INVITED REVIEW ARTICLE

Current status and prospects of breast cancer imaging-based diagnosis using artificial intelligence

Chikako Sekine[1] · **Jun Horiguchi**[1]

Abstract

Breast imaging has several modalities, each unique in terms of its imaging position, evaluation index, and imaging method. Breast diagnosis is made by combining a large number of past imaging features with the clinical course and histological findings. Artificial intelligence (AI), which extracts the features from image data and evaluates them based on comprehensive analysis, has been making rapid progress in this regard. Many previous studies have demonstrated the usefulness and development potential of AI, such as machine learning and deep learning, in breast imaging. However, despite studies showing the good performance of AI models, their overall utilization remains low, since a large amount of diverse imaging data is required, and prospective verification is necessary to prove its high reproducibility and robustness. Sharing information and collaborating with multiple institutions to collect and verify images of different conditions and backgrounds are vital. If image diagnosis using AI can indeed ensure a more detailed diagnosis, such as breast cancer subtypes or prognosis, it can help develop personalized medicine, which is urgently required. The positive results of AI research, using such image information, can make each modality more valuable than ever. The current review summarized the results of previous studies using AI in each evaluation field and discussed the related future prospects.

Keywords Breast cancer diagnosis · Breast imaging · Artificial intelligence

Introduction

Breast cancer is the most common cancer in women worldwide [1]. To reduce breast cancer mortality, a screening program is essential for identifying early and small breast cancers. Mammography (MG) is the only breast cancer screening modality that has been shown to decrease mortality and is widely adopted across the world [2]. Ultrasound (US) screening has attracted attention recently due to its ability to improve breast cancer detection rate in young Asian women, who often have dense breasts [3]. Magnetic resonance imaging (MRI) has proven useful in populations with high risk of breast cancer [4, 5]. The field of breast imaging has undergone transformative improvements since the inception of MG for screening purposes in the early 1960s. Spatial and temporal resolutions have improved remarkably in all imaging modalities. The most significant change has been in digitalization. Digitalization allowed much more information to be preserved and shared, and facilitated image processing. The images are treated not only as pictures but also as data.

With improvements in molecular biology techniques, diagnosis based on biological and pathological factors would be possible. A correct therapeutic decision is guided by a detailed diagnosis that includes receptor status, proliferation index, nodal status, or immunocyte activity. Although tissue sampling specimens have traditionally contributed to this information, images can play an essential role in whole-tumor evaluation considering the heterogeneity of breast cancer. Imaging technology using artificial intelligence (AI) has developed into a noninvasive method to provide biological and pathological information.

✉ Chikako Sekine
   chikako.s@jikei.ac.jp

1   Department of Breast Surgery, International University
    of Health and Welfare, Narita Hospital, 852 Hatakeda
    Narita, Chiba 286-0124, Japan

Artificial intelligence in breast imaging

AI is a vast field that analyzes data and makes inferences based on knowledge, judgment, problems, and solutions. Technologies used in breast imaging research can be divided into two types, namely machine learning (ML) and deep learning (DL) (Fig. 1). ML is a subset of AI in which computers are trained perform functions without being explicitly programmed by humans to complete those tasks. ML commonly uses features and inputs from human programmers as the basis for learning. With ML, relatively good models can be created, even in small-scale studies with fewer than 1000 target images. Various ML and learning methods have been proposed one after another to create highly accurate prediction models. DL increases the number of layers of ML algorithms to perform more complex and extensive data analyses [6]. Convolutional neural networks (CNN) are the most commonly used DL tools for image-based diagnosis. The CNN approach has resulted in breakthroughs in image processing, including breast imaging, over the last few decades. Therefore, many image research and clinical applications use CNNs to perform clinically meaningful tasks, such as classification, segmentation, and detection.

Breast density and risk prediction

Breast density reflects the amount of fibroglandular tissue in the breasts. It on MG is an independent risk factor for breast cancer [7]. Evaluation of breast density is important to recognize the breast cancer risk and the possibility that noncalcified lesions may be masked in the fibroglandular
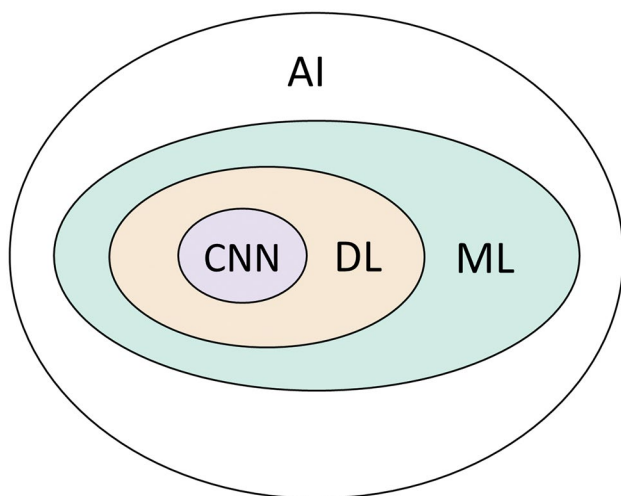
tissue. The European Society of Breast Imaging (EUSOBI) recommends that women be informed about their breast densities and contrast-enhanced breast MRI be conducted in women aged 50–70 years with extremely dense breasts [8]. Although several guidelines exist for evaluating breast density, the most widely used method uses the density score from American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) [9]. The ACR BI-RADS Atlas 5th edition classifies breast density into 4 categories, namely a: "The breasts are almost entirely fatty"; b: "There are scattered areas of fibroglandular density"; c: "The breasts are heterogeneously dense, which may obscure small masses"; d: "The breasts are extremely dense, which lowers the sensitivity of mammography". These 4 categories are associated with breast cancer risk [10]. However, variability remains across radiologists, and consistency is low, even among individual radiologists [11] . Studies using AI have demonstrated beneficial results in terms of both conformity and reproducibility. The DL model demonstrated excellent classification accuracy for dense and non-dense breasts [12]. It exhibited good agreement with the density assessments of an experienced radiologist. Another DL model revealed that breast density estimation can be assessed using any type of mammography, including full-field digital mammography (FFDM), digital breast tomosynthesis (DBT), and synthesized 2D mammograms [13]. Although inter- and intra-reader variability is obvious, visual categorization models are commonly used to estimate the models. Gastounioti et al. stated that this discrepancy is mainly due to the lack of large datasets with ground-truth density estimations [14]. Many AI applications, such as Quantra™ (Hologic, USA), Intel-liMammo® densityai™ (Densitas, Canada), and Volpara TruDensity® (Volpara Imaging, New Zealand), are already commercially available. A study comparing mammographic assessments in the available models showed a strong association between breast cancer risk and automatically measured breast density [15]. Interestingly, visual density assessment demonstrated a strong relationship with cancer, despite the known inter-observer variability.

Some studies have directly predicted the risk of developing breast cancer. An accurate prediction of breast cancer risk is required for personalized screening. Several breast cancer risk prediction models using personal health data, such as age, race, hormone usage, and prior cancer history, have been proposed and investigated in randomized trials [16, 17]. Mirai model [18], a DL-mammography-based risk model, indicated that a DL model with mammographic features added to the previous risk prediction model could increase the accuracy of cancer risk prediction.

Fig. 1 Diagram explaining the relationships in different techniques in the AI field. *AI* artificial intelligence, *ML* machine learning, *DL* deep learning, *CNN* convolutional neural network

Cancer detection

Ideal breast cancer screening should have high sensitivity and specificity, without invasion or cost. Although MG is most widely used around the world, its sensitivity and specificity are insufficient, especially for dense breasts [19]. Several types of US, MRI, and other modalities have been used to detect early-stage breast cancer till date [20, 21]. However, AI algorithms for cancer detection could achieve highly accurate screening and reduce the burden on interpreters. Computer-aided detection (CAD) software for mammography was introduced in the 1990s [22]. The Food and Drug Administration (FDA) approved the first commercial CAD system as a second opinion for screening mammography in 1998. However, CAD increased recall rates [23], and there was no evidence that CAD when applied to digital MG significantly improved screening performance [24]. Instead, many studies have been published on AI algorithms that demonstrated excellent performance in breast cancer detection. McKinney et al. used an AI-based breast cancer detection algorithm, which was trained on larger and more representative datasets in the United Kingdom and the United States, and proved its performance to be better than that of radiologists (AUC: 0.81–0.89) [25]. Other retrospective studies with large European populations confirmed the usefulness of AI-based cancer detection systems [26]. Although many studies have demonstrated the potential of AI in providing highly accurate screening and reducing the workload of MG interpreters, there are several limitations. Most of these studies were retrospective and included small populations. Therefore, prospective randomized controlled trials have had a significant impact in this field.

The Mammography Screening with Artificial Intelligence (MASAI) trial, a randomized controlled trial, revealed that AI-supported MG screening resulted in a similar cancer detection rate compared to standard double reading, with a lower screen-reading workload [27]. A large number of subjects and many years would be required to demonstrate the usefulness of screening image-based diagnoses using AI. However, previous reports had shown positive results when AI was introduced in the medical examination business. In future, AI may be commonly used to reduce the burden on image readers.

Diagnosis and characterization of cancer

When the area or lesion of interest is identified, whether it is benign or malignant is ascertained, and which subtypes of breast cancer are needed. If AI can provide information with high accuracy, unnecessary biopsies can be avoided, treatment preparations can be quickly initiated, and therapy effect monitoring can be promptly performed. Conventionally, the morphological characteristics related to benign or malignant lesions are recognized. For example, a spiculated margin or linear calcification is highly suggestive of a malignancy. These morphological characteristics are supported by verification in a large number of previous cases. Radiomics is a method for analyzing not only the features visible to the human eye, but also those derived from precise calculations (Fig. 2). Radiomics treats images as data and extracts tens to hundreds of types of "radiomic features" (Fig. 3). Radiomic features can be divided into three representative groups, namely morphological, histogram, and texture. In addition, transforming images using techniques, such as wavelet or Fourier transformation, can increase the number of features.
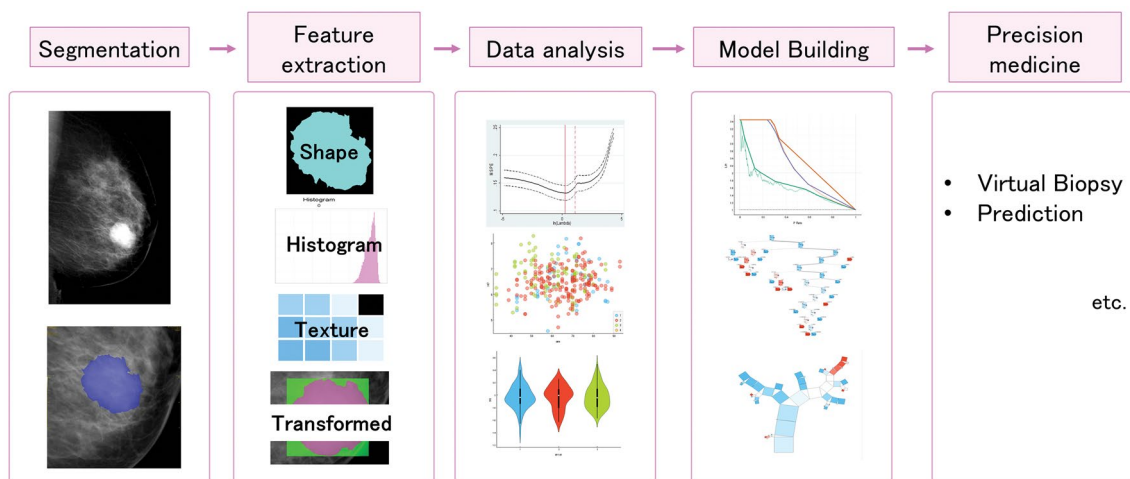

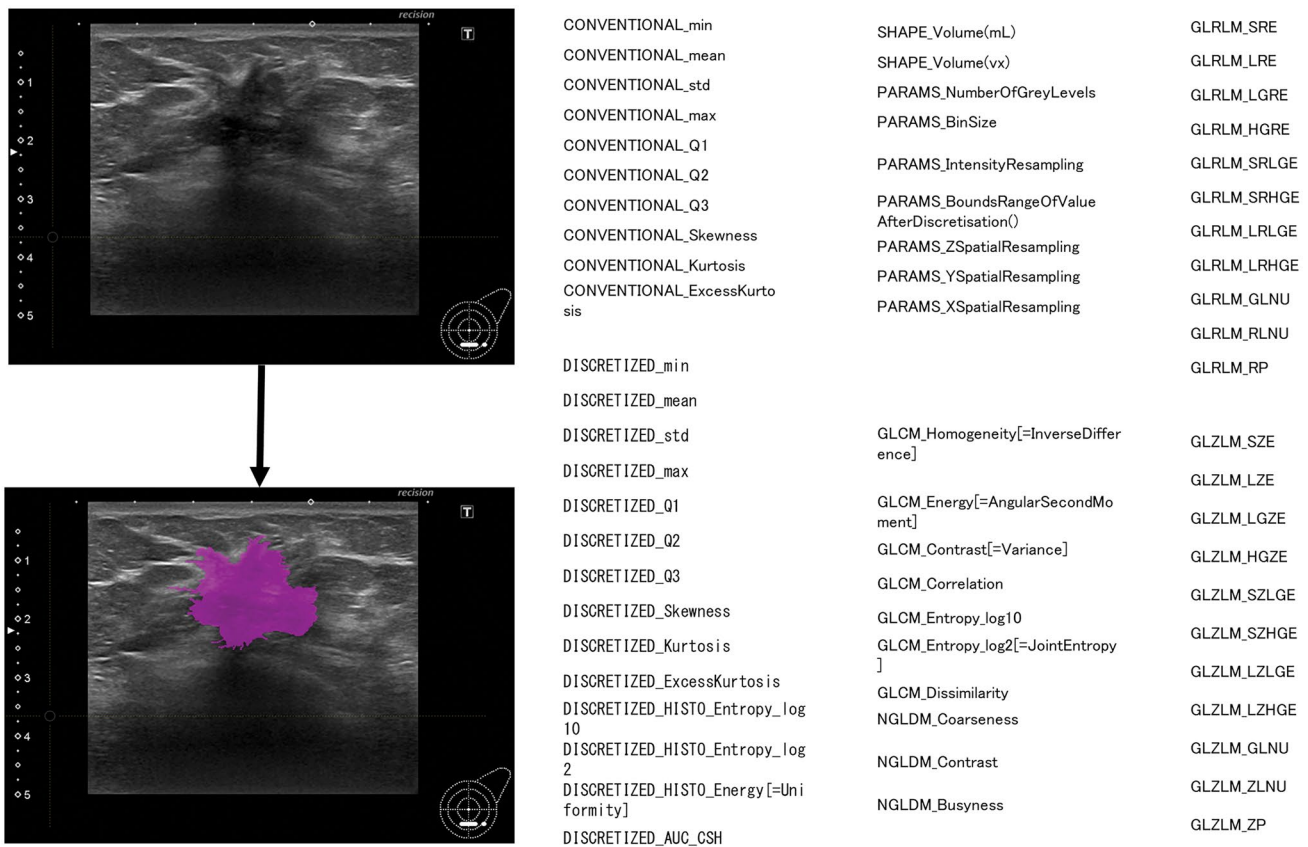**Fig. 2** Radiomics workflow in breast imaging

Fig. 3 Segmentation and examples of radiomics features

To calculate the radiomic features, complex calculations are performed using the pixel values of each pixel in the image. Radiomics comprehensively analyzes the features collected in these ways and uses ML or DL to predict the clinical or histopathological features. In research using radiomics, studies have used US [28] and MRI [29] to determine whether a lesion is benign or malignant. The studies that used ML demonstrated high accuracy. A contrast-enhanced mammography-radiomics model was used to predict the breast cancer characteristics [30]. The accuracies of whether hormone receptor positivity or negativity was 95.6%, and of the tumor grade was 77.8%.

There have also been reports on predicting lesion characteristics using ML or DL without using radiomics. Herent et al. used a DL model that utilizes MRI to identify lesions, determined whether the lesion was benign or malignant, and classified their histological subtypes [31]; the overall Area under the curve (AUC) was 0.817. Fleury et al. used five ML methods to determine whether lesions were benign or malignant using ultrasound images. In this study, support vector machine (SVM) showed the highest AUC of 0.840 [32].

Since new targeted therapies have been developed for specific molecular subtypes, an appropriate therapeutic choice would improve therapeutic response. The breast is an organ that is relatively easily approached for biopsy. Therefore, if images are used to simply determine whether they are benign or malignant, or to classify subtypes of breast cancer, the accuracy would be equivalent to that of a biopsy, and more safety and simplicity can be achieved. However, unlike a biopsy, imaging has the advantage of being able to evaluate an entire lesion. Breast cancer is a clinically and biologically heterogeneous disease [33]; if we can verify the differences in treatment effects and prognosis due to the biological diversity of tumors, the significance of diagnostic imaging will increase dramatically.

Treatment response

Neoadjuvant systemic therapy (NST) can reduce tumor size and allow minimally invasive surgery. If the therapeutic effect can be predicted before starting an NST, it can greatly contribute to drug selection and be useful in precision medicine. Although magnetic resonance imaging (MRI) is the most accurate modality for determining the response to NST, its determination occurs after NST [34]. In recent research, AI has shown the potential to predict NST responses beforehand. In a study using MG, the DL model predicted

pathological complete response (pCR), represented by an area under the curve (AUC) of 0.71 [35]. In a small prospective study using MRI prior to and after two cycles of NST, the high accuracy of several ML models was reported [36]. A radiomics study using pre-treatment ultrasound and digital breast tomosynthesis (DBT) demonstrated that a multimodal algorithm significantly improved the assessment of response to NST than an algorithm using only clinical variables [37]. In addition, a radiomics model using only pre-treatment T2 non-contrast images predicted responder or non-responder status with an AUC of 0.87 [38]. There are several limitations to each study that used AI. However, AI approaches have a significant advantage in predicting the response to NST using images before the therapy, or with images without contrast agents. These studies demonstrated the potential of AI in clinical precision medicine.

Breast cancer prognosis

Following the diagnosis of breast cancer, the most important application of the radiomic AI model would be in the prediction of breast cancer prognosis. Conventionally, some clinical features have been used as predictive factors, such as age, tumor size, axillary lymph node (ALN) metastasis, hormone receptor status, human epidermal receptor 2 (HER2) status, and Ki-67 index. ALN status is critical for predicting disease-free survival and overall survival in patients with breast cancer [39]. Sentinel lymph node biopsy (SLNB) is a widely accepted method that provides an accurate diagnosis of axillary lymph node metastasis and avoids unnecessary axillary lymph node dissection (ALND) [40, 41]. Although SLNB is less invasive than ALND, it is associated with complications. Therefore, AI-based trials that predict ALN metastasis using noninvasive methods have been conducted. The ML model using MG [42] and the DL model using US images [43] and MRI [44] predicted appreciable ALN metastasis. A CNN using MRI divided the patients into three Oncotype DX Recurrence Score groups, with an accuracy of 0.81%. [36] An ML model using MRI predicted disease-specific survival with an AUC of 0.83 [45]. Radiomics showed that addition of radiomics to the conventional radiological prediction workflow improves the prognostic value of breast imaging. A prospective study [46] using radiomics features extracted from pre-treatment US images identified recurrence with an accuracy of 82%. Radiomics features using MRI were associated with disease-free survival in patients with breast cancer [47].

Several approaches have been used, based on clinical characteristics or multigene assays, for prognosis prediction. AI-based prediction models that use breast images acquired in clinical settings may achieve repeatable and cost-effective decisions.

Perspectives on AI in clinical breast cancer imaging

Various AI models exist in breast imaging diagnosis, from those already in commercial use to those still in the research stage. AI models are already being used to assist radiologists in their diagnoses. They are known to be useful, but it will take some time before AIs that can replace radiologists are developed because several imaging studies using AI have been conducted for breast cancer. Most of these were retrospective studies, and large-scale prospective studies would be required in future to evaluate the usefulness and reproducibility of the developed AI. The most important factor in creating a highly accurate AI model is to acquire a large number of high-quality images and associated clinical information. In breast cancer treatment, diagnosis is often performed using a combination of several imaging modalities, such as mammography, ultrasound, and MRI. An AI diagnosis or interpretation of results that combine modalities is likely to be necessary in future. Diagnosing based on a combination of modalities requires more case data than diagnosing based on a single modality. To handle such large amounts of data, a system for sharing data, including image information, would be necessary. Image databases and program codes published online are useful for AI image research; however, the data sources are still insufficient. In addition, each imaging equipment manufacturer has its own platform; however, the environments in which it can be used are limited. More data and knowledge should be shared to implement better AI models in future. By having AI perform simple image-based diagnosis, radiologists may be able to concentrate on diagnosing each patient based on optimal image interpretation.

Conclusion

Research on breast cancer images using AI has been conducted widely, ranging from risk prediction to breast cancer prognosis. The demand for diagnostic imaging is likely to increase because of the associated advantage of observing the entire lesion. Additionally, minimally invasive imaging modalities, such as ultrasound or non-contrast MRI, may attract more attention than ever before. We look forward to AI models with reproducibility and robustness and an environment that makes them easy to use.

Declarations

Conflict of interest All authors have no conflicts of interest to report.

References

1. International Agency for Research on Cancer Cancer tomorrow. 2020. https://gco.iarc.fr/tomorrow/en
2. Lauby-Secretan B, Scoccianti C, Loomis D et al (2015) Breast-cancer screening—Viewpoint of the IARC Working Group. N Engl J Med 372:2353–2358
3. Ohuchi N, Suzuki A, Sobue T et al (2016) Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomised controlled trial. Lancet 387:341–348
4. Saadatmand S, Geuzinge HA, Rutgers EJT et al (2019) MRI versus mammography for breast cancer screening in women with familial risk (FaMRIsc): a multicentre, randomised, controlled trial. Lancet Oncol 20:1136–1147. https://doi.org/10.1016/S1470-2045(19)30275-X
5. Mann RM, Kuhl CK, Moy L (2019) Contrast-enhanced MRI for breast cancer screening. J Magn Reson Imag 50:377–390. https://doi.org/10.1002/jmri.26654
6. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539
7. Boyd NF, Guo H, Martin LJ et al (2007) Mammographic density and the risk and detection of breast cancer. N Engl J Med 356:227–236. https://doi.org/10.1056/NEJMoa062790
8. Mann RM, Athanasiou A, Baltzer PAT et al (2022) Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). Eur Radiol 32:4036–4045. https://doi.org/10.1007/s00330-022-08617-6
9. Spak DA, Plaxco JS, Santiago L et al (2017) BI-RADS®, 5th ed.: summary of changes. Diagn Interv Imag 98:179–190. https://doi.org/10.1016/j.diii.2017.01.001
10. Bodewes FTH, van Asselt AA, Dorrius MD et al (2022) Mammographic breast density and the risk of breast cancer: a systematic review and meta-analysis. Breast 66:62–68. https://doi.org/10.1016/j.breast.2022.09.007
11. Spayne MC, Gard CC, Skelly J et al (2012) Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study. Breast J 18:326–333. https://doi.org/10.1111/j.1524-4741.2012.01250.x
12. Lehman CD, Yala A, Schuster T et al (2019) Mammographic breast density assessment using deep learning: clinical implementation. Radiology 290:52–58. https://doi.org/10.1148/radiol.2018180694
13. Rigaud B, Weaver OO, Dennison JB et al (2022) Deep learning models for automated assessment of breast density using multiple Mammographic image types. Cancers (Basel) 14:5003. https://doi.org/10.3390/cancers14205003
14. Gastounioti A, Desai S, Ahluwalia VS et al (2022) Artificial intelligence in mammographic phenotyping of breast cancer risk: a narrative review. Breast Cancer Res 24:14
15. Astley SM, Harkness EF, Sergeant JC et al (2018) A comparison of five methods of measuring mammographic density: a case-control study. Breast Cancer Res 20:10. https://doi.org/10.1186/s13058-018-0932-z
16. Paci E, Mantellini P, Giorgi Rossi P et al (2013) Tailored breast screening trial (TBST). Epidemiol Prev 37:317–327
17. Esserman LJ, WISDOM Study and Athena Investigators (2017) The WISDOM study: Breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer 3:34
18. Yala A, Lehman C, Schuster T et al (2019) A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 292:60–66
19. Kim WH, Chang JM, Lee J et al (2017) Diagnostic performance of tomosynthesis and breast ultrasonography in women with dense breasts: a prospective comparison study. Breast Cancer Res Treat 162:85–94
20. Comstock CE, Gatsonis C, Newstead GM et al (2020) Comparison of abbreviated breast MRI vs digital breast tomosynthesis for breast cancer detection among women with dense breasts undergoing screening. JAMA 323:746–756. https://doi.org/10.1001/jama.2020.0572
21. Lebron-Zapata L, Jochelson MS (2018) Overview of breast cancer screening and diagnosis. PET Clin 13:301–323. https://doi.org/10.1016/j.cpet.2018.02.001
22. Rao VM, Levin DC, Parker L et al (2010) How widely is computer-aided detection used in screening and diagnostic mammography? J Am Coll Radiol 7:802–805
23. Le EPV, Wang Y, Huang Y et al (2019) Artificial intelligence in breast imaging. Clin Radiol 74:357–366
24. Lehman CD, Wellman RD, Buist DSM et al (2015) Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 175:1828–1837
25. McKinney SM, Sieniek M, Godbole V et al (2020) International evaluation of an AI system for breast cancer screening. Nature 577:89–94
26. Larsen M, Aglen CF, Lee CI et al (2022) Artificial intelligence evaluation of 122969 mammography examinations from a population-based screening program. Radiology 303:502–511
27. Lång K, Josefsson V, Larsson AM et al (2023) Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (Masai): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol 24:936–944
28. Lee SE, Han K, Kwak JY et al (2018) Radiomics of US texture features in differential diagnosis between triple-negative breast cancer and fibroadenoma. Sci Rep 8:13546. https://doi.org/10.1038/s41598-018-31906-4
29. Whitney HM, Li H, Ji Y et al (2020) Harmonization of radiomic features of breast lesions across international DCE-MRI datasets. J Med Imag (Bellingham) 7:012707. https://doi.org/10.1117/1.JMI.7.1.012707
30. Herent P, Schmauch B, Jehanno P et al (2019) Detection and characterization of MRI breast lesions using deep learning. Diagn Interv Imag 100:219–225
31. Fleury E, Marcomini K (2019) Performance of machine learning software to classify breast lesions using BI-RADS radiomic features on ultrasound images. Eur Radiol Exp 3:34
32. Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. Nature 490:61–70
33. Fowler AM, Mankoff DA, Joe BN (2017) Imaging neoadjuvant therapy response in breast cancer. Radiology 285:358–375
34. Skarping I, Larsson M, Förnvik D (2022) Analysis of mammograms using artificial intelligence to predict response to neoadjuvant chemotherapy in breast cancer patients: Proof of concept. Eur Radiol 32:3131–3141
35. Tahmassebi A, Wengert GJ, Helbich TH et al (2019) Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant Chemotherapy and survival outcomes in Breast Cancer patients. Invest Radiol 54:110–117
36. Cai L, Gibbons CS, Nees J et al (2024) Can multi-modal radiomics using pretreatment ultrasound and tomosynthesis predict response to neoadjuvant systemic treatment in breast cancer? Eur Radiol 34:2560–2573. https://doi.org/10.1007/s00330-023-10238-6

37. Kolios C, Sannachi L, Dasgupta A et al (2021) MRI texture features from tumor core and margin in the prediction of response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. Oncotarget 12:1354–1365

38. Kolarik D, Pecha V, Skovajsova M et al (2013) Predicting axillary sentinel node status in patients with primary breast cancer. Neoplasma 60:334–342

39. Lyman GH, Giuliano AE, Somerfield MR et al (2005) American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. J Clin Oncol 23:7703–7720

40. Naik AM, Fey J, Gemignani M, et al. (2004) The risk of axillary relapse after sentinel lymph node biopsy for breast cancer is comparable with that of axillary lymph node dissection: A follow-up study of 4008 procedures. Ann Surg 240:462–8; discussion 468

41. Haraguchi T, Goto Y, Furuya Y et al (2023) Use of machine learning with two-dimensional synthetic mammography for axillary lymph node metastasis prediction in breast cancer: a preliminary study. Transl Cancer Res 12:1232–1240

42. Zhou LQ, Wu XL, Huang SY et al (2020) Lymph Node Metastasis Prediction from Primary Breast Cancer US Images Using Deep Learning. Radiology 294:19–28

43. Chen Y, Wang L, Dong X et al (2023) Deep Learning Radiomics of Preoperative Breast MRI for Prediction of Axillary Lymph Node Metastasis in Breast Cancer. J Digit Imag 36:1323–1331

44. Ha R, Chang P, Mutasa S et al (2019) Convolutional neural network using a breast MRI tumor dataset can predict Oncotype Dx recurrence score. J Magn Reson Imag 49:518–524

45. Dasgupta A, Bhardwaj D, Dicenzo D et al (2021) Radiomics in predicting recurrence for patients with locally advanced breast cancer using quantitative ultrasound. Oncotarget 12:2437–2448

46. Mazurowski MA, Saha A, Harowicz MR et al (2019) Association of distant recurrence-free survival with algorithmically extracted MRI characteristics in breast cancer. J Magn Reson Imag 49:e231–e240