in reading time. Readers reduced their reading time in cases with a low score, leading to a potential reduction of overall reading time of 4.5% for a screening data set (albeit the general reading times in that study were longer than those normally needed in clinical practice). Alternative approaches that classify mammograms as negative by AI systems alone, without human input, have been proposed. These approaches would have a much larger effect on workflow efficiency.

A human-like AI system could, for example, be used as a fully independent second reader of screening mammograms. A second human reader would only arbitrate discrepancies between the first human reader and the AI system, thus halving the workload for any screening program in which double reading is standard. While tempting, it should be noted that the actual effect of such an approach on recall policy and positive predictive values for recall and biopsy is still unknown. Following earlier studies with conventional CAD systems, a single reading plus CAD approach was not adopted due to a slightly lower sensitivity and higher recall rate (65). The increased performance of the AI system might not completely overcome this, as it is also dependent on the behavior of the human reader who eventually determines whether the finding is suspicious and who will recall the patient for additional imaging. The idea to dismiss mammograms that are categorized as very likely normal without any human reader interpretation is the logical next step. Such preselection of normal cases may be based on case-based AI scores as described earlier but will likely
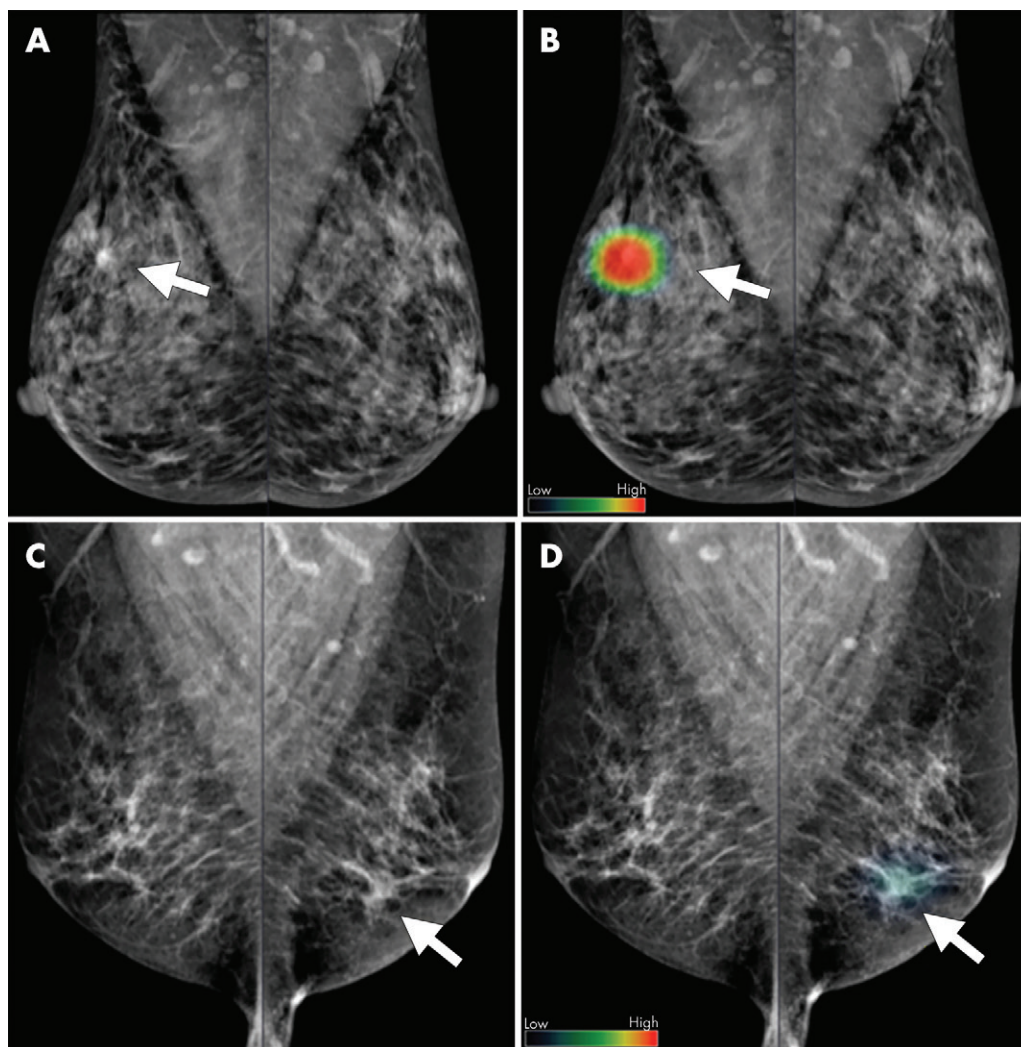
Figure 6: Bilateral mediolateral oblique mammograms in two women with breast cancer (arrow). Mammograms were obtained, *A, C,* without and, *B, D,* with the output of a convolutional neural network–based cancer detection system. The likelihood of cancer presence is given as a heat map. *A, B,* Images in one patient with a relatively obvious spiculated mass found to be a grade 2 invasive ductal breast cancer. *C, D,* Images in another patient with a much more subtle asymmetry found to be an invasive lobular carcinoma. (Images courtesy of Beomseok Suh, PhD.)

examinations. Conventional CAD systems may standardize the reporting of breast density by using either the projected white areas from the processed mammograms directly or a volumetric calculation of the amount of fibroglandular tissue from raw mammograms (94,95). Several studies showed that automated quantitative assessment of breast density is more robust than human evaluations, especially when evaluated over time (96,97).

However, the prediction of risk for the development of breast cancer with use of automated measures seems to be inferior to that of visual assessment (95,98,99). This might have to do with the subconscious incorporation of fibroglandular tissue distribution and texture features by radiologists. Automated risk prediction becomes better when incorporating texture features (100–103), and the integration of texture features with deep learning may strongly improve their discriminatory power (104). Several studies have shown that deep learning–based classification of fibroglandular density categories is closer to radiologist observations than classic feature-based techniques (88,105–107). Moreover, Lehman et al (89) showed that the vast majority (94%) of deep learning–based density classifications is accepted by reporting radiologists. Although the predictive value of AI-based density estimations still must be assessed in modeling studies (108), it is anticipated that the risk assessment with such density estimations, when trained on sufficiently large databases, is similar to that of radiologists.

result in dismissal of a small fraction of cancers by the computer alone. Ethical considerations and cost-effectiveness will determine whether such an approach might be viable in the future. Table 2 lists the differences and potential use of deep learning–based AI systems compared with conventional CAD systems in the detection of cancer. Table 3 provides an overview of these CNN-based AI systems and their current clinical applications.

Quantitative and Reproducible Assessment of Breast Density to Stratify Risk for Breast Cancer

Another important task for CAD systems is to provide an accurate and reproducible assessment of mammographic breast density (88–91). Mammographic breast density may mask an underlying cancer. In addition, dense breast tissue is an independent risk factor for the development of breast cancer (92,93). Consequently, breast density assessment is commonly used for stratification of women for supplemental screening

Applying Deep Learning Algorithms to DBT

The issue of image normalization is even more important in DBT than in mammography. DBT images acquired with machines from different vendors have differences in angular range, acquisition technique, pixel binning, and reconstruction technique (14). Therefore, the difference in the appearance of the normal breast parenchyma with DBT images from different

Table 2: Potential Use of CAD for Cancer Diagnosis

Task	Conventional CAD	AI-based CAD	Advantages of AI
Prevention of overlook errors	Yes, prompts potential abnormalities	Yes, provides prompts or heat map to show most suspicious abnormalities	Fewer false-positive findings
Interactive decision support	Possible, but not commonly used	Possible	Provides feedback on areas not deemed abnormal enough to prompt
Determination of cancer likelihood	No, only lesion-based likelihoods of malignancy are provided	Yes, case-based scores show likelihood of cancer presence	Confirmation of normality may increase confidence and speed of evaluation
Independent second reading	No, CAD findings lead to improved detection but also increased recall	Potentially, as sensitivity is on par with that of radiologists	Strong reduction of workload when double reading is standard, but clinical validation is still required; may increase recall
Dismissal of normal cases (independent first reading)	No, absence of malignancy is unreliable	Potentially, as sensitivity is on par with that of radiologists	May lead to dismissal of some human-detected cancers, potential is dependent on cost-benefit evaluation

Note.— AI = artificial intelligence, CAD = computer-aided diagnosis.

vendors is much larger than that with full-field mammography. This is an important consideration for training deep learning models. In addition, available training data sets for DBT are much smaller, which implies that other techniques that work with a relative paucity of data must be used to improve performance. To manage this issue, transfer learning can be applied. Transfer learning is based on the assumption that if two learning tasks are similar, a network trained to solve a task with more data available can be reused for a task with fewer training data available (109). Most commonly, transfer learning is implemented by copying the parameters of the network trained with a lot of data into the network that is intended to solve the task for which fewer data are available. Subsequently, the second network is only trained for a very short time to prevent overfitting. In the context of breast imaging, this technique was, for example, used for classifying breast density by using a network originally designed for performing Breast Imaging Reporting and Data System classification (88). Current cancer detection systems for DBT are largely based on adaptations of networks originally trained on mammograms to allow the image patterns learned from mammography to be transferred to the analysis of DBT images. However, the depth dimension in tomosynthesis has a poor spatial resolution and therefore only a limited influence on the detection accuracy per anatomic slice (110). It is therefore to be expected that the performance of AI for DBT is somewhat behind the performance for mammography (110).

Current CNN-based systems for DBT already improve upon features that are manually identified and labeled by humans (111,112). With larger training data sets, these algorithms will improve and become indispensable in the evaluation of DBT because the potential gain in workflow efficiencies will be much higher due to the longer reading times of DBT examinations. Proposed detection systems largely work with conventional prompts placed on the synthetic mammogram. These prompts guide the reader to the most suspicious section in the DBT volume when clicked. More advanced integration of AI with DBT is expected, and potential applications are listed in Table 4. These

applications start at the level of scatter correction and denoising to reduce the radiation dose (113,114). Basic reconstruction of a DBT volume is based on (filtered) back-projection, a technique that is commonly used for CT. However, studies have shown that more sophisticated iterative techniques considerably improve image quality (115–117), which also improves subsequent automated cancer detection with CNNs (117). It seems that deep learning–based techniques may further optimize the quality of the reconstructed images (118,119). In the future, synthetic mammograms will be generated from the tomosynthesis data by using deep learning techniques, as current synthetic mammograms may, at best, be comparable to full-field digital mammograms (120,121). The use of machine learning to generate synthetic mammograms may enhance suspicious findings in the DBT volume so that they become more conspicuous. In addition, they may even remove normal tissue that may mask eventual relevant findings. The use of a multiplanar reconstruction fitted through the most suspicious lesions detected by a conventional CAD system in a DBT examination improved reader performance compared to that with full-field digital mammography (122). A commercially available synthetic mammography system on which lesions detected in the DBT volume are enhanced has also been evaluated (Fig 7). In an initial reader study, readers performed equally well with and without CAD, but the average reduction in reading time was 23.5% (123). In addition, James et al (124) found that radiologist performance increased substantially when they compared CAD-enhanced synthetic mammograms with conventional synthetic mammograms.

Radiomics

Radiomics, an expansion of CAD, is defined as the conversion of images to minable data by means of digital decoding of radiologic images into quantitative features (125). In radiomics analysis, the tumor is segmented from its background and various tumor features (eg, intensity, shape, size or volume, and texture patterns) are extracted. Once large high-quality and well-curated data sets are available, they can be used for data mining, which refers to the process of discovering patterns in

Table 3: Summary of Recent Results for Digital Mammography and DBT and AI Applications

Result and Study	Algorithm	Amount of Training	Clinical Results
Deep learning–based AI systems may achieve near-human performance			
Kooi et al, 2017 (84)	CNN-based AI program shows equal performance for mass detection on a patch level	Training set of 39 872 images from 6433 women. Trained on 334 752 positive and 853 800 negative patches. Validation on 4218 images from 710 women. Test set of 18 182 images of 2064 women	Accuracy was 0.93 on test set. Model had equal performance (0.85) to that of radiologists on a subset of patches for mass detection
Geras et al, 2017 (71)	Multiview deep CNN	Approximately 200 000 breast cancer screening examinations (1 000 000 images)	Increasing the amount of training cases to close to 1 million still improved the quality of a CNN-based classifier. Downscaling the images led to a loss of accuracy
Ribli et al, 2018 (69)	Faster *R*-CNN; the base CNN was a VGG16 network (Oxford, England), which is a 16-layer-deep CNN	The public Digital Database for Screening Mammography contains 2620 digitized screen-film screening mammographic examinations; Semmelweis University (Budapest, Hungary) contains 847 FFDM images from 214 examinations; INbreast data set contains 115 FFDM cases	By using the public INbreast database, the model achieved an AUC of 0.95. Also, this approach achieved 2nd place in the Digital Mammography Dialogue for Reverse Engineering Assessments and Methods, or DREAM, Challenge, with an AUC of 0.85
Kim et al, 2018 (83)	CNN-based AI system; deep CNN specialized for images	Digital mammograms from five institutions (4339 cancer cases and 24 768 normal cases) were included	Sensitivity of 75%, specificity of 90%, and accuracy of 90% on a case level
Rodriguez-Ruiz et al, 2019 (82)	Deep learning CNN, feature classifiers, and image analysis algorithms to detect calcifications (20,21) and soft-tissue lesions	Trained, validated, and tested on a private database with more than 9000 cancer cases and 180 000 normal cases	Model achieved noninferior performance compared with 101 radiologists in a series of data sets from various vendors with varying abnormalities, each read by multiple radiologists
Quality and efficiency of radiologist's readings may improve with use of deep learning–based systems for mammography and DBT			
Wu et al, 2017 (88)	Deep learning–based density classification	Approximately 200 000 breast cancer screening examinations (1 000 000 images)	CNN achieved agreement with radiologist's interpretation on the same level as that of radiologists with themselves
Lehman et al, 2018 (89)	Deep learning–based density classification; deep CNN (ResNet-18) with PyTorch (2018, version 0.31; *pytorch.org*)	The model was trained on 41 479 mammograms obtained in 27 684 women and tested on a held-out test set of 8677 mammograms in 5741 women. Then, the algorithm was implemented in clinical practice on 10 763 mammograms	The system provided density assessment in good agreement with radiologist's interpretation, and its score was accepted in 94% of cases
Rodriguez-Ruiz et al, 2019 (85)	CNN-based CAD system	Trained, validated, and tested on a private database with more than 9000 cancer cases and 180 000 normal cases	Radiologists improved detection performance in a reader study of 240 cases (100 malignant) when using the model for reading mammograms (AUC = 0.89 vs 0.87, *P* < .01)

Table 3 (continues)

Table 3 (continued): Summary of Recent Results for Digital Mammography and DBT and AI Applications

Result and Study	Algorithm	Amount of Training	Clinical Results
Deep learning–based AI systems may achieve near-human performance			
Wu et al, 2019 (90)	CNN-based breast cancer screening classifier (code and parameters available at *https://github.com/nyukat/ breast_cancer_classifier*)	Approximately 200 000 breast cancer screening examinations (1 000 000 images)	Model achieved superior performance to 14 breast radiologists in terms of AUC; a hybrid model, averaging probability of malignancy predicted by a radiologist with a prediction of our neural network, was more accurate than either of the two separately both in terms of AUC and precision-recall AUC
Conant et al, 2019 (91)	Deep learning–based system, provides outlines of detected soft tissue and calcific lesions in the DBT sections and calibrated lesion-level and case-level scores	12 000 cases, including 4000 biopsy-proven cancers; reader study	Use of AI for DBT interpretation improved performance (AUC = 0.85 vs 0.80, $P < .01$) while reducing reading time by 52.7%

Note.—AI = artificial intelligence, AUC = area under the receiver operating characteristic curve, CAD = computer-aided diagnosis, CNN = convolutional neural network, DBT = digital breast tomosynthesis, FFDM = full-field digital mammography.

Table 4: AI Solutions for Issues in DBT

Issues with DBT	Potential AI Solution
Radiation	Deep learning–based reconstruction of synthetic images may lead to a reduction in radiation dose
More images	Generation of AI-enhanced synthetic mammograms optimized for a specific task
Calcifications not always clearly visible	Enhance the conspicuity of calcifications
Longer reading times compared with digital mammography	Confirmation of absence of lesions; help localize a finding detected on multiple images
Reduces the effect of tissue superimposition and detects more benign and malignant abnormalities	Improved lesion classification with the three-dimensional information available in the DBT volume
Limited sensitivity in dense breast tissue	Deep learning algorithms may remove normal fibroglandular tissue from composite images to "see through" the dense breast tissue

Note.—AI = artificial intelligence, DBT = digital breast tomosynthesis.

large data sets. This process can use AI, machine learning, or statistical approaches (126). The goal of quantitative radiomics is to yield predictive image-based phenotypes of breast cancer with the aim of better classifying the tumor to improve treatment and prognosis, in line with precision medicine. Furthermore, radiogenomics (ie, imaging genomics) aims to find associations between imaging data and clinical data, molecular data, genomic data, and outcome data (127). Most radiomic studies extract data from breast MRI to determine the cancer phenotype and, in particular, heterogeneity (128,129). However, several studies have shown correlations between mammographic characteristics and biologic profiles of breast cancers (130,131). Consequently, mammographic data may be used to gain insight into breast cancer phenotypes. In a recent study, Shi et al (132)

showed that a CNN detected occult invasion in patients with ductal carcinoma in situ, achieving an area under the receiver operating characteristic curve of 0.70 in a very small database of digital mammograms. This is slightly better than that achieved by Li et al (133), who used more conventional radiomics feature extraction techniques. Another recent study using such feature extraction (134) showed that parenchymal texture features of the contralateral breast may be used to improve the differentiation between benign and malignant lesions. Using a similar approach, Yang et al (135) achieved a classification accuracy of 84% in predicting lymph node involvement from mammographic characteristics of the primary tumor. Another recent study (136) reported that radiomics features of the parenchyma from DBT in women with occult breast cancer in dense breasts
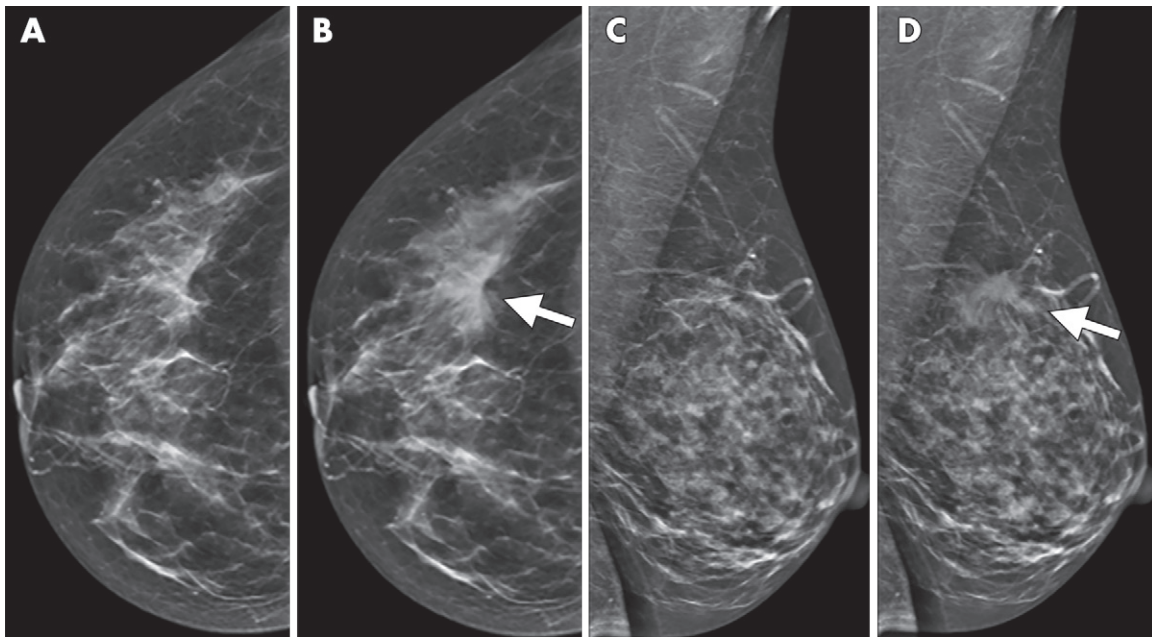
Figure 7: Examples of artificial intelligence (AI)–enhanced synthetic mammograms. *A,* Normal synthetic craniocaudal mammogram of right breast. *B,* AI-enhanced craniocaudal acquisition clearly shows an invasive ductal carcinoma (arrow) that is hardly visible in *A. C,* Normal synthetic mediolateral oblique mammogram and, *D,* AI-enhanced version. The invasive ductal carcinoma in *D* (arrow) is hardly visible in *C.* (Image courtesy of Corinne Balleyguier, PhD.)

differ from those in women without cancer, thus yielding the possibility to predict breast cancer risk estimation. It may be possible to further optimize therapy by using automated extraction of mammographic features of cancer, although it remains to be seen whether these features are complementary to clinical and histopathologic information alone.

Future Directions

A shortcoming of currently used neural network models is that they only evaluate the most recent examination. Although it is possible to make a reasonably good assessment this way, it is evident that this does not take into account all the information a radiologist would rely on to evaluate a difficult examination. Previous mammograms, and images obtained with different imaging modalities, can be exploited to improve the quality of prediction of neural networks. A network that could learn by using these data would be especially useful in diagnosing very early stages of cancer, as even subtle changes in the breast tissue are difficult for a radiologist to perceive. In addition, nonimaging-based patient characteristics, such as demographic information, history of cancer, and genetic information, may be integrated into the model. Given a sufficiently large data set, neural networks could use these pieces of information in conjunction with the image data to identify women at high risk of cancer. Similarly, in patients with breast cancer, AI may allow for highly personalized therapy, commonly referred to as precision medicine, using deep learning–based radiomics assessment. Data on the effect of AI systems on clinical performance and patient outcome are limited. Studies evaluating such end points are vital for the positioning of these techniques in health care, especially because policy-level issues such as reimbursement and liability have yet to be defined.

As AI becomes an important tool for radiologists, it will become fully integrated in the different imaging modalities (137). To be efficient in this role, a deep neural network must be able to explain its decision in a form that is comprehensible to humans. Clinical implementation of AI systems is limited by the machine's current inability to explain its decisions and actions to human users. This must be addressed. Beyond improving the understanding of the predictions made by the neural networks, indicating important parts of the mammogram could be used for planning and analysis of subsequent imaging examinations such as US or MRI.

Conclusions

The development and implementation of artificial intelligence (AI) for mammography has been ongoing for several decades. Because of the advances in deep learning, the speed of implementation and the clinical value of AI have markedly increased. AI will play an important role for mammography and digital breast tomosynthesis (DBT) in all steps—from image generation and denoising to risk prediction, cancer detection, and, ultimately, therapy selection and outcome prediction. Compared with classic computer-aided detection systems based on manually crafted features, deep learning–based systems have a better performance—approaching that of radiologists for specific tasks. Still, there are also residual shortcomings of the novel AI solutions. These include the need for very large and well-curated data sets to train and validate algorithms and a necessity to devise continuous quality control systems as the algorithms are versatile and may evolve over time when more data become available. External validation studies are urgently needed. Although many recent studies are promising and report strong results, we must look at them critically and recognize their limitations in several aspects. First, almost all works only re-

port the area under the receiver operating characteristic curve in detecting malignancy. Although the area under the receiver operating characteristic curve is the most widely applied metric for measuring a classifier's performance, it is sensitive to class distribution. Studies that use test data of different class distributions should not be compared by using the area under the receiver operating characteristic curve. Second, very few studies explain the data distribution used for training and testing in enough detail. Little is known about how accurate these different networks are for different types of findings. We also do not know how well different networks would work when applied to data acquired with different machines or to data acquired for a population of different demographic characteristics. Finally, few studies have been performed to evaluate how the advances in AI can be implemented in a manner that maximizes their clinical impact, which must be the ultimate target. Even with these limitations, it is expected that AI will play a major role in the evaluation of mammography and DBT in the near future, particularly in the screening setting.

References

1. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380(9855): 1778–1786.
2. Tabár L, Yen AM, Wu WY, et al. Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs. Breast J 2015;21(1):13–20.
3. Expert Panel on Breast Imaging, Mainiero MB, Moy L, et al. ACR Appropriateness Criteria Breast Cancer Screening. J Am Coll Radiol 2017;14(11S):S383–S390.
4. Sardanelli F, Aase HS, Álvarez M, et al. Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, the Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey. Eur Radiol 2017;27(7):2737–2743.
5. Expert Panel on Breast Imaging, Moy L, Heller SL, et al. ACR Appropriateness Criteria Palpable Breast Masses. J Am Coll Radiol 2017;14(5S):S203–S224.
6. Sardanelli F, Fallenberg EM, Clauser P, et al. Mammography: an update of the EUSOBI recommendations on information for women. Insights Imaging 2017;8(1):11–18.
7. National Ambulatory Medical Care Survey: 2015 State and National Summary Tables. Available at: https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2015_namcs_web_tables.pdf. Accessed October 1, 2018.
8. Wing P, Langelier MH. Workforce shortages in breast imaging: impact on mammography utilization. AJR Am J Roentgenol 2009;192(2):370–378.
9. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992;184(3):613–617.
10. Weber RJ, van Bommel RM, Louwman MW, et al. Characteristics and prognosis of interval cancers after biennial screen-film or full-field digital screening mammography. Breast Cancer Res Treat 2016;158(3):471–483.
11. Whang JS, Baker SR, Patel R, Luk L, Castro A 3rd. The causes of medical malpractice suits against radiologists in the United States. Radiology 2013;266(2):548–554.
12. Arleo EK, Saleh M, Rosenblatt R. Lessons Learned from Reviewing Breast Imaging Malpractice Cases. J Am Coll Radiol 2016;13(11S):R58–R60.
13. Vedantham S, Karellas A, Vijayaraghavan GR, Kopans DB. Digital Breast Tomosynthesis: State of the Art. Radiology 2015;277(3):663–684.
14. Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys 2013;40(1):014301.
15. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 2013;14(7):583–589.
16. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 2014;311(24):2499–2507.
17. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 2013;267(1):47–56.
18. Tagliafico AS, Calabrese M, Bignotti B, et al. Accuracy and reading time for six strategies using digital breast tomosynthesis in women with mammographically negative dense breasts. Eur Radiol 2017;27(12):5179–5184.
19. 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(7):1954–1965.
20. Winsberg F, Elkin M, Macy J Jr, Bordaz V, Weymouth W. Detection of Radiographic Abnormalities in Mammograms by Means of Optical Scanning and Computer Analysis. Radiology 1967;89(2):211–215.
21. Food and Drug Administration. M1000 ImageChecker. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?ID=319829. Published 1998. Accessed October 1, 2018.
22. Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology 2000;215(2):554–562.
23. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology 2001;220(3):781–786.
24. Birdwell RL, Ikeda DM, O'Shaughnessy KF, Sickles EA. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. Radiology 2001;219(1):192–202.
25. Destounis SV, DiNitto P, Logan-Young W, Bonaccio E, Zuley ML, Willison KM. Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience. Radiology 2004;232(2):578–584.
26. Keen JD, Keen JM, Keen JE. Utilization of Computer-Aided Detection for Digital Screening Mammography in the United States, 2008 to 2016. J Am Coll Radiol 2018;15(1 Pt A):44–48.
27. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med 2007;356(14):1399–1409.
28. Lehman CD, Wellman RD, Buist DS, et al. Diagnostic Accuracy of Digital Screening Mammography with and without Computer-Aided Detection. JAMA Intern Med 2015;175(11):1828–1837.
29. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems 2012;25.
30. Zheng YJ, Sheng WG, Sun XM, Chen SY. Airline Passenger Profiling Based on Fuzzy Deep Machine Learning. IEEE Trans Neural Netw Learn Syst 2017;28(12):2911–2923.
31. Bojarski M, Del Testa D, Dworakowski D, et. al. End to End Learning for Self-Driving Cars. 2016.
32. Sun Y, Wang X, Tang X. Deep Learning Face Representation by Joint Identification-Verification. https://arxiv.org/abs/1406.4773. Published 2014. Accessed October 1, 2018.
33. Chen C, Seff A, Kornhauser A, Xiao J. DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving. 2015 IEEE International Conference on Computer Vision (ICCV), 2015; 2722–2730.
34. Betancur J, Commandeur F, Motlagh M, et al. Deep Learning for Prediction of Obstructive Disease from Fast Myocardial Perfusion SPECT: A Multicenter Study. JACC Cardiovasc Imaging 2018;11(11):1654–1663.
35. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer. JAMA 2017;318(22):2199–2210.
36. Ting DSW, Cheung CY, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images from Multiethnic Populations with Diabetes. JAMA 2017;318(22):2211–2223.
37. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP. Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. Radiology 2018;287(1):313–322.
38. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology 2017;284(2):574–582.

39. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15(11):e1002686.

40. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

41. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18(8):500–510.

42. Choy G, Khalilzadeh O, Michalski M, et al. Current Applications and Future Impact of Machine Learning in Radiology. Radiology 2018;288(2):318–328.

43. Oliver A, Freixenet J, Martí J, et al. A review of automatic mass detection and segmentation in mammographic images. Med Image Anal 2010;14(2):87–110.

44. Bria A, Karssemeijer N, Tortorella F. Learning from unbalanced data: a cascade-based approach for detecting clustered microcalcifications. Med Image Anal 2014;18(2):241–252.

45. Bocchi L, Coppini G, Nori J, Valli G. Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks. Med Eng Phys 2004;26(4):303–312.

46. Regentova E, Zhang L, Zheng J, Veni G. Microcalcification detection based on wavelet domain hidden Markov tree model: study for inclusion to computer aided diagnostic prompting system. Med Phys 2007;34(6):2206–2219.

47. Li L, Zheng Y, Zhang L, Clark RA. False-positive reduction in CAD mass detection using a competitive classification strategy. Med Phys 2001;28(2):250–258.

48. Masotti M, Lanconelli N, Campanini R. Computer-aided mass detection in mammography: false positive reduction via gray-scale invariant ranklet texture features. Med Phys 2009;36(2):311–316.

49. Wu YT, Wei J, Hadjiiski LM, et al. Bilateral analysis based false positive reduction for computer-aided mass detection. Med Phys 2007;34(8):3334–3344.

50. Elter M, Horsch A. CADx of mammographic masses and clustered microcalcifications: a review. Med Phys 2009;36(6):2052–2068.

51. Gao Y, Geras KJ, Lewin AA, Moy L. New Frontiers: An Update on Computer-Aided Diagnosis for Breast Imaging in the Age of Artificial Intelligence. AJR Am J Roentgenol 2019;212(2):300–307.

52. Giger ML. Machine Learning in Medical Imaging. J Am Coll Radiol 2018;15(3 Pt B):512–520.

53. Cole EB, Zhang Z, Marques HS, et al. Assessing the stand-alone sensitivity of computer-aided detection with cancer cases from the Digital Mammographic Imaging Screening Trial. AJR Am J Roentgenol 2012;199(3):W392–W401.

54. Yang SK, Moon WK, Cho N, et al. Screening mammography–detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. Radiology 2007;244(1):104–111.

55. Hupse R, Samulski M, Lobbes M, et al. Standalone computer-aided detection compared to radiologists' performance for the detection of mammographic masses. Eur Radiol 2013;23(1):93–100.

56. Singh SP, Urooj S. An Improved CAD System for Breast Cancer Diagnosis Based on Generalized Pseudo-Zernike Moment and Ada-DEWNN Classifier. J Med Syst 2016;40(4):105.

57. Ikeda DM, Birdwell RL, O'Shaughnessy KF, Sickles EA, Brenner RJ. Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography. Radiology 2004;230(3):811–819.

58. Ciatto S, Del Turco MR, Risso G, et al. Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography. Eur J Radiol 2003;45(2):135–138.

59. Helvie MA, Hadjiiski L, Makariou E, et al. Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection: pilot clinical trial. Radiology 2004;231(1):208–214.

60. Gur D, Stalder JS, Hardesty LA, et al. Computer-aided detection performance in mammographic examination of masses: assessment. Radiology 2004;233(2):418–423.

61. Khoo LA, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. Radiology 2005;237(2):444–449.

62. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. Radiology 2005;236(2):451–457.

63. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: interpretation with computer-aided detection—prospective evaluation. Radiology 2006;239(2):375–383.

64. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. AJR Am J Roentgenol 2006;187(1):20–28.

65. Gilbert FJ, Astley SM, Gillan MG, et al. Single reading with computer-aided detection for screening mammography. N Engl J Med 2008;359(16):1675–1684.

66. Guerriero C, Gillan MG, Cairns J, Wallis MG, Gilbert FJ. Is computer aided detection (CAD) cost effective in screening mammography? A model based on the CADET II study. BMC Health Serv Res 2011;11(1):11.

67. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, Mass: MIT Press, 2016.

68. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–444.

69. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. Sci Rep 2018;8(1):4165.

70. Sage Bionetworks iso. DM Challenge mammography. https://www.synapse.org/#!Synapse:syn9773040/wiki/426908. Accessed October 1, 2018.

71. Geras KJ, Wolfson S, Shen Y, et al. High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks. arXiv:1703.07047v3 2017. https://arxiv.org/abs/1703.07047. Accessed October 1, 2018.

72. Kyono T, Gilbert FJ, van der Schaar M. MAMMO: A Deep Learning Solution for Facilitating Radiologist-Machine Collaboration in Breast Cancer Diagnosis. https://arxiv.org/abs/1811.02661. Published 2018. Accessed October 1, 2018.

73. Zhu W, Lou Q, Vang YS, Xie X. Deep Multi-instance Networks with Sparse Label Assignment for Whole Mammogram Classification. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017. Cham, Switzerland: Springer, 2017; 603–611.

74. Lotter W, Sorensen G, Cox D. A Multi-Scale CNN and Curriculum Learning Strategy for Mammogram Classification. In: Cardoso M, et al, eds. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2017, ML-CDS 2017. Lecture Notes in Computer Science, vol 10553. Cham, Switzerland: Springer, 2017; 169–177.

75. Buelow T, Heese HS, Grewer R, Kutra D, Wiemker R. Inter- and intra-observer variations in the delineation of lesions in mammograms. Medical Imaging 2015: Image Perception, Observer Performance, and Technology Assessment. Bellingham, Wash: International Society for Optics and Photonics, 2019; 941605.

76. Cole EB, Pisano ED, Zeng D, et al. The effects of gray scale image processing on digital mammography interpretation performance. Acad Radiol 2005;12(5):585–595.

77. Gastounioti A, Oustimov A, Keller BM, et al. Breast parenchymal patterns in processed versus raw digital mammograms: A large population study toward assessing differences in quantitative measures across image representations. Med Phys 2016;43(11):5862–5877.

78. Elsken T, Metzen JH, Hutter F. Neural Architecture Search: A Survey. https://arxiv.org/abs/1808.05377. Published 2018. Accessed October 1, 2018.

79. Fong R, Vedaldi A. Interpretable Explanations of Black Boxes by Meaningful Perturbation. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), 2017.

80. Dabkowski P, Gal Y. Real Time Image Saliency for Black Box Classifiers. arXiv:1705.07857. https://arxiv.org/abs/1705.07857. Published 2017. Accessed October 1, 2018.

81. Zolna K, Geras KJ, Cho K. Classifier-agnostic saliency map extraction. https://arxiv.org/abs/1805.08249. Published 2018. Accessed October 1, 2018.

82. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. J Natl Cancer Inst doi: 10.1093/jnci/djy222. Published online March 5, 2019. Accessed October 1, 2018.

83. Kim EK, Kim HE, Han K, et al. Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study. Sci Rep 2018;8(1):2762.

84. Kooi T, Litjens G, van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 2017;35:303–312.

85. Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. Radiology 2019;290(2):305–314.

86. Hupse R, Samulski M, Lobbes MB, et al. Computer-aided detection of masses at mammography: interactive decision support versus prompts. Radiology 2013;266(1):123–129.

87. 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):5902–5909.

88. Wu N, Geras KJ, Shen Y, et al. Breast Density Classification with Deep Convolutional Neural Networks. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018; 6682–6686.

89. Lehman CD, Yala A, Schuster T, et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. Radiology 2019;290(1):52–58.

90. Wu N, Phang J, Park J, et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. https://arxiv.org/abs/1903.08297. Published 2019. Accessed October 1, 2018.

91. Conant E, Toledano A, Periaswamy S, et al. Improving Accuracy and Efficiency with Concurrent Use of Artificial Intelligence for Digital Breast Tomosynthesis

Screening. Radiological Society of North America 2018 Scientific Assembly and Annual Meeting, Chicago, Ill, 2018.

92. Vourtsis A, Berg WA. Breast density implications and supplemental screening. Eur Radiol 2019;29(4):1762–1777.

93. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density. JAMA Oncol 2018;4(9):e180174.

94. Alonzo-Proulx O, Mawdsley GE, Patrie JT, Yaffe MJ, Harvey JA. Reliability of automated breast density measurements. Radiology 2015;275(2):366–376.

95. Astley SM, Harkness EF, Sergeant JC, et al. A comparison of five methods of measuring mammographic density: a case-control study. Breast Cancer Res 2018;20(1):10.

96. Sprague BL, Conant EF, Onega T, et al. Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice: A Multicenter Observational Study. Ann Intern Med 2016;165(7):457–464.

97. Sartor H, Lång K, Rosso A, Borgquist S, Zackrisson S, Timberg P. Measuring mammographic density: comparing a fully automated volumetric assessment versus European radiologists' qualitative classification. Eur Radiol 2016;26(12):4354–4360.

98. Jeffers AM, Sieh W, Lipson JA, et al. Breast Cancer Risk and Mammographic Density Assessed with Semiautomated and Fully Automated Methods and BI-RADS. Radiology 2017;282(2):348–355.

99. Brandt KR, Scott CG, Ma L, et al. Comparison of Clinical and Automated Breast Density Measurements: Implications for Risk Prediction and Supplemental Screening. Radiology 2016;279(3):710–719.

100. Wanders JOP, van Gils CH, Karssemeijer N, et al. The combined effect of mammographic texture and density on breast cancer risk: a cohort study. Breast Cancer Res 2018;20(1):36.

101. Wang C, Brentnall AR, Cuzick J, Harkness EF, Evans DG, Astley S. A novel and fully automated mammographic texture analysis for risk prediction: results from two case-control studies. Breast Cancer Res 2017;19(1):114.

102. Winkel RR, von Euler-Chelpin M, Nielsen M, et al. Mammographic density and structural features can individually and jointly contribute to breast cancer risk assessment in mammography screening: a case-control study. BMC Cancer 2016;16:414.

103. Kontos D, Winham SJ, Oustimov A, et al. Radiomic Phenotypes of Mammographic Parenchymal Complexity: Toward Augmenting Breast Density in Breast Cancer Risk Assessment. Radiology 2019;290(1):41–49.

104. Gastounioti A, Oustimov A, Hsieh MK, Pantalone L, Conant EF, Kontos D. Using Convolutional Neural Networks for Enhanced Capture of Breast Parenchymal Complexity Patterns Associated with Breast Cancer Risk. Acad Radiol 2018;25(8):977–984.

105. 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(1):314–321.

106. Li S, Wei J, Chan HP, et al. Computer-aided assessment of breast density: comparison of supervised deep learning and feature-based statistical learning. Phys Med Biol 2018;63(2):025005.

107. Lee J, Nishikawa RM. Automated mammographic breast density estimation using a fully convolutional network. Med Phys 2018;45(3):1178–1190.

108. Chan HP, Helvie MA. Deep Learning for Mammographic Breast Density Assessment and Beyond. Radiology 2019;290(1):59–60.

109. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? https://arxiv.org/abs/1411.1792. Published 2014. Accessed October 1, 2018.

110. Samala RK, Chan HP, Hadjiiski LM, Helvie MA, Richter C, Cha K. Evolutionary pruning of transfer learned deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis. Phys Med Biol 2018;63(9):095005.

111. Yousefi M, Krzyżak A, Suen CY. Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. Comput Biol Med 2018;96:283–293.

112. Kim DH, Kim ST, Chang JM, Ro YM. Latent feature representation with depth directional long-term recurrent learning for breast masses in digital breast tomosynthesis. Phys Med Biol 2017;62(3):1009–1031.

113. Liu J, Zarshenas A, Qadir A, et al. Radiation dose reduction in digital breast tomosynthesis (DBT) by means of deep-learning-based supervised image processing. In: Angelini ED, Landman BA, eds. Proceedings of SPIE: medical imaging 2018—image processing. Vol 10574. Bellingham, Wash: International Society for Optics and Photonics, 2018; 105740F.

114. Shen H, Liu J, Fu L. Self-learning Monte Carlo with deep neural networks. Phys Rev B Condens Matter Mater Phys 2018;97:205140.

115. Kim YS, Park HS, Lee HH, et al. Comparison study of reconstruction algorithms for prototype digital breast tomosynthesis using various breast phantoms. Radiol Med (Torino) 2016;121(2):81–92.

116. Garrett JW, Li Y, Li K, Chen GH. Reduced anatomical clutter in digital breast tomosynthesis with statistical iterative reconstruction. Med Phys 2018;45(5):2009–2022.

117. Rodriguez-Ruiz A, Teuwen J, Vreemann S, et al. New reconstruction algorithm for digital breast tomosynthesis: better image quality for humans and computers. Acta Radiol 2018;59(9):1051–1059.

118. Cheng L. Accelerated Iterative Image Reconstruction Using a Deep Learning Based Leapfrogging Strategy. https://www.researchgate.net/publication/315702276_Accelerated_Iterative_Image_Reconstruction_Using_a_Deep_Learning_Based_Leapfrogging_Strategy. Accessed October 1, 2018.

119. Ayyagari D, Ramesh N, Yatsenko D, Tasdizen T, Atria C. Image reconstruction using priors from deep learning. In: Angelini ED, Landman BA, eds. Proceedings of SPIE: medical imaging 2018—image processing. Vol 10574. Bellingham, Wash: International Society for Optics and Photonics, 2018; 105740H.

120. Choi JS, Han BK, Ko EY, et al. Comparison between two-dimensional synthetic mammography reconstructed from digital breast tomosynthesis and full-field digital mammography for the detection of T1 breast cancer. Eur Radiol 2016;26(8):2538–2546.

121. Mariscotti G, Durando M, Houssami N, et al. Comparison of synthetic mammography, reconstructed from digital breast tomosynthesis, and digital mammography: evaluation of lesion conspicuity and BI-RADS assessment categories. Breast Cancer Res Treat 2017;166(3):765–773.

122. van Schie G, Mann R, Imhof-Tas M, Karssemeijer N. Generating Synthetic Mammograms from Reconstructed Tomosynthesis Volumes. IEEE Trans Med Imaging 2013;32(12):2322–2331.

123. Balleyguier C, Arfi-Rouche J, Levy L, 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.

124. James JJ, Giannotti E, Chen Y. Evaluation of a computer-aided detection (CAD)–enhanced 2D synthetic mammogram: comparison with standard synthetic 2D mammograms and conventional 2D digital mammography. Clin Radiol 2018;73(10):886–892.

125. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5(1):4006.

126. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278(2):563–577.

127. Pinker K, Chin J, Melsaether AN, Morris EA, Moy L. Precision Medicine and Radiogenomics in Breast Cancer: New Approaches toward Diagnosis and Treatment. Radiology 2018;287(3):732–747.

128. Li H, Zhu Y, Burnside ES, et al. MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays. Radiology 2016;281(2):382–391.

129. Braman NM, Etesami M, Prasanna P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res 2017;19(1):57 [Published correction appears in Breast Cancer Res 2017;19(1):80.] https://doi.org/10.1186/s13058-017-0846-1.

130. Elias SG, Adams A, Wisner DJ, et al. Imaging features of HER2 overexpression in breast cancer: a systematic review and meta-analysis. Cancer Epidemiol Biomarkers Prev 2014;23(8):1464–1483.

131. Woodard GA, Ray KM, Joe BN, Price ER. Qualitative Radiogenomics: Association between Oncotype DX Test Recurrence Score and BI-RADS Mammographic and Breast MR Imaging Features. Radiology 2018;286(1):60–70.

132. Shi B, Grimm LJ, Mazurowski MA, et al. Prediction of Occult Invasive Disease in Ductal Carcinoma in Situ Using Deep Learning Features. J Am Coll Radiol 2018;15(3 Pt B):527–534.

133. Li J, Song Y, Xu S, et al. Predicting underestimation of ductal carcinoma in situ: a comparison between radiomics and conventional approaches. Int J CARS 2019;14(4):709–721.

134. Li H, Mendel KR, Lan L, Sheth D, Giger ML. Digital Mammography in Breast Cancer: Additive Value of Radiomics of Breast Parenchyma. Radiology 2019;291(1):15–20.

135. Yang J, Wang T, Yang L, et al. Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Cancer Using Mammography-Based Radiomics Method. Sci Rep 2019;9(1):4429.

136. Tagliafico AS, Valdora F, Mariscotti G, et al. An exploratory radiomics analysis on digital breast tomosynthesis in women with mammographically negative dense breasts. Breast 2018;40:92–96.

137. Bluemke DA. Editor's Note: Publication of AI Research in Radiology. Radiology 2018;289(3):579–580.