

Review Article



Artificial Intelligence in Breast Cancer Diagnosis and Personalized Medicine

Jong Seok Ahn ^{1,*}, Sangwon Shin ^{1,*}, Su-A Yang ¹, Eun Kyung Park ¹,
Ki Hwan Kim ¹, Soo Ick Cho ¹, Chan-Young Ock ¹, Seokhwi Kim ^{2,3}

¹Lunit Inc., Seoul, Korea

²Department of Pathology, Ajou University School of Medicine, Suwon, Korea

³Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea



Received: Sep 5, 2023

Revised: Sep 25, 2023

Accepted: Oct 6, 2023

Published online: Oct 17, 2023

Correspondence to

Seokhwi Kim

Department of Pathology, Ajou University
School of Medicine, 164 World cup-ro,
Yeongtong-gu, Suwon 16499, Korea.
Email: seokhwikim@ajou.ac.kr

*These authors contributed equally to this work.

© 2023 Korean Breast Cancer Society

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Jong Seok Ahn

<https://orcid.org/0000-0003-1189-5981>

Sangwon Shin

<https://orcid.org/0009-0003-2626-3052>

Su-A Yang

<https://orcid.org/0009-0006-5875-4966>

Eun Kyung Park

<https://orcid.org/0000-0002-5388-9515>

Ki Hwan Kim

<https://orcid.org/0000-0001-7684-235X>

Soo Ick Cho

<https://orcid.org/0000-0003-3414-9869>

Chan-Young Ock

<https://orcid.org/0000-0002-6755-9728>


ABSTRACT

Breast cancer is a significant cause of cancer-related mortality in women worldwide. Early and precise diagnosis is crucial, and clinical outcomes can be markedly enhanced. The rise of artificial intelligence (AI) has ushered in a new era, notably in image analysis, paving the way for major advancements in breast cancer diagnosis and individualized treatment regimens. In the diagnostic workflow for patients with breast cancer, the role of AI encompasses screening, diagnosis, staging, biomarker evaluation, prognostication, and therapeutic response prediction. Although its potential is immense, its complete integration into clinical practice is challenging. Particularly, these challenges include the imperatives for extensive clinical validation, model generalizability, navigating the “black-box” conundrum, and pragmatic considerations of embedding AI into everyday clinical environments. In this review, we comprehensively explored the diverse applications of AI in breast cancer care, underlining its transformative promise and existing impediments. In radiology, we specifically address AI in mammography, tomosynthesis, risk prediction models, and supplementary imaging methods, including magnetic resonance imaging and ultrasound. In pathology, our focus is on AI applications for pathologic diagnosis, evaluation of biomarkers, and predictions related to genetic alterations, treatment response, and prognosis in the context of breast cancer diagnosis and treatment. Our discussion underscores the transformative potential of AI in breast cancer management and emphasizes the importance of focused research to realize the full spectrum of benefits of AI in patient care.

Keywords: Artificial Intelligence; Breast Neoplasms; Diagnostic Imaging; Pathology; Precision Medicine

INTRODUCTION

The incidence of breast cancer has been increasing, with breast cancer surpassing lung cancer as the most commonly diagnosed cancer in women [1,2]. However, the combination of advancements in early detection through screening and the emergence of personalized treatment strategies has led to a decline in breast cancer mortality rates [1,3]. The introduction of innovative treatment options tailored to specific tumor characteristics has significantly improved patient outcomes.

Seokhwi Kim 
<https://orcid.org/0000-0001-7646-5064>

Conflict of Interest

Jong Seok Ahn, Sangwon Shin, Su-A Yang, Eun Kyung Park, Ki Hwan Kim, Soo Ick Cho, and Chan-Young Ock are employed by Lunit Inc. and/or have stock/stock options in Lunit Inc. No other disclosures were reported.

Data Availability

In accordance with the ICMJE data sharing policy, the authors have agreed to make the data available upon request.

Author Contributions

Conceptualization: Ock CY, Kim S; Writing - original draft: Ahn JS, Shin S, Yang SA; Writing, review & editing: Park EK, Kim KH, Cho SI, Ock CY, Kim S.

Despite advances in diagnostic modalities, the current workflow for breast cancer diagnosis is not without errors [4-7]. From the challenges posed by overdiagnosis in low-risk groups to the strenuous workload shouldered by the dwindling number of radiologists and pathologists, the system often finds itself stretched thin. Additionally, inherent discrepancies in test interpretations, limited test accessibility, and high costs further underscore the pressing need for an overhaul [8-12].

The digital era is ushering in transformative changes in the clinical domain, particularly within radiology and pathology. Artificial intelligence (AI), with its potential to extract intricate details from images, automate workflows, and offer predictive insights, presents a promising avenue to address existing limitations [13-16]. Applications of AI are already reshaping practices, from aiding in radiological workflows to predicting long-term disease risks.

In this comprehensive review, we aimed to scrutinize the current diagnostic landscape of breast cancer and highlight its strengths and limitations. We explored the burgeoning role of AI in this field and examined its applications, challenges, and prospects. Through this review, we hope to shed light on the immense potential of AI and the trajectory it sets for breast cancer diagnosis and treatment in the future.

LIMITATIONS OF CURRENT CLINICAL WORKFLOW FOR BREAST CANCER

The diagnostic workflow for breast cancer typically involves several stages, including screening, diagnostic imaging, biopsy, pathologic diagnosis, staging, additional testing (e.g., evaluation of biomarkers), and treatment decision-making (e.g., neoadjuvant chemotherapy [NAC], surgery, adjuvant therapy, hormonal therapy and targeted therapy) [17,18].

Although clinical workflow has improved significantly over the years, it still has several limitations. One of the key challenges for current breast screening programs is the lack of optimal approach for all individuals, as breast cancer risk varies among different individuals. These programs can potentially lead to overdiagnosis in low-risk individuals. The Independent United Kingdom (UK) Panel on Breast Cancer Screening reviewed three randomized controlled trials (RCTs) and found an approximately 19% overdiagnosis risk among UK women [4]. Overdiagnosis can result in unnecessary treatment, emotional distress, and increased healthcare costs, thereby affecting the quality of life of affected individuals.

High-risk individuals can be targeted through current risk stratification models for breast cancer screening, which identify eligible women for supplemental screening or preventive interventions. Guidelines by the American Cancer Society recommend annual breast magnetic resonance imaging (MRI) for women with a lifetime risk of 20% or greater [19]. However, these risk models are primarily based on factors, such as reproductive history, family history of breast cancer, previous benign breast disease, and genetic factors, which may not always be readily available during routine screening workflows. Additionally, some risk models may have limited discriminatory performance, with area under the curve (AUC) values below 0.7, and only a few models incorporate mammographic density, which is an important risk factor. Additionally, women identified to be at high risk for breast cancer based on these models are significantly more likely to be diagnosed with breast cancers that have a favorable prognosis, thereby having less impact on the overall disease burden of the population [20-22].

Another problem is the workload and labor requirements of clinicians. The world is currently witnessing a significant global shortage of radiologists and pathologists, compared to the escalating utilization of radiologic and pathologic tests. From a radiological perspective, there is a growing demand for scans that require more time for interpretation, such as computed tomography (CT) and MRI. However, there is a constant shortage of clinicians available to read these scans. The UK estimates a 40% shortage of radiologists by 2027 [5]. Similarly, Metter et al. [6] reported a decline in the number of active pathologists in the United States (US), resulting in a 41.73% increase in the diagnostic workload per pathologist during the same period.

Moreover, the majority of these tasks are time-consuming and labor-intensive. For example, radiologists may opt to utilize digital breast tomosynthesis (DBT) or additional imaging modalities, such as MRI, to improve cancer detection [23]. Likewise, pathologists may acquire additional slide sections from the tissue blocks. However, these practices can increase the workload and labor time, incur additional costs, and introduce delays in patient reporting [7].

Additionally, there are inherent difficulties and variations in interpreting radiological and pathological tests [8-10]. Significant discrepancies persisted among pathologists and radiologists, with inter-reader agreement ranging between 75%–88%, particularly for specific types of diagnoses [8,11]. Notably, experience level affects diagnostic accuracy; more experienced radiologists demonstrate higher sensitivity in identifying breast abnormalities on mammograms than less experienced radiologists [24]. However, training new radiologists or pathologists is becoming increasingly challenging owing to time constraints imposed by rising workloads and a shortage of experienced mentors to provide supervision. Without a solution to effectively develop new clinicians into proficient and skilled practitioners within a short period, inter-reader variability remains a formidable challenge, impacting the correct interpretation of images, disease management, and potentially delaying diagnosis and treatment decision-making [25].

Lastly, limitations arise as certain tests require specialized facilities and they incur high costs, thereby restricting accessibility for some hospitals and patients to obtain test results. For example, radiological tests, such as contrast-enhanced mammography (MMG) and MRI, in conjunction with gene assays, such as Oncotype Dx (ODX), provide valuable assistance in more accurate diagnosis and treatment decision-making for breast cancer. However, these assays are highly expensive and often necessitate either patient travel to specialized facilities or transportation of specimens to these facilities, leading to time-intensive procedures. Additionally, their tissue-destructive processes can pose challenges in obtaining additional biomarkers or genetic tests [12].

As mentioned previously, multiple limitations require effective and scalable care. Therefore, we believe that AI technology can be instrumental in addressing these limitations.

EMERGENCE OF AI IN RADIOLOGY AND PATHOLOGY

Radiology and pathology specialties are witnessing the introduction of digital workflows and AI, which offer promising prospects for addressing some of the aforementioned limitations. In the era of precision medicine, the demand for more precise and comprehensive diagnostic

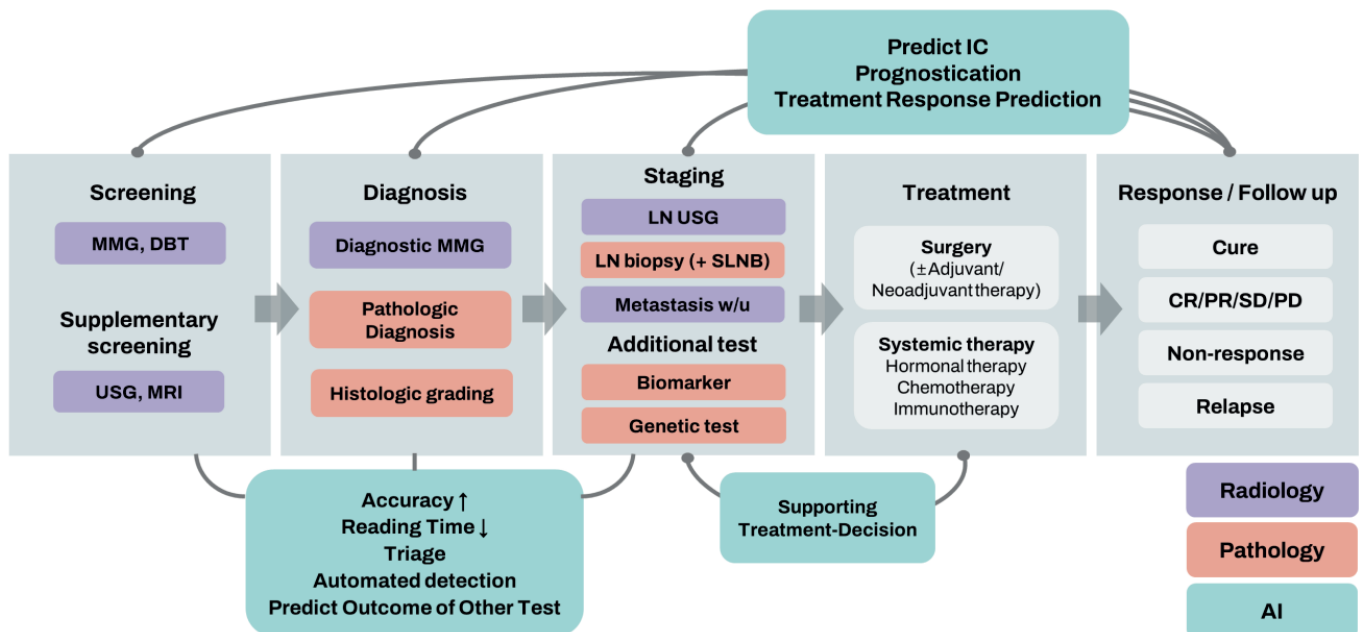


Figure 1. Diagnostic flow chart of breast cancer and application of artificial intelligence.

Potential application of AI in the diagnostic workflow of patients with breast cancer.

IC = interval cancer; MMG = mammography; DBT = digital breast tomosynthesis; USG = ultrasonography; MRI = magnetic resonance imaging; LN = lymph node; SLNB = sentinel lymph node biopsy; W/U = workup; CR = complete remission; PR = partial response; SD = stable disease; PD = progressive disease; AI = artificial intelligence.

tests is increasing. In this regard, the emergence of digital images and Picture Archiving and Communication Systems (PACS) in radiology, as well as whole slide images (WSI) and digital pathology, have brought significant changes to routine practice.

AI systems have advanced rapidly over the last 20 years, transitioning from machine learning (ML) to deep learning (DL), and now to transformer models capable of leveraging multimodal information as inputs. Among the popular DL architectures for imaging analysis, convolutional neural networks (CNNs) are widely used, owing to their ability to extract spatial and contextual information from images through multiple layers and convolutional operations [26]. CNNs trained on extensively labeled datasets can perform tasks, such as segmentation, prediction, and detection, with high accuracy and efficiency [27]. Transfer learning in DL is a valuable method for establishing baseline capabilities for image-related tasks. By leveraging pretrained models on large datasets, transfer learning allows the transfer of learned features and representations to new tasks with limited labeled data, thereby improving performance and reducing the need for extensive training [28]. In practical applications, AI systems support radiologists by managing workflows and detecting suspicious lesions [29]. Certain systems outperform humans in predicting long-term breast cancer risk and prognosticating breast cancers [30]. Similarly, in breast pathology, AI algorithms have been applied to tasks, including cancer detection, classification, histologic grading, lymph node (LN) metastasis detection, biomarker quantification, and genetic abnormality prediction, such as *BRCA* mutation [31]. **Figure 1** summarizes and illustrates the potential areas in which AI can be integrated into the diagnostic workflow for patients with breast cancer.

SCREENING OF BREAST CANCER

Breast imaging and computer aided detection (CAD) in MMG screening

Breast imaging has played a pivotal role in reshaping early cancer detection and reducing mortality in breast cancer since Salomon developed the first mammogram in 1913 [32]. The introduction of imaging modalities, such as MMG, has significantly improved the initial detection of cancer, precise segmentation and characterization of cancerous lesions, as well as monitoring treatment response and post-treatment follow-up [33]. Consequently, two-dimensional (2D) MMG has been widely adopted as the gold standard screening test for women in many countries worldwide, offering cost-effective and scalable approaches [34].

However, standard breast screening using 2D MMG is not without caveats. There is inherent difficulty in detecting cancers using 2D MMG because of overlapping structures in pathological regions. In particular, the difficulty in interpreting mammograms increases with breast density. Dense breasts have a higher proportion of glandular and fibrous tissues, which can mask potential tumors and make them appear similar to normal breast tissue on mammograms. Because breast cancer may be missed or diagnosed at a later stage in women with dense breasts (characterized by $\geq 75\%$ density on MMG), they are at a higher risk of breast cancer [35]. Therefore, attention should be directed towards the development and evaluation of alternative techniques for such cases.

To improve breast cancer detection, several different solutions, such as double reading or arbitration in breast screening, DBT, and the use of CAD systems, have emerged.

Some European countries have implemented double-reading as a strategy for breast cancer screening, which requires two radiologists to reach a consensus on patient recall. In cases of disagreement, arbitration involved a third radiologist, typically with more experience, making the final decision. These methods aimed to enhance cancer detection and reduce recall rates. However, given the previously mentioned shortage of radiologists, existing double reading screening programs are expected to experience major challenges for survival [36].

In regions, such as the US, alternative forms of MMG, such as DBT, are being explored. DBT, also known as three-dimensional MMG, provides a more detailed and clearer view of breast tissues. Thus, the chances of detecting cancerous lesions are increased, while reducing false-positive findings and recall rates compared to 2D MMG [37,38]. Nonetheless, the drawback of the added three-dimensional information is that it requires radiologists to review multiple image slices, leading to an increased reading time [39].

The introduction of digital imaging has significantly increased data availability, prompting the development of CAD algorithms to assist with data interpretation. The first CAD algorithm obtained a US Food and Drug Administration (FDA) clearance in 1998, which showed improved cancer detection in 2D MMG by presenting regions of interest (ROIs) to radiologists, thereby aiding focused reviews [40]. Traditional CAD approaches rely on rule-based methodologies that incorporate domain knowledge to extract handcrafted radiologic features from abnormal tissues, such as pixel values, surrounding pixels, texture, and shape [41]. These features are utilized to make a final decision by comparing them with all detected lesions, thereby enhancing the specificity of lesion detection [42]. Feature engineering plays a crucial role in these systems, with support vector machine classifiers frequently employed [43]. The traditional CAD system demonstrated promising results and was considered

effective, with the potential to replace a human reader in a double-reading screening workflow when used as an aid alongside a single human reader. However, this was at the cost of an incremental increase in recall rates [44].

Many prospective studies that followed did not prove the efficacy of CAD algorithms [45]. A major issue with traditional CAD was its low specificity, which resulted in numerous false-positive findings [46]. For example, it faced challenges in differentiating benign calcifications from those associated with ductal carcinoma *in situ* (DCIS). When used in a screening setting, where cancer incidence is usually approximately 0.5% and the majority of examinations are normal [47], many unnecessary additional reviews and tests had to be conducted, which pushed the biopsy rates up to 19.7% [48], resulting in increased fatigue for the radiologists and a reluctance to use CAD [49].

AI for breast cancer detection: digital MMG and DBT

Despite the limitations of traditional CAD systems, the growing volume of medical scans, shortage of radiologists, and imperative need for early and accurate cancer detection have underscored the need for an improved CAD system. Rapid advancements in AI and DL techniques have provided opportunities for the development of sophisticated CAD systems that can detect subtle signs and features that may not be readily apparent to the human eye.

The development of AI-CAD begins with the collection of a large dataset representing the target population and imaging device. Human readers then collaborate to identify and label lesions in mammograms based on confirmed pathological reports for breast cancer detection [50]. Using these labeled images, AI-CAD self-learns the features that will be used for training, which makes it critically different from the traditional CAD, which only learns human-derived features. To further improve the performance of the algorithm, internal validation was conducted using a dataset separate from the training data to prevent overfitting [51]. The result is an AI-CAD system that can achieve high cancer detection rates while maintaining high specificity and performs significantly better than the traditional CAD [52]. This transformative technology has the potential to enhance accuracy, improve efficiency, and reduce diagnostic variability in breast cancer screening. This can help alleviate the burden on radiologists and facilitate timely and accurate diagnoses.

AI can be integrated into the workflow of 2D breast screening under various scenarios. These include using AI as a standalone system to replace a human reader, and concurrent reading with AI-CAD or AI for triaging normal cases (**Figure 2**). In double-reading screening, AI may assume the role of a second reader or CAD for one or both readers. Alternatively, AI may also act to pre-screen normal cases and reduce the workload for radiologists, or employ a rule-in rule-out approach to remove low-risk cases and refer high-risk cases for another reading by radiologists. When choosing how AI will be implemented into a workflow, factors, such as target sensitivity, specificity, recall rate, and reading workflow in the target country must be considered. An example of the AI output is shown in **Figure 3**.

AI stand-alone performance was assessed to simulate a scenario in which AI entirely replaces a human reader. Numerous studies have demonstrated that AI can perform equivalently or even better than humans [53]. In a systematic review and meta-analysis of 16 studies, standalone AI performed equally well or better than individual radiologists in digital MMG interpretation, based on sensitivity, specificity, and AUC metrics. AI also outperforms radiologists in DBT interpretation, but further evidence is required for a more comprehensive assessment [54].

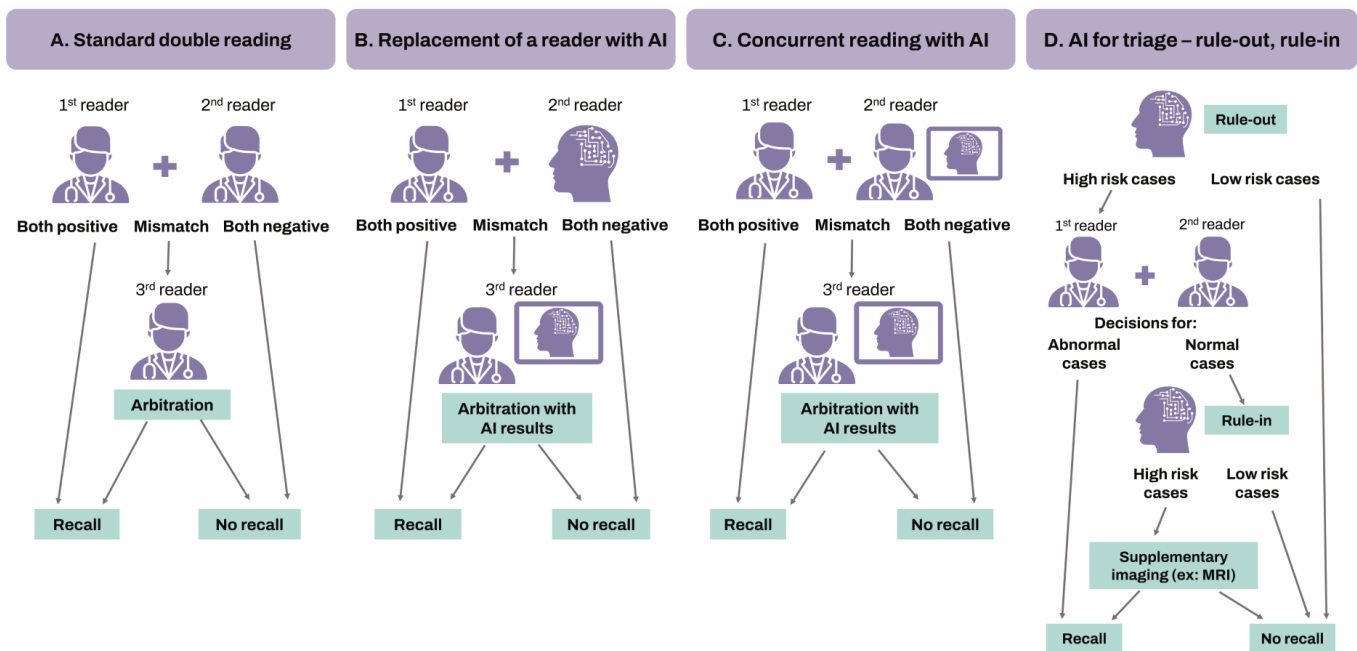


Figure 2. Various workflow scenarios for artificial intelligence usage in two-dimensional breast screening. (A) Standard double reading with arbitration. (B) Standalone AI as a replacement of a second reader. (C) Concurrent reading by the second reader. (D) AI in a rule-out rule-in approach. AI = artificial intelligence.

This underscores the potential of AI in independent mammographic screening, which is particularly important for countries that employ double reading, as replacing a human reader with AI can lead to significant reductions in the required human resources.

The selection of an optimal AI output score, known as the threshold score or operating point, is important for the implementation of AI algorithms for diagnostic decision-making [55]. Although AI algorithms often have a default threshold score, it is important to recognize that different scenarios may require different scores. Factors, such as the specific workflow in which the AI was used or the goals of the screening program should be considered when determining the most suitable algorithm threshold score.

For example, Dembrower et al. [55] compared the sensitivity and workload of standalone AI versus a combination of AI and radiologist. When the sensitivity of the standalone AI was matched with that of a human radiologist, it showed a potential relative sensitivity approximately 5% higher than that for the combined sensitivity of the AI and radiologist, which also matched that of the two radiologists. However, the workload involved in the consensus discussions for the standalone AI scenario was nearly double that of the combined AI reader approach. This suggests that the combined AI-reader scenario and associated AI algorithm threshold may be more suitable for screening programs aimed at reducing the workload, while maintaining similar sensitivity compared to having two readers [55].

Similar to the traditional CAD, concurrent reading occurs when AI is used as a prompt. In a reader study by Kim et al. [56], it has been shown that the diagnostic accuracy of the readers increases with AI, and the incremental range is dependent on the experience level of the readers. When aided by AI, general radiologists (GR) showed improvement in their performance for detecting cancers in 2D MMG (AI unaided GR area under the receiver

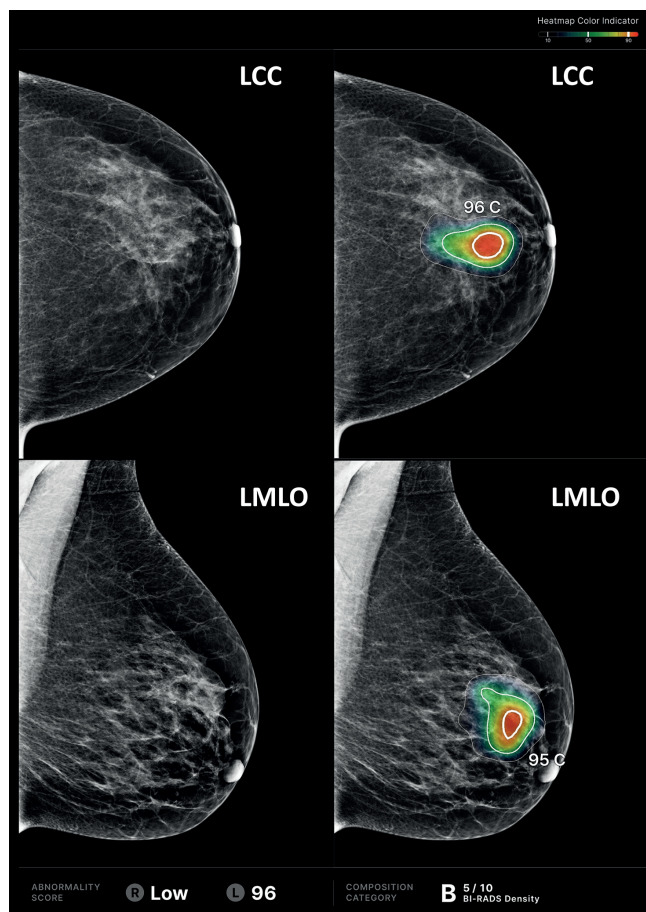


Figure 3. Example of artificial intelligence application in two-dimensional breast mammography. The figures show the original two views (LCC and LMLO), as well as the AI outputs generated using the Lunit INSIGHT MMG (Lunit Inc.). These AI outputs display abnormality scores to indicate a cancerous lesion and heat maps for localization. A density score was provided, according to the BI-RADS category on a scale of 1–10. LCC = left craniocaudal; LMLO = left mediolateral oblique; AI = artificial intelligence; MMG = mammography; BI-RADS = Breast Imaging Reporting and Data System.

operating characteristic curve [AUROC], 0.772; 95% confidence interval [CI], 0.729–0.816 vs. AI aided GR AUROC, 0.869; 95% CI, 0.834–0.903) to a level similar to that of a breast radiologist (BR) (AI unaided BR AUROC, 0.847; 95% CI, 0.809–0.886 vs. AI aided BR AUROC, 0.892; 95% CI, 0.859–0.926) [56]. This is important, as it shows potential for use in countries where BRs are scarce and GRs report mammograms.

In another reader study for DBT, it has also been observed that AI use not only increased the radiologist performance (0.795 without AI to 0.852 with AI) but also reduced the reading time by up to 50% (from 64.1 seconds without AI to 30.4 seconds with AI) [57].

AI triaging is another method for testing AI algorithms. Because the majority of screening mammograms are negative for malignancy, removing even a portion of normal examinations can significantly reduce the workload. Dembrower et al. [29] showed that AI can be set at a threshold at which 60% of the cases can be safely removed from the worklist without risking missing cancer cases. Other studies have reported similar results, with a 47% reduction in workload, resulting in only 7% missed cancers [58]. Additionally, a “rule-in” approach can

be employed, where cases labeled as benign by human readers but assigned a high score by AI are automatically recalled for further testing. This workflow, combined with the “rule-out” approach, can significantly reduce the workload, while increasing the detection of subsequent interval cancers (ICs) and next-round detected cancers [29].

Retrospective studies utilize existing data representing target populations and allow various simulations to test AI algorithms. The radiologists’ decisions and histopathological data were necessary for comparison. It is common practice to establish the ground truth based on at least two consecutive screening episodes to detect screen-detected cancers, ICs, and next-round detected cancers. Promising results have been achieved; however, most retrospective studies are limited to the validation of AI algorithm performance in an enriched cohort or multiple-reader multiple-case analysis [59].

A recent topic of interest in AI cancer-detection algorithms is improving the detection of ICs. ICs are often aggressive forms of cancer associated with higher mortality rates, and the risk of death from IC is 3.5 times higher than that of non-ICs [60]. Despite previous efforts, IC accounts for approximately 30% of detected breast cancers, and its attempts to improve IC detection have been unsuccessful [61]. However, AI algorithms have shown promise in detecting ICs. Hickman et al. [62] demonstrated that a standalone AI can detect 23.7% of ICs, even when set at a 96% threshold, thereby potentially allowing for a significant increase in IC detection.

With the abundance of available retrospective evidence, ongoing efforts are being made worldwide to conduct prospective clinical trials. Results of several prospective trials investigating the use of AI in 2D breast screening are emerging. For example, the ScreenTrustCAD study conducted in Sweden examined the impact of replacing one reader in a double-reading setting. The results were highly positive, showing that in a prospective interventional study based on a large population, a single reader with AI can achieve a superior cancer detection rate, while maintaining the recall rate compared with traditional double readers [63]. In this scenario, the effects of AI on arbitration can only be prospectively evaluated. In another RCT conducted in Sweden called the Mammography Screening with Artificial Intelligence trial, the clinical safety of using AI as a detection support in MMG screening was investigated [64]. In an intervention group, examinations were first classified by AI into high- and low-risk groups, which were then double- or single-read, respectively, by radiologists with AI support. Interim analysis results demonstrated that AI-supported screening showed not only comparable cancer detection rates to a control group’s standard double reading, but also a significantly reduced screen-reading workload. This RCT showed that employing AI in MMG screening could be a safe and effective alternative to standard double reading in Europe. The trial will continue for two more years to assess the primary endpoint of the IC rate [64]. Other studies, such as the AI-STREAM in South Korea, are also actively investigating the effects of AI in single-reader concurrent reading settings [65].

Prospective trials are indeed essential, as they provide valuable insights into the performance of AI algorithms in real clinical settings and capture the challenges that may arise in these environments. A pitfall of retrospective trials is that cancer-enriched datasets that do not reflect the real-life prevalence of cancer are often used. Therefore, AI performance from these skewed studies may not necessarily be replicated in prospective studies or real life [66]. Prospective trials, on the other hand, allow the evaluation of AI algorithms in out-of-distribution scenarios, providing a more realistic assessment of their performance. However,

the disadvantage of prospective studies is their high cost and lengthy time frame, which makes it difficult to conduct them frequently.

A potential solution for addressing the challenges of conducting prospective trials for every use case and geographical area is to leverage large-scale retrospective studies using extensive datasets. By collecting a sufficient sample size and incorporating data from multiple centers, these retrospective studies can better account for the variability encountered in real-life scenarios. National initiatives, such as the Swedish Validation of Artificial Intelligence for Breast Imaging project, exemplify this approach by establishing comprehensive multicenter databases for external validation, allowing independent and simulated testing of AI algorithms [67]. By combining the insights gained from prospective and retrospective trials, it is possible to ensure the cost-effectiveness, scalability, and safe adoption of AI in breast screening, which benefit both patients and healthcare systems.

AI in supplemental breast cancer screening with MRI/ultrasound

Supplemental imaging techniques, such as DBT, MRI, handheld ultrasound, and automated breast ultrasound (ABUS), are commonly used as additional tools to traditional MMG to enhance cancer detection in women with dense breasts. Attempts have also been made to apply AI to these modalities to enhance their performance.

For example, Shen et al. [68] demonstrated that implementing an AI system was beneficial to the radiologists' diagnostic process for identifying breast cancer using ultrasound. The use of AI led to a significant reduction in false-positive rates by 37.3% and biopsy requests by 27.8%, while maintaining sensitivity. Moreover, a standalone AI system outperformed an average of ten board-certified BRs, with an AUROC improvement of 0.038 (95% CI, 0.028–0.052; $p < 0.001$) [68]. This suggests that the AI system can not only assist radiologists in improving the accuracy, consistency, and efficiency of breast ultrasound diagnosis but also performs better than human experts [68].

MRI enhancement using AI algorithms focuses on improving the acquisition time, which is a critical problem in this modality. 'Fast MRI challenge' is a research initiative aimed to develop and evaluate MRI techniques using AI to expedite MRI image acquisition without compromising image quality. The results from this challenge have shown that AI can successfully reconstruct missing data in accelerated magnetic resonance images, while maintaining acceptable data quality for radiologists [69].

Finally, as CAD systems, AI algorithms have demonstrated usefulness in conjunction with supplemental imaging techniques. CAD-ABUS helps radiologists achieve a significant reduction in reading time, while maintaining accuracy in detecting suspicious lesions [70]. Additionally, in the case of MRI, DL-based CAD systems have shown a significantly higher average sensitivity in early phase scans where abbreviated MRI protocols are utilized [71]. This highlights the potential of AI in playing an increasingly important role in the future, particularly in the interpretation of supplemental images.

DIAGNOSIS OF BREAST CANCER

A confirmatory diagnosis of breast cancer can only be established through the histopathological examination of breast specimens, which are commonly obtained via core

needle biopsy. Pathologic reports typically provide information not only on the presence of tumor cells but also on whether the tumor is an invasive carcinoma or carcinoma in situ, histological subtype (ductal vs. lobular), histologic grade (such as the Bloom-Richardson grade [BRG]), as well as the status of biomarker expression and other relevant findings. The clinical management and prognosis of the disease may vary, depending on the specific pathologic outcomes identified in the report [17,72].

Given the fundamental role of pathologic diagnosis in disease management, the accuracy of pathologist's diagnoses is important, and discrepancies in pathologic diagnoses can have a profound impact on diagnostic accuracy and treatment decision-making processes [73]. Nevertheless, inter-observer variation was noted in breast biopsy specimens, resulting in an overall concordance rate of 75.3%. The lowest concordance rate of 48% was reported for DCIS specimens and atypical hyperplasia [8].

AI for breast cancer diagnosis

Several researches have focused on developing AI algorithms for breast cancer detection and classification [74-77]. Cruz-Roa et al. [74] developed a CNN model that can classify patches containing invasive ductal carcinoma from the WSI of breast cancer and estimated the degree of infiltration and extent of invasive foci from the WSI using the ConvNet classifier. Han et al. [75] reported a DL model with an average accuracy of 93.2% across eight classes (four benign and four malignant) in a test dataset.

Furthermore, an image analysis challenge called Breast Cancer Histology (BACH) challenge aimed to automate breast tissue histology classification from hematoxylin and eosin (H&E)-stained microscopic images and WSIs. The best-performing model reached pathologist-level accuracy, with AI assistance increasing average accuracy from 0.80 to 0.88 and improving mean interobserver concordance from 0.83 to 0.90 [76].

One example of a commercially available AI algorithm is the GALEN Breast. This AI model utilized breast biopsy specimens to detect cancer cells and classified them into different breast cancer subtypes. The AUC values of this model based on a large-scale, multi-institutional dataset were reported as 0.99 and 0.98 for the detection of invasive carcinoma and DCIS, respectively [77].

AI for histologic grading in breast cancer

In addition to extensive efforts in cancer classification, various attempts have been made toward histologic grading. The BRG is widely recognized as a prognostic factor in breast cancer, and is derived from the assessment of three morphological components: tubule formation, nuclear pleomorphism, and mitotic count [72].

Romo-Bucheli et al. [78] developed a DL model capable of detecting tubule nuclei in WSI obtained from patients with estrogen receptor (ER)-positive breast cancer by introducing a tubule formation indicator (TFI) based on tubule nuclei to total nuclei ratio. Notably, the TFI correlated with the corresponding risk categories from ODX.

Similarly, Whitney et al. [12] showed that an ML classifier utilizing quantitative features related to nuclear shape, texture, and architecture could predict ODX risk categories for early stage ER-positive breast cancer patients (accuracy between 75%–86%). In another study conducted on H&E tissue microarray cohorts of early stage LN-negative, ER-positive breast

cancer patients, a model using nuclear shape and orientation stratified short- and long-term survival outcomes, indicating high-risk group outcomes independent of T stage, histologic grade, and nuclear grade [79].

Indeed, the majority of studies have focused on predicting mitotic counts. Similar to the approaches used for tubule formation and nuclear pleomorphism, a study demonstrated a DL model capable of counting mitotic figures from H&E-stained slides of early stage ER-positive breast tumors. This model achieved an accuracy of 83.19% in distinguishing ODX risk groups [80]. Furthermore, several mitosis detection contests have been conducted, including the 2012 International Conference on Pattern Recognition (ICPR) MITOSIS detection contest, 2014 ICPR MITOS-ATYPIA challenge, and mitosis detection task in the 2016 Tumor Proliferation Assessment Challenge 2016 contest [81-83]. These challenges have stimulated the development of diverse AI models that have shown noteworthy performance when applied to these datasets and have unveiled the potential for precise mitosis counting using AI models [84-86]. Additionally, a specific DL algorithm utilizing Faster R-CNN with a ResNet-101 backbone network exhibited not only high accuracy in mitosis counting tasks but also the potential to decrease pathologists' reading time by 27.8% [87]. Recently, Nateghi et al. [88] introduced a fully automated system that encompasses ROI identification, mitosis counting from WSI, and prediction of tumor proliferation scores. When evaluated using the TUPAC-16 dataset, this system outperformed all previous methods in these tasks [88].

Moreover, a model allows for more refined stratification within the breast recurrence group (Nottingham histological grade [NHG] 2), a category that encompasses approximately half of all patients with breast cancer. This model divides patients with NHG 2 into two distinct groups based on their recurrence-free survival rates, thereby revealing a subset with increased recurrence risk and morphological characteristics akin to the NHG 3 group [89].

AI for preoperative breast cancer evaluation

AI may also play a significant role in preoperative breast cancer evaluation, including the prediction of LN metastasis. Zhou et al. [90] demonstrated that LN metastasis can be predicted with an AUROC accuracy of 0.90 based on primary breast cancer ultrasound images. Additionally, AI has been utilized to predict occult invasive diseases and improve the prediction of DCIS upstaging, thereby aiding decision-making and appropriate treatment planning [91]. Other preoperative uses of AI include the prediction of molecular subtypes based on radiomic features. This prediction is important because of the heterogeneous nature of breast cancer, where single biopsy specimens may not accurately represent the disease, especially when immunohistochemistry (IHC) test results are uncertain [92]. Radiomic features can provide a better overall representation of cancer, enabling more accurate classification and treatment. For example, previous studies have shown that combining MMG and ultrasound images to extract features has high accuracy in discriminating luminal and non-luminal diseases [93].

STAGING OF BREAST CANCER

As in the case of managing most solid tumors, an assessment based on the eighth edition of the American Joint Committee on Cancer TNM staging system is of paramount importance. This staging system serves as a vital determinant of treatment decisions and prognosis because the disease can vary significantly at different stages [17,94]. Sentinel LNs (SLNs) evaluation plays a critical role in breast cancer clinical staging and subsequent treatment planning [95]. SLN

biopsy (SLNB) is a widely used procedure in breast cancer staging for assessing whether cancer cells have spread to regional LNs, particularly axillary LNs. However, the false-negative rates of SLNB are almost 10% [96], indicating that additional imaging modalities, such as ultrasound, MRI, or positron emission tomography/CT, need to be considered to improve the accuracy of nodal staging and identify cases where SLNB alone may be insufficient.

AI for LN metastasis detection in histologic slides

AI applications for the evaluation of pathologic LN metastases have largely concentrated on SLNs in breast cancers [97]. Various models have been developed to detect LN metastases in cytokeratin IHC-stained WSIs, with one model achieving a sensitivity of 100% and specificity of 68.9%, indicating no false negatives [98-100]. Despite being more time-consuming and complex, IHC-based methods are particularly beneficial in certain scenarios, such as in patients who have undergone NAC, where the nodal tissue might exhibit drug-induced changes or an inflammatory/fibrotic response.

The CAMELYON16 challenge has spurred significant advancements in the automated detection of LN metastases from H&E-stained slides. An AI with a top-performing model that achieved an AUC of 0.994 surpassed the AUC values of pathologists under time constraints [101]. Another notable model demonstrated state-of-the-art performance and significantly reduced false-negative results for predicting the pathologic nodal status on the CAMELYON17 dataset by combining a patch-level CNN-based metastasis detection model and slide-level LN classifier [102]. Google's AI model, known as LYNA, improves pathologists' interpretations and sensitivity, particularly for SLN micrometastasis detection, while also reducing the average reading time [103,104].

However, only a few studies have explored AI implementation for intraoperative evaluation assistance with frozen sections because evaluation is challenging, owing to low-quality slides and time constraints [97]. One study demonstrated the possibility of developing an AI model for frozen sections using formalin-fixed paraffin-embedded datasets from CAMELYON 16 via transfer learning [105]. In the HeLP 2018 challenge in H&E-stained frozen tissue sections of SLNs of breast cancer patients, the best-performing algorithms achieved an AUC of 0.805 and a processing time of 10.8 min, but AI model accuracy was similarly reduced by factors, such as micrometastasis, neoadjuvant therapy, and invasive lobular carcinoma [106]. Nevertheless, the potential of AI to overcome these challenges is evident in the field of gastric cancer. For example, one study revealed the possibility of confidently using an AI model for LN screening after neoadjuvant therapy with a positive predictive value of 94.44% [107], whereas another study showed increased sensitivity for micrometastases and isolated tumor cells, along with a shorter review time [108].

PREDICTIVE OR PROGNOSTIC FACTORS FOR BREAST CANCER

Biomarkers in breast cancer

Tissue biomarkers have gained importance in personalized medicine, aiding in disease diagnosis, prognosis prediction, and selection of patients who may derive specific therapeutic benefits. Breast cancer management involves key tissue biomarkers, including ER, progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2), and Ki-67. Ongoing research has investigated novel biomarkers, such as programmed death ligand-1

(PD-L1) and tumor-infiltrating lymphocytes (TIL). Despite the importance of biomarker assessment, several studies have demonstrated intra- and inter-laboratory variability in the assessment of ER, PgR, HER2, and Ki-67. This could influence treatment decisions regarding hormonal and anti-HER2 targeted therapies [9].

AI in biomarkers of breast cancer

Assessing hormone receptor (HR) status via IHC can help identify patients who are likely to benefit from endocrine therapies, such as tamoxifen. Samples with at least 1% ER- or PgR-positive tumor nuclei were deemed positive, with quantification achievable by reporting the percentage of positive cells or utilizing scoring systems, such as the Allred or H-score [109,110].

Since then, several studies have explored automated quantitative digital imaging analysis (DIA) techniques. Although manual interpretation of IHC staining is inherently subjective and time-consuming, studies have shown a strong correlation between manual and DIA-based scoring of ER and PgR IHC staining in breast cancer. Notably, the utilization of DIA has demonstrated improved reproducibility compared with manual scoring methods [111-117]. Moreover, efforts have been made to integrate DIA algorithms into routine digital pathology workflows and ensure the robustness of AI models, and promising results have been reported [117,118]. Additionally, some AI models are promising in predicting ER and PgR status using only H&E-stained slides, thereby eliminating the need for specific IHC staining [119-121]. Notably, a DL model based on ShuffleNet was developed to predict molecular alterations and biomarkers in various solid tumor types, including breast cancer [122].

HER2 status, determined by IHC with or without *in situ* hybridization (ISH), is essential for identifying candidates for anti-HER2 therapies, such as trastuzumab. Levels of HER2 are classified based on the proportion and intensity of stained tumor cells [123]. In an effort to quantify HER2 IHC slides, a study reported an overall agreement of 92.3% between software-based analysis and pathologist assessment by evaluating cell membrane connectivity [124]. Another study demonstrated the potential of an AI-powered HER2 analyzer to mitigate interobserver variability and aid pathologists in achieving a consistent evaluation of HER2 expression levels [125]. Furthermore, several studies have used AI models to predict the amplification state of HER2 by analyzing digitized fluorescence ISH (FISH) images [126-128].

Subsequently, several AI models have been developed to exclusively predict the HER2 status using H&E-stained slides, including the HEROCHE challenge [119,121,122,129,130]. In this challenge, 21 international teams presented their AI models, and the best-performing model exhibited a classification accuracy of 0.68 in terms of F1 score. It is important to note that the choice of evaluation metric may influence the performance of the models. Beyond simply predicting the HER2 status, some studies have shown associations between AI models and therapeutic responses. Farahmand et al. [131] developed a CNN classifier, achieving an AUC of 0.80 in predicting the response to trastuzumab therapy based on HER2 status. Another intriguing application of AI was observed in patients who achieved a pathologic complete response (pCR) after NAC with anti-HER2 agents, revealing a notably higher proportion of tumor cells with intense HER2 staining. This insight suggests that AI models may be pivotal in providing nuanced information for predicting responses in patients with HER2-positive early breast cancer undergoing NAC [132].

Furthermore, quantitative assessment of HER2 IHC using AI algorithms is not limited to breast cancer. It has demonstrated promise in reducing inter-observer variability and

forecasting prognosis or treatment outcomes in other cancer types, including urothelial carcinoma, biliary tract cancer, and colorectal cancer [133-135].

Despite the consistent treatment benefit of cyclin-dependent kinase 4 and 6 inhibitors demonstrated in a recent phase III clinical trial regardless of the Ki-67 index, Ki-67 serves as a cell proliferation marker and prognostic and predictive biomarker for breast cancer [136,137]. However, the reproducibility of Ki-67 assessment remains a longstanding challenge [9,138,139].

Similar to other biomarker evaluations, Ki-67 is also evaluated using IHC, and several DIA platforms showed promising results. A comparative study revealed excellent reproducibility among the three DIA platforms and reference standards, with the platforms demonstrating indistinguishable capabilities for predicting patient outcomes in breast cancer [140]. Furthermore, another study revealed that incorporating AI support in the evaluation of Ki-67, ER, and PgR expression led to a slight improvement in pathologist agreement, with 95.8% of the AI analysis results for Ki-67 confirmed by most of the pathologists [117].

Recently, immunotherapy combined with chemotherapy has demonstrated efficacy in specific breast cancer subsets, necessitating the use of predictive biomarkers like PD-L1. Validation of PD-L1 IHC expression was provided by the KEYNOTE-355 trial, revealing improved survival outcomes in patients with metastatic triple-negative breast cancer (TNBC) exhibiting a PD-L1 combined positive score ≥ 10 [141].

In terms of applying AI, a study utilizing a digital pathology platform for PD-L1 scoring in breast cancer showed that an AI algorithm could predict chemotherapy outcomes in patients with TNBC, as well as in those with HER2-positive and ER-negative breast cancer [142]. The potential utility of AI as an aid to pathologists was highlighted in a multi-institutional ring study that showed that the degree of agreement among pathologists when assessing PD-L1 expression levels could be improved with AI assistance [143]. Moreover, the DL model was able to predict PD-L1 status from H&E-stained images, indicating the potential role of AI in clinical practice for decision support and quality assurance. AI can enhance patient management strategies by identifying cases susceptible to misinterpretation [144]. A representative example of the application of AI in PD-L1 assessment is shown in **Figure 4A**.

The significance of TILs within the tumor microenvironment (TME) continues to increase because of their correlation with improved prognosis and their predictive value for chemotherapy and immunotherapy responses in breast cancer [145,146]. However, the concordance rate for manual TILs assessment among pathologists remains suboptimal [147,148].

Several computational approaches have been suggested to address interobserver variability, including the recommendation of the International Immuno-Oncology Working Groups to incorporate a computational approach in TIL assessment [149]. Additionally, one AI model proposed novel immunogradient indicators by computing TIL density profiles across the tumor-stroma interface zone, demonstrating robust prognostic stratification that outperforms traditional clinical and pathologic variables [150]. Another AI model quantified stromal TIL scores and provided valuable assistance to pathologists, particularly when discordant interpretations arose. This model enhanced the concordance rate among the pathologists. Furthermore, the prediction of NAC response in patients with TNBC and HER2-positive breast cancer has been enhanced with the assistance of AI [151]. Using an identical

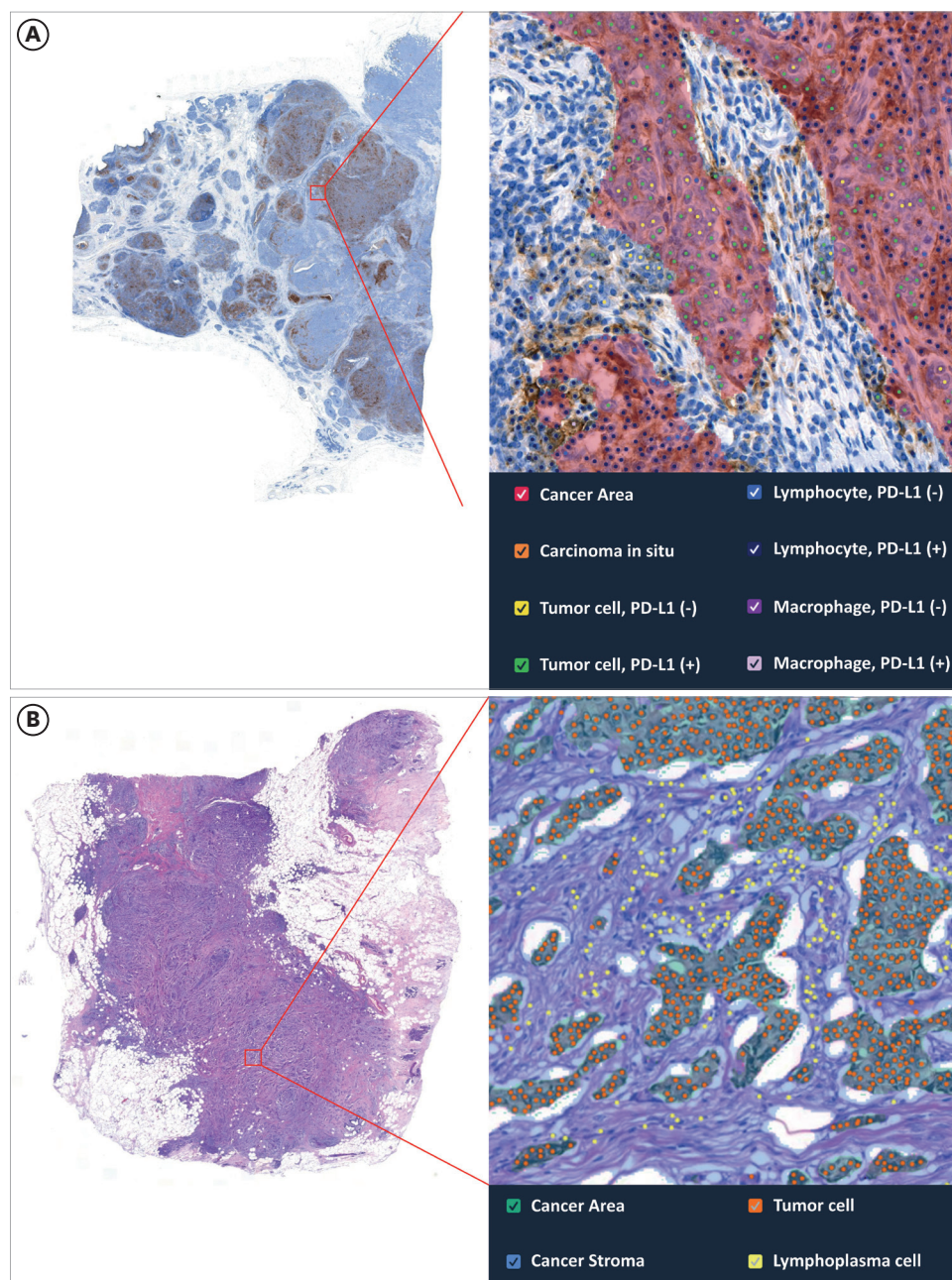


Figure 4. Example of artificial intelligence application in whole slide images.

(A) PD-L1 IHC-stained WSI. (B) H&E-stained WSI. Both figures illustrate the original WSI at low and high magnification with the AI outputs using the Lunit SCOPE PD-L1 (Lunit Inc.) for (A) and Lunit SCOPE IO (Lunit Inc.) for (B). Each cell type (\pm PD-L1 positivity) identified by the AI model is represented by colored dots, while the AI-segmented areas are depicted with colored patches.

PD-L1 = programmed cell death-ligand 1; IHC = immunohistochemistry; WSI = whole slide image; H&E = hematoxylin and eosin; AI = artificial intelligence; IO = immune-oncology.

AI model, the density of TIL was spatially analyzed, leading to determination of the immune phenotype (IP). One study revealed varying TIL densities and IPs across different molecular subtypes of breast cancer, suggesting a distinct immune landscape [152]. A representative example of the application of AI to the spatial analysis of TIL is shown in **Figure 4B**. An additional AI model has proposed digital stromal TILs and digital tumor-associated stroma

scores, based on the spatial relationships among TME components, showing prognostic significance in predicting disease-specific survival in patients with TNBC [153].

Beyond breast cancer, AI-powered TIL spatial evaluations are gaining traction in colorectal cancer, with promising implications for anti-HER2 therapy response prediction. This AI algorithm also enables the assessment of macrophage and fibroblast cell densities within the TME, potentially forecasting anti-HER2 therapy outcomes [135]. Another pan-carcinoma investigation revealed diminished intratumoral and stromal TIL densities in HER2-amplified tumors, as assessed using an AI model, alluding to a correlation between HER2 amplification and low immune infiltration [154].

AI for breast cancer risk stratification and genetic alteration prediction

Mammographic density, measured using the Breast Imaging Reporting and Data System (BI-RADS) category, has been investigated extensively, and it has been found that breast density is a strong risk factor for breast cancer [155]. Consequently, new breast screening strategies, such as those explored in the Dense Tissue and Early Breast Neoplasm Screening trials, now consider a woman's breast density to evaluate her risk [156]. However, the current standard for measuring breast density relies heavily on the subjective judgment of radiologists, which leads to significant inter-reader variability. This highlights the need for more objective and standardized approaches for assessing breast density to enhance screening accuracy and consistency.

Objective and consistent density measurements are crucial for individual risk stratification, leading to the development of automated assessment tools, such as Volpara, which calculates the volumetric breast density percentage of each mammogram on a continuous scale [157]. Another alternative is to develop density AI models trained using labeled data provided by radiologists. These AI models can provide automated and standardized breast density measurements, which are not only used to assess the risk of developing breast cancer but also as predictive surrogate markers for therapy response in high-risk patients [158]. Further research is necessary to determine the most suitable assessment tool and how to effectively integrate this information into routine clinical practice.

Traditional risk prediction models, such as the Tyrer-Cuzick model, also consider breast density as a part of the risk factors [159]. AI models have been incorporated to enhance the existing breast cancer prediction models [160]. A recent study by Arasu et al. [161] demonstrated that multiple AI models outperformed the Breast Cancer Surveillance Consortium (BCSC) risk model in predicting five-year breast cancer risk, with significantly better performance (AUC, 0.63–0.67 for AI models vs. AUC, 0.61 for BCSC model).

Additionally, AI algorithms can not only be trained on human-extracted features but can also analyze breast parenchymal patterns that may not be discernible to the human eye. Kim et al. [162] developed a model that utilizes Imaging Biomarkers in MMG, which are parenchymal patterns observed in high-risk individuals. This model can accurately predict cancer occurrence, even when trained solely with the unaffected breasts of patients with cancer. These models enable accurate short- and long-term risk predictions using MMGs from a single time point [163]. Another example is the ML model called Mirai, which performed better than previous DL models in identifying both five-year breast cancer risk and high-risk patients across diverse populations [164]. There is also ProFound AI, an AI-CAD-based concurrent-read predictive model for DBT cases, which helps reduce the workload and time

required to enhance radiologists' cancer detection performance. These models may be able to determine screening methods and frequencies for each individual.

The potential for the direct prediction of genetic alterations using AI models has been suggested, akin to the prediction of HER2 FISH status using AI models. The ShuffleNet-based DL algorithm consistently infers a wide range of genetic mutations, molecular tumor subtypes, gene expression signatures, and pathology biomarkers from H&E-stained slides across 14 of the most common solid tumor types, and detects mutations, such as *PIK3CA* and *MAP2K4* in breast cancer [122]. The ML model could predict molecular features, including DNA methylation, gene expression, copy number alterations, and somatic mutations. Additionally, AI models have been developed to predict germline *BRCA* mutation status and chromosomal instability status, both of which have a prognostic value [165,166]. Several studies have developed AI models to predict ODX risk scores, offering both prognostic and predictive insights for adjuvant systemic therapy, which can classify ODX risk categories by quantifying tubule nuclei and mitotic counts [78,80,167]. Similarly, Cho et al. [168] reported that an AI model could classify the ODX risk score with a cutoff value of 25. The predicted high-risk groups demonstrated significantly lower survival outcomes in patients with early stage HR-positive breast cancer, further underscoring the potential of AI for cancer prognostication and management [168].

AI in predicting clinical outcomes and treatment response

AI has been used to monitor and assess the prognosis of breast cancer. AI algorithms in conjunction with MRI scans were employed to evaluate the anticipated response to adjuvant and neoadjuvant treatments based on pretreatment imaging. By analyzing the imaging features and patterns, AI can assist in predicting treatment responses and optimizing treatment strategies to improve patient outcomes [169]. A similar endeavor occurs with ultrasonography, where AI predicts the response to NAC and helps forecast the overall breast cancer prognosis [170,171]. Additionally, AI has emerged as a potential tool for assessing the response to chemotherapy in post-treatment MRIs and predicting recurrence risk [172,173]. In the future, AI algorithms could analyze medical images, such as MRIs, and provide quantitative assessments and predictions that could assist radiologists and oncologists in their decision-making processes.

Turning the spotlight to pathology, the wealth of information extracted from pathological slides is a gold mine for predicting treatment responses and broader clinical outcomes. For example, an AI algorithm proposes a novel recurrence score (RS) with the potential to serve as a viable alternative to the more expensive 21-gene assays. This model analyzed different aspects of the cancer and surrounding tissues as well as the density of TILs and could help predict which high-risk patients would benefit from adjuvant chemotherapy. This suggests that the RS from the AI model may serve as a predictive biomarker for adjuvant chemotherapy responses [174]. In a comparative study of ML models utilizing clinical and pathological data, the random forest model demonstrated the highest performance, with an AUC of 0.88, for predicting pCR following NAC in patients with locally advanced or high-risk early breast cancer [175]. Recently, a CNN-based model trained on H&E-stained WSIs from core biopsies of TNBC patients after NAC was reported to have a positive predictive value of 73.7% for pCR [176]. Huang et al. [177] developed an AI-based automatic WSI feature extraction pipeline, named IMPRESS, using WSIs stained with both H&E and multiplex IHC (PD-L1, CD8+, and CD163+). ML models using features from IMPRESS and clinical variables accurately predicted the NAC response in patients with HER2+ or TNBC, surpassing a model trained with manually generated pathological features, suggesting that it may be a preferred method for

developing algorithms to predict treatment responses in the future. Upon external validation, these models produced promising results, especially for the HER2+ subtype (AUC = 0.90 for HER2+, and 0.59 for TNBC) [177]. Furthermore, a multi-omics ML model, trained on a combination of clinical, DNA, RNA, digital pathology, and treatment features, showed an AUC of 0.87 in predicting pCR following NAC, with or without HER2-targeted therapy [178].

CURRENT ADAPTATIONS, LIMITATIONS, AND HURDLES

As of October 2022, numerous AI systems operating as Medical Devices have obtained regulatory approval worldwide, including at least 521 devices that have received clearance from the US FDA [179].

However, there are several limitations for AI to be addressed in the field of breast cancer. The clinical validation of AI models in the real world is a primary concern. Despite the remarkable performance of AI algorithms in numerous studies, it is indispensable for these algorithms to undergo clinical validation with large-scale datasets before integrating into clinical practice. Recent attempts at clinical validation, particularly those focusing on treatment outcomes [151,153,168], are often constrained by retrospective designs that might introduce unexpected biases, underscoring the need for prospective studies [16]. These prospective studies are crucial to fully comprehend the impact of AI implementation on clinical practice and ensure that AI tools are both effective and reliable in a clinical setting.

Second, integrating AI models into real-world clinical settings presents challenges beyond validation, such as utility and usability. For utility, AI models should undergo rigorous validation through RCTs that assess a range of clinical endpoints. These endpoints should include not only overall survival but also disease control, toxicity reduction, improved quality of life, and decreased healthcare resource utilization. On the usability front, AI models need to be tested in real-world settings for time efficiency, user satisfaction, and acceptance of AI recommendations. Additionally, the incorporation of a feedback mechanism through post-market surveillance is essential to identify potential weaknesses and areas for enhancement, thereby ensuring the continuous improvement of these models [14]. Real-time monitoring systems for physicians and AI algorithm developers are necessary to ensure the safe delivery of care. Additionally, seamless integration with existing workflows, such as PACS, is necessary for the efficient use of AI tools.

A third concern is the generalizability or robustness of the AI model, which refers to its consistent performance across various datasets, including the one on which they were trained. Several strategies to address this include using datasets with a wide array of preanalytic and analytic factors to enhance model robustness, although acquiring large-scale datasets with manual annotations presents challenges in the development of AI algorithms [77,180]. To address this issue, innovative strategies, such as unsupervised learning and Generative Adversarial Networks, are being utilized [181,182]. Yet another challenge that arises is the risk of misrepresenting health concerns in minority populations, owing to the creation of AI models largely based on data from majority populations. This situation can potentially exacerbate health disparities [183,184].

Fourth, most AI algorithms are often considered black boxes because it is often unclear which features can be recognized within them. In this regard, the approaches to develop explainable

AI algorithms could build trust among clinicians, provide transparency in the decision-making process, and alleviate various types of biases [185]. In contrast, Ghassemi et al. [186] suggested that enthusiastic internal and external validation of AI algorithms could be a direct means of achieving goals associated with explainability.

Finally, issues related to reimbursement must be considered, particularly if AI systems start to replace certain roles traditionally performed by physicians. These broader discussions on reimbursement and the impact of AI on healthcare systems should take place at the national or screening program levels to ensure equitable and effective implementation [187].

To facilitate discussions and engagement among clinicians, the wider medical team, responsible national agencies, and hospitals regarding the integration of AI in healthcare, it is important to explore the overarching challenges. One significant challenge is the substantial investment required for AI development, including the development of algorithms and necessary IT infrastructure. This investment encompasses not only the initial costs, but also the ongoing maintenance and updating of AI systems.

CONCLUSION

The innovative intersection of AI and breast cancer care promises to revolutionize disease screening, disease diagnosis, biomarker evaluation, prognostication, and treatment strategies by overcoming human limitations and achieving remarkable precision and efficiency. However, the journey towards the full-scale clinical adoption of AI is not without hurdles. Key challenges encompass clinical validation, ensuring algorithmic robustness across diverse datasets, grappling with the 'black box' enigma of AI, and navigating the complex terrain of regulatory, legal, and economic considerations. Moreover, addressing potential biases, particularly those that negatively affect minor populations, and assessing performance using reliable metrics are critical for building equitable and trustworthy AI systems.

Looking ahead, the future of AI in breast cancer care is contingent on our collective ability to overcome these challenges. The design and implementation of large-scale prospective studies are essential for validating the clinical efficacy of AI algorithms. Developing models that are transparent and interpretable and fostering strategies to improve the generalizability of AI systems will facilitate their wider acceptance. Moreover, creating regulatory frameworks and ethical guidelines will ensure the responsible integration of AI in healthcare. Furthermore, the issue of cost, both initial and ongoing maintenance, is a significant barrier to AI adoption. Hence, future initiatives should focus on devising sustainable financing models to mitigate the financial burden of AI development.

Although the path ahead is marked by complexity, the potential benefits of integrating AI into breast cancer management are too significant to be ignored. By navigating these challenges with careful deliberation, we have the opportunity to drastically improve patient outcomes, reduce health disparities, and set the stage for a new chapter in precision medicine. As we continue to explore and innovate, the integration of AI in breast cancer care redefines our approach to screening, diagnosis, and treatment in unimaginable ways.

ACKNOWLEDGMENTS

This study was supported by Lunit Inc. We sincerely thank Seulki Kim for diligent proofreading of the manuscript.

REFERENCES

1. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al. Breast cancer statistics, 2022. *CA Cancer J Clin* 2022;72:524-41.
[PUBMED](#) | [CROSSREF](#)
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209-49.
[PUBMED](#) | [CROSSREF](#)
3. Taylor C, McGale P, Probert J, Broggio J, Charman J, Darby SC, et al. Breast cancer mortality in 500 000 women with early invasive breast cancer diagnosed in England, 1993–2015: population based observational cohort study. *BMJ* 2023;381:e074684.
[PUBMED](#) | [CROSSREF](#)
4. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013;108:2205-40.
[PUBMED](#) | [CROSSREF](#)
5. The Royal College of Radiologists. RCR Clinical Radiology Workforce Census 2022. London: The Royal College of Radiologists; 2022.
6. Metter DM, Colgan TJ, Leung ST, Timmons CF, Park JY. Trends in the US and Canadian pathologist workforces from 2007 to 2017. *JAMA Netw Open* 2019;2:e194337.
[PUBMED](#) | [CROSSREF](#)
7. Connor SJ, Lim YY, Tate C, Entwistle H, Morris J, Whiteside S, et al. A comparison of reading times in full-field digital mammography and digital breast tomosynthesis. *Breast Cancer Res* 2012;14:P26.
8. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson AN, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA* 2015;313:1122-32.
[PUBMED](#) | [CROSSREF](#)
9. Acs B, Fredriksson I, Rönnlund C, Hagerling C, Ehinger A, Kovács A, et al. Variability in breast cancer biomarker assessment and the effect on oncological treatment decisions: a nationwide 5-year population-based study. *Cancers (Basel)* 2021;13:1166.
[PUBMED](#) | [CROSSREF](#)
10. Fernandez AI, Liu M, Bellizzi A, Brock J, Fadare O, Hanley K, et al. Examination of low ERBB2 protein expression in breast cancer tissue. *JAMA Oncol* 2022;8:1-4.
[PUBMED](#) | [CROSSREF](#)
11. Kim SH, Lee EH, Jun JK, Kim YM, Chang YW, Lee JH, et al. Interpretive performance and inter-observer agreement on digital mammography test sets. *Korean J Radiol* 2019;20:218-24.
[PUBMED](#) | [CROSSREF](#)
12. Whitney J, Corredor G, Janowczyk A, Ganesan S, Doyle S, Tomaszewski J, et al. Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer. *BMC Cancer* 2018;18:610.
[PUBMED](#) | [CROSSREF](#)
13. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med* 2022;28:31-8.
[PUBMED](#) | [CROSSREF](#)
14. Kann BH, Hosny A, Aerts HJ. Artificial intelligence for clinical oncology. *Cancer Cell* 2021;39:916-27.
[PUBMED](#) | [CROSSREF](#)
15. Niazi MK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* 2019;20:e253-61.
[PUBMED](#) | [CROSSREF](#)
16. Hickman SE, Baxter GC, Gilbert FJ. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. *Br J Cancer* 2021;125:15-22.
[PUBMED](#) | [CROSSREF](#)

17. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up†. *Ann Oncol* 2019;30:1194-220.
[PUBMED](#) | [CROSSREF](#)
18. Gradishar WJ, Moran MS, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast cancer, version 3.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2022;20:691-722.
[PUBMED](#) | [CROSSREF](#)
19. Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007;57:75-89.
[PUBMED](#) | [CROSSREF](#)
20. Tice JA, Miglioretti DL, Li CS, Vachon CM, Gard CC, Kerlikowske K. Breast density and benign breast disease: risk assessment to identify women at high risk of breast cancer. *J Clin Oncol* 2015;33:3137-43.
[PUBMED](#) | [CROSSREF](#)
21. Gail MH. Choosing breast cancer risk models: importance of independent validation. *J Natl Cancer Inst* 2020;112:433-5.
[PUBMED](#) | [CROSSREF](#)
22. Holm J, Li J, Darabi H, Eklund M, Eriksson M, Humphreys K, et al. Associations of breast cancer risk prediction tools with tumor characteristics and metastasis. *J Clin Oncol* 2016;34:251-8.
[PUBMED](#) | [CROSSREF](#)
23. Gilbert FJ, Tucker L, Gillan MG, Willsher P, Cooke J, Duncan KA, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme--a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health Technol Assess* 2015;19:i-xxv, 1-136.
[PUBMED](#) | [CROSSREF](#)
24. Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology* 2009;253:641-51.
[PUBMED](#) | [CROSSREF](#)
25. Redondo A, Comas M, Macià F, Ferrer F, Murta-Nascimento C, Maristany MT, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol* 2012;85:1465-70.
[PUBMED](#) | [CROSSREF](#)
26. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
[PUBMED](#) | [CROSSREF](#)
27. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90.
[CROSSREF](#)
28. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Richter CD, Cha KH. Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. *IEEE Trans Med Imaging* 2019;38:686-96.
[PUBMED](#) | [CROSSREF](#)
29. Dembrower K, Wählin E, Liu Y, Salim M, Smith K, Lindholm P, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *Lancet Digit Health* 2020;2:e468-74.
[PUBMED](#) | [CROSSREF](#)
30. Ravichandran K, Braman N, Janowczyk A, Madabhushi A. A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI. *Proceedings of the SPIE*; 2018 Feb 10-15; Houston, TX. Bellingham, WA: SPIE; 2018.
31. Liu Y, Han D, Parwani AV, Li Z. Applications of artificial intelligence in breast pathology. *Arch Pathol Lab Med* 2023;147:1003-13.
[CROSSREF](#)
32. Picard JD. History of mammography. *Bull Acad Natl Med* 1998;182:1613-20.
[PUBMED](#)
33. World Health Organization. WHO position paper on mammography screening. Geneva: World Health Organization; 2014.
34. Ahern CH, Shen Y. Cost-effectiveness analysis of mammography and clinical breast examination strategies: a comparison with current guidelines. *Cancer Epidemiol Biomarkers Prev* 2009;18:718-25.
[PUBMED](#) | [CROSSREF](#)
35. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356:227-36.
[PUBMED](#) | [CROSSREF](#)

36. Taylor-Phillips S, Jenkinson D, Stinton C, Wallis MG, Dunn J, Clarke A. Double reading in breast cancer screening: cohort evaluation in the CO-OPS trial. *Radiology* 2018;287:749-57.
[PUBMED](#) | [CROSSREF](#)
37. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR Am J Roentgenol* 2013;200:1401-8.
[PUBMED](#) | [CROSSREF](#)
38. Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;311:2499-507.
[PUBMED](#) | [CROSSREF](#)
39. Gur D, Abrams GS, Chough DM, Ganott MA, Hakim CM, Perrin RL, et al. Digital breast tomosynthesis: observer performance study. *AJR Am J Roentgenol* 2009;193:586-91.
[PUBMED](#) | [CROSSREF](#)
40. Giger ML, Chan HP, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. *Med Phys* 2008;35:5799-820.
[PUBMED](#) | [CROSSREF](#)
41. Nagi J, Kareem SA, Nagi F, Ahmed SK. Automated breast profile segmentation for ROI detection using digital mammograms. *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*; 2010 Nov 30-Dec 2; Kuala Lumpur, Malaysia. Piscataway, NJ: IEEE; 2010. 87-92 p.
42. 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;36:311-6.
[PUBMED](#) | [CROSSREF](#)
43. El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. *IEEE Trans Med Imaging* 2002;21:1552-63.
[PUBMED](#) | [CROSSREF](#)
44. Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008;359:1675-84.
[PUBMED](#) | [CROSSREF](#)
45. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med* 2015;175:1828-37.
[PUBMED](#) | [CROSSREF](#)
46. Fenton JJ, Abraham L, Taplin SH, Geller BM, Carney PA, D'Orsi C, et al. Effectiveness of computer-aided detection in community mammography practice. *J Natl Cancer Inst* 2011;103:1152-61.
[PUBMED](#) | [CROSSREF](#)
47. Salim M, Dembrower K, Eklund M, Lindholm P, Strand F. Range of radiologist performance in a population-based screening cohort of 1 million digital mammography examinations. *Radiology* 2020;297:33-9.
[PUBMED](#) | [CROSSREF](#)
48. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007;356:1399-409.
[PUBMED](#) | [CROSSREF](#)
49. Kohli A, Jha S. Why CAD failed in mammography. *J Am Coll Radiol* 2018;15:535-7.
[PUBMED](#) | [CROSSREF](#)
50. Sun L, Wang J, Hu Z, Xu Y, Cui Z. Multi-view convolutional neural networks for mammographic image classification. *IEEE Access* 2019;7:126273-82.
[CROSSREF](#)
51. Mutasa S, Sun S, Ha R. Understanding artificial intelligence based radiology studies: what is overfitting? *Clin Imaging* 2020;65:96-9.
[PUBMED](#) | [CROSSREF](#)
52. Lee SE, Yoon JH, Hong H, Son NH, Kim EK. One-on-one comparison between conventional CAD and AI-CAD applied to screening mammography. *J Korean Soc Breast Screen* 2023;20:19-29.
53. Do YA, Jang M, Yun B, Shin SU, Kim B, Kim SM. Diagnostic performance of artificial intelligence-based computer-aided diagnosis for breast microcalcification on mammography. *Diagnostics (Basel)* 2021;11:1409.
[PUBMED](#) | [CROSSREF](#)
54. Yoon JH, Strand F, Baltzer PA, Conant EF, Gilbert FJ, Lehman CD, et al. Standalone AI for breast cancer detection at screening digital mammography and digital breast tomosynthesis: a systematic review and meta-analysis. *Radiology* 2023;307:e222639.
[PUBMED](#) | [CROSSREF](#)

55. Dembrower K, Salim M, Eklund M, Lindholm P, Strand F. Implications for downstream workload based on calibrating an artificial intelligence detection algorithm by standalone-reader or combined-reader sensitivity matching. *J Med Imaging (Bellingham)* 2023;10:S22405.
[PUBMED](#) | [CROSSREF](#)
56. Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health* 2020;2:e138-48.
[PUBMED](#) | [CROSSREF](#)
57. Conant EF, Toledano AY, Periaswamy S, Fotin SV, Go J, Boatsman JE, et al. Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis. *Radiol Artif Intell* 2019;1:e180096.
[PUBMED](#) | [CROSSREF](#)
58. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Teuwen J, Broeders M, Gennaro G, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. *Eur Radiol* 2019;29:4825-32.
[PUBMED](#) | [CROSSREF](#)
59. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89-94.
[PUBMED](#) | [CROSSREF](#)
60. Niraula S, Biswanger N, Hu P, Lambert P, Decker K. Incidence, characteristics, and outcomes of interval breast cancers compared with screening-detected breast cancers. *JAMA Netw Open* 2020;3:e2018179.
[PUBMED](#) | [CROSSREF](#)
61. Wanders AJ, Mees W, Bun PA, Janssen N, Rodríguez-Ruiz A, Dalmış MU, et al. Interval cancer detection using a neural network and breast density in women with negative screening mammograms. *Radiology* 2022;303:269-75.
[PUBMED](#) | [CROSSREF](#)
62. Hickman S, Gilbert F. Investigating a stand-alone ai system prompt accuracy for interval cancer detection in screening mammography. *RSNA 2022 Empowering Patients and Partners in Care; 2022 Nov 27-Dec 11; Chicago, IL. Oak Brook, IL: RSNA; 2023.*
63. Dembrower K, Crippa A, Colón E, Eklund M, Strand F; ScreenTrustCAD Trial Consortium. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health* 2023;5:E703-11.
[CROSSREF](#)
64. Lång K, Josefsson V, Larsson AM, Larsson S, Högberg C, Sartor H, et al. 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* 2023;24:936-44.
[PUBMED](#) | [CROSSREF](#)
65. Chang YW, An JK, Choi N, Ko KH, Kim KH, Han K, et al. Artificial intelligence for breast cancer screening in mammography (AI-STREAM): a prospective multicenter study design in Korea using AI-based CAdE/x. *J Breast Cancer* 2022;25:57-68.
[PUBMED](#) | [CROSSREF](#)
66. Han SS, Kim YJ, Moon IJ, Jung JM, Lee MY, Lee WJ, et al. Evaluation of artificial intelligence-assisted diagnosis of skin neoplasms: a single-center, paralleled, unmasked, randomized controlled trial. *J Invest Dermatol* 2022;142:2353-2362.e2.
[PUBMED](#) | [CROSSREF](#)
67. Cossío F, Schurz H, Engström M, Barck-Holst C, Tsirikoglou A, Lundström C, et al. VAI-B: a multicenter platform for the external validation of artificial intelligence algorithms in breast imaging. *J Med Imaging (Bellingham)* 2023;10:061404.
[PUBMED](#) | [CROSSREF](#)
68. Shen Y, Shamout FE, Oliver JR, Witowski J, Kannan K, Park J, et al. Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. *Nat Commun* 2021;12:5645.
[PUBMED](#) | [CROSSREF](#)
69. Knoll F, Murrell T, Sriram A, Yakubova N, Zbontar J, Rabbat M, et al. Advancing machine learning for MR image reconstruction with an open competition: overview of the 2019 fastMRI challenge. *Magn Reson Med* 2020;84:3054-70.
[PUBMED](#) | [CROSSREF](#)
70. van Zelst JC, Tan T, Clauser P, Domingo A, Dorrius MD, Drieling D, et al. Dedicated computer-aided detection software for automated 3D breast ultrasound; an efficient tool for the radiologist in supplemental screening of women with dense breasts. *Eur Radiol* 2018;28:2996-3006.
[PUBMED](#) | [CROSSREF](#)

71. Dalmış MU, Vreemann S, Kooi T, Mann RM, Karssemeijer N, Gubern-Mérida A. Fully automated detection of breast cancer in screening MRI using convolutional neural networks. *J Med Imaging (Bellingham)* 2018;5:014502.
[PUBMED](#) | [CROSSREF](#)
72. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991;19:403-10.
[PUBMED](#) | [CROSSREF](#)
73. Badve SS. Artificial intelligence in breast pathology - dawn of a new era. *NPJ Breast Cancer* 2023;9:5.
[PUBMED](#) | [CROSSREF](#)
74. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NN, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. *Sci Rep* 2017;7:46450.
[PUBMED](#) | [CROSSREF](#)
75. Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S. Breast cancer multi-classification from histopathological images with structured deep learning model. *Sci Rep* 2017;7:4172.
[PUBMED](#) | [CROSSREF](#)
76. Polónia A, Campelos S, Ribeiro A, Aymore I, Pinto D, Biskup-Fruzynska M, et al. Artificial intelligence improves the accuracy in histologic classification of breast lesions. *Am J Clin Pathol* 2021;155:527-36.
[PUBMED](#) | [CROSSREF](#)
77. Sandbank J, Bataillon G, Nudelman A, Krasnitsky I, Mikulinsky R, Bien L, et al. Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. *NPJ Breast Cancer* 2022;8:129.
[PUBMED](#) | [CROSSREF](#)
78. Romo-Bucheli D, Janowczyk A, Gilmore H, Romero E, Madabhushi A. Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images. *Sci Rep* 2016;6:32706.
[PUBMED](#) | [CROSSREF](#)
79. Lu C, Romo-Bucheli D, Wang X, Janowczyk A, Ganesan S, Gilmore H, et al. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. *Lab Invest* 2018;98:1438-48.
[PUBMED](#) | [CROSSREF](#)
80. Romo-Bucheli D, Janowczyk A, Gilmore H, Romero E, Madabhushi A. A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. *Cytometry A* 2017;91:566-73.
[PUBMED](#) | [CROSSREF](#)
81. Roux L, Racoceanu D, Loménie N, Kulikova M, Irshad H, Klossa J, et al. Mitosis detection in breast cancer histological images an ICPR 2012 contest. *J Pathol Inform* 2013;4:8.
[PUBMED](#) | [CROSSREF](#)
82. 22nd International Conference on Pattern Recognition (ICPR). 2014. MITOS & ATYPIA 14 Contest. <https://mitos-atyfia-14.grand-challenge.org/Home/>. Accessed August 10th, 2023.
83. Veta M, Heng YJ, Stathonikos N, Bejnordi BE, Beca F, Wollmann T, et al. Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. *Med Image Anal* 2019;54:111-21.
[PUBMED](#) | [CROSSREF](#)
84. Li C, Wang X, Liu W, Latecki LJ. DeepMitosis: mitosis detection via deep detection, verification and segmentation networks. *Med Image Anal* 2018;45:121-33.
[PUBMED](#) | [CROSSREF](#)
85. Sebai M, Wang X, Wang T. MaskMitosis: a deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images. *Med Biol Eng Comput* 2020;58:1603-23.
[PUBMED](#) | [CROSSREF](#)
86. Mahmood T, Arsalan M, Owais M, Lee MB, Park KR. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *J Clin Med* 2020;9:749.
[PUBMED](#) | [CROSSREF](#)
87. Pantanowitz L, Hartman D, Qi Y, Cho EY, Suh B, Paeng K, et al. Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses. *Diagn Pathol* 2020;15:80.
[PUBMED](#) | [CROSSREF](#)
88. Nateghi R, Danyali H, Helfroush MS. A deep learning approach for mitosis detection: application in tumor proliferation prediction from whole slide images. *Artif Intell Med* 2021;114:102048.
[PUBMED](#) | [CROSSREF](#)

89. Wang Y, Acs B, Robertson S, Liu B, Solorzano L, Wählby C, et al. Improved breast cancer histological grading using deep learning. *Ann Oncol* 2022;33:89-98.
[PUBMED](#) | [CROSSREF](#)
90. Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, et al. Lymph node metastasis prediction from primary breast cancer US images using deep learning. *Radiology* 2020;294:19-28.
[PUBMED](#) | [CROSSREF](#)
91. Hou R, Grimm LJ, Mazurowski MA, Marks JR, King LM, Maley CC, et al. Prediction of upstaging in ductal carcinoma in situ based on mammographic radiomic features. *Radiology* 2022;303:54-62.
[PUBMED](#) | [CROSSREF](#)
92. Zardavas D, Irrthum A, Swanton C, Piccart M. Clinical management of breast cancer heterogeneity. *Nat Rev Clin Oncol* 2015;12:381-94.
[PUBMED](#) | [CROSSREF](#)
93. Zhang T, Tan T, Han L, Appelman L, Veltman J, Wessels R, et al. Predicting breast cancer types on and beyond molecular level in a multi-modal fashion. *NPJ Breast Cancer* 2023;9:16.
[PUBMED](#) | [CROSSREF](#)
94. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, et al. Breast cancer-major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017;67:290-303.
[PUBMED](#) | [CROSSREF](#)
95. Apple SK. Sentinel lymph node in breast cancer: review article from a pathologist's point of view. *J Pathol Transl Med* 2016;50:83-95.
[PUBMED](#) | [CROSSREF](#)
96. Krag DN, Anderson SJ, Julian TB, Brown AM, Harlow SP, Ashikaga T, et al. Technical outcomes of sentinel-lymph-node resection and conventional axillary-lymph-node dissection in patients with clinically node-negative breast cancer: results from the NSABP B-32 randomised phase III trial. *Lancet Oncol* 2007;8:881-8.
[PUBMED](#) | [CROSSREF](#)
97. Caldonazzi N, Rizzo PC, Eccher A, Girolami I, Fanelli GN, Naccarato AG, et al. Value of artificial intelligence in evaluating lymph node metastases. *Cancers (Basel)* 2023;15:2491.
[PUBMED](#) | [CROSSREF](#)
98. Holtén-Rossing H, Talman MM, Jylling AM, Laenkholm AV, Kristensson M, Vainer B. Application of automated image analysis reduces the workload of manual screening of sentinel lymph node biopsies in breast cancer. *Histopathology* 2017;71:866-73.
[PUBMED](#) | [CROSSREF](#)
99. Weaver DL, Krag DN, Manna EA, Ashikaga T, Harlow SP, Bauer KD. Comparison of pathologist-detected and automated computer-assisted image analysis detected sentinel lymph node micrometastases in breast cancer. *Mod Pathol* 2003;16:1159-63.
[PUBMED](#) | [CROSSREF](#)
100. Clarke GM, Peressotti C, Holloway CM, Zubovits JT, Liu K, Yaffe MJ. Development and evaluation of a robust algorithm for computer-assisted detection of sentinel lymph node micrometastases. *Histopathology* 2011;59:116-28.
[PUBMED](#) | [CROSSREF](#)
101. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199-210.
[PUBMED](#) | [CROSSREF](#)
102. Lee B, Paeng K. A robust and effective approach towards accurate metastasis detection and pN-stage classification in breast cancer. 21st International Conference on Medical Image Computing and Computer-Assisted Intervention 2018; 2018 Sep 16-20; Granada, Spain. Cham: Springer Nature Switzerland; 2018. 841-50 p.
103. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists. *Arch Pathol Lab Med* 2019;143:859-68.
[PUBMED](#) | [CROSSREF](#)
104. Steiner DF, MacDonald R, Liu Y, Truszowski P, Hipp JD, Gammage C, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *Am J Surg Pathol* 2018;42:1636-46.
[PUBMED](#) | [CROSSREF](#)
105. Kim YG, Kim S, Cho CE, Song IH, Lee HJ, Ahn S, et al. Effectiveness of transfer learning for enhancing tumor classification with a convolutional neural network on frozen sections. *Sci Rep* 2020;10:21899.
[PUBMED](#) | [CROSSREF](#)

106. Kim YG, Song IH, Lee H, Kim S, Yang DH, Kim N, et al. Challenge for diagnostic assessment of deep learning algorithm for metastases classification in sentinel lymph nodes on frozen tissue section digital slides in women with breast cancer. *Cancer Res Treat* 2020;52:1103-11.
[PUBMED](#) | [CROSSREF](#)
107. Hu Y, Su F, Dong K, Wang X, Zhao X, Jiang Y, et al. Deep learning system for lymph node quantification and metastatic cancer identification from whole-slide pathology images. *Gastric Cancer* 2021;24:868-77.
[PUBMED](#) | [CROSSREF](#)
108. Huang SC, Chen CC, Lan J, Hsieh TY, Chuang HC, Chien MY, et al. Deep neural network trained on gigapixel images improves lymph node metastasis detection in clinical settings. *Nat Commun* 2022;13:3347.
[PUBMED](#) | [CROSSREF](#)
109. Allison KH, Hammond ME, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *J Clin Oncol* 2020;38:1346-66.
[PUBMED](#) | [CROSSREF](#)
110. Rai PD, Vagha S, Shukla S, Bhake A. Comparison of various scoring systems by immunohistochemistry for evaluating hormone receptors (estrogen receptor and progesterone receptor) in carcinoma of breast. *J Datta Meghe Inst Med Sci Univ* 2020;15:202-8.
[CROSSREF](#)
111. Diaz LK, Sahin A, Sneige N. Interobserver agreement for estrogen receptor immunohistochemical analysis in breast cancer: a comparison of manual and computer-assisted scoring methods. *Ann Diagn Pathol* 2004;8:23-7.
[PUBMED](#) | [CROSSREF](#)
112. Gokhale S, Rosen D, Sneige N, Diaz LK, Resetkova E, Sahin A, et al. Assessment of two automated imaging systems in evaluating estrogen receptor status in breast carcinoma. *Appl Immunohistochem Mol Morphol* 2007;15:451-5.
[PUBMED](#) | [CROSSREF](#)
113. Rexhepaj E, Brennan DJ, Holloway P, Kay EW, McCann AH, Landberg G, et al. Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of oestrogen and progesterone receptor levels in breast cancer. *Breast Cancer Res* 2008;10:R89.
[PUBMED](#) | [CROSSREF](#)
114. Turbin DA, Leung S, Cheang MC, Kennecke HA, Montgomery KD, McKinney S, et al. Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: a tissue microarray study of 3,484 cases. *Breast Cancer Res Treat* 2008;110:417-26.
[PUBMED](#) | [CROSSREF](#)
115. Faratian D, Kay C, Robson T, Campbell FM, Grant M, Rea D, et al. Automated image analysis for high-throughput quantitative detection of ER and PR expression levels in large-scale clinical studies: the TEAM Trial Experience. *Histopathology* 2009;55:587-93.
[PUBMED](#) | [CROSSREF](#)
116. Bolton KL, Garcia-Closas M, Pfeiffer RM, Duggan MA, Howat WJ, Hewitt SM, et al. Assessment of automated image analysis of breast cancer tissue microarrays for epidemiologic studies. *Cancer Epidemiol Biomarkers Prev* 2010;19:992-9.
[PUBMED](#) | [CROSSREF](#)
117. Abele N, Tiemann K, Krech T, Wellmann A, Schaaf C, Länger F, et al. Noninferiority of artificial intelligence-assisted analysis of Ki-67 and estrogen/progesterone receptor in breast cancer routine diagnostics. *Mod Pathol* 2023;36:100033.
[PUBMED](#) | [CROSSREF](#)
118. Shafi S, Kellough DA, Lujan G, Satturwar S, Parwani AV, Li Z. Integrating and validating automated digital imaging analysis of estrogen receptor immunohistochemistry in a fully digital workflow for clinical use. *J Pathol Inform* 2022;13:100122.
[PUBMED](#) | [CROSSREF](#)
119. Shamai G, Binenbaum Y, Slossberg R, Duek I, Gil Z, Kimmel R. Artificial intelligence algorithms to assess hormonal status from tissue microarrays in patients with breast cancer. *JAMA Netw Open* 2019;2:e197700.
[PUBMED](#) | [CROSSREF](#)
120. Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer* 2018;4:30.
[PUBMED](#) | [CROSSREF](#)

121. Naik N, Madani A, Esteva A, Keskar NS, Press MF, Ruderman D, et al. Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. *Nat Commun* 2020;11:5727.
[PUBMED](#) | [CROSSREF](#)
122. Kather JN, Heij LR, Grabsch HI, Loeffler C, Ehle A, Muti HS, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Can* 2020;1:789-99.
[PUBMED](#) | [CROSSREF](#)
123. Wolff AC, Somerfield MR, Dowsett M, Hammond MEH, Hayes DF, McShane LM, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology-College of American Pathologists guideline update. *Arch Pathol Lab Med* 2023;147:993-1000.
[CROSSREF](#)
124. Brüggemann A, Eld M, Lelkaitis G, Nielsen S, Grunkin M, Hansen JD, et al. Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains. *Breast Cancer Res Treat* 2012;132:41-9.
[PUBMED](#) | [CROSSREF](#)
125. Jung M, Song SG, Cho SI, Jung W, Oum C, Song H, et al. Artificial intelligence-powered human epidermal growth factor receptor 2 (HER2) analyzer in breast cancer as an assistance tool for pathologists to reduce interobserver variation. *J Clin Oncol* 2022;40:e12543.
[CROSSREF](#)
126. Furrer D, Jacob S, Caron C, Sanschagrin F, Provencher L, Diorio C. Validation of a new classifier for the automated analysis of the human epidermal growth factor receptor 2 (HER2) gene amplification in breast cancer specimens. *Diagn Pathol* 2013;8:17.
[PUBMED](#) | [CROSSREF](#)
127. Zakrzewski F, de Back W, Weigert M, Wenke T, Zeugner S, Mantey R, et al. Automated detection of the HER2 gene amplification status in Fluorescence in situ hybridization images for the diagnostics of cancer tissues. *Sci Rep* 2019;9:8231.
[PUBMED](#) | [CROSSREF](#)
128. Xue T, Chang H, Ren M, Wang H, Yang Y, Wang B, et al. Deep learning to automatically evaluate HER2 gene amplification status from fluorescence in situ hybridization images. *Sci Rep* 2023;13:9746.
[PUBMED](#) | [CROSSREF](#)
129. Anand D, Kurian NC, Dhage S, Kumar N, Rane S, Gann PH, et al. Deep learning to estimate human epidermal growth factor receptor 2 status from hematoxylin and eosin-stained breast tissue images. *J Pathol Inform* 2020;11:19.
[PUBMED](#) | [CROSSREF](#)
130. Conde-Sousa E, Vale J, Feng M, Xu K, Wang Y, Della Mea V, et al. HEROHE challenge: predicting HER2 status in breast cancer from hematoxylin-eosin whole-slide imaging. *J Imaging* 2022;8:213.
[PUBMED](#) | [CROSSREF](#)
131. Farahmand S, Fernandez AI, Ahmed FS, Rimm DL, Chuang JH, Reisenbichler E, et al. Deep learning trained on hematoxylin and eosin tumor region of Interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. *Mod Pathol* 2022;35:44-51.
[PUBMED](#) | [CROSSREF](#)
132. Cho SY, Lim Y, Cho SI, Kim S, Park G, Song S, et al. Artificial intelligence (AI) - powered human epidermal growth factor receptor-2 (HER2) and tumor-infiltrating lymphocytes (TIL) analysis for HER2-positive early breast cancer patients treated with HER2-targeted neoadjuvant chemotherapy (NAC). *Ann Oncol* 2022;33:S610.
[CROSSREF](#)
133. Kim M, Lee JL, Shin SJ, Bae WK, Lee HJ, Byun JH, et al. Phase II study of a trastuzumab biosimilar in combination with paclitaxel for HER2-positive recurrent or metastatic urothelial carcinoma: KCSG GU18-18. *ESMO Open* 2023;8:101588.
[PUBMED](#) | [CROSSREF](#)
134. Lee CK, Chon HJ, Chon J, Lee MA, Im HS, Jang JS, et al. Trastuzumab plus FOLFOX for gemcitabine/cisplatin refractory HER2-positive biliary tract cancer: a multi-institutional phase II trial of the Korean Cancer Study Group (KCSG-HB19-14). *J Clin Oncol* 2022;40:4096.
[CROSSREF](#)
135. Imai M, Nakamura Y, Okamoto W, Kato T, Esaki T, Kato K, et al. Artificial intelligence (AI)-powered HER2 quantification continuous score (QCS) and tumor microenvironment (TME) analysis in HER2-amplified metastatic colorectal cancer (mCRC) treated with pertuzumab plus trastuzumab. *JCO Glob Oncol* 2023;9:34.
136. Harbeck N, Rastogi P, Martin M, Tolaney SM, Shao ZM, Fasching PA, et al. Adjuvant abemaciclib combined with endocrine therapy for high-risk early breast cancer: updated efficacy and Ki-67 analysis from the monarchE study. *Ann Oncol* 2021;32:1571-81.
[PUBMED](#) | [CROSSREF](#)

137. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol* 2010;11:174-83.
[PUBMED](#) | [CROSSREF](#)
138. Polley MY, Leung SC, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. An international Ki67 reproducibility study. *J Natl Cancer Inst* 2013;105:1897-906.
[PUBMED](#) | [CROSSREF](#)
139. Yamamoto S, Chishima T, Mastubara Y, Adachi S, Harada F, Toda Y, et al. Variability in measuring the Ki-67 labeling index in patients with breast cancer. *Clin Breast Cancer* 2015;15:e35-9.
[PUBMED](#) | [CROSSREF](#)
140. Acs B, Pelekanou V, Bai Y, Martinez-Morilla S, Toki M, Leung SC, et al. Ki67 reproducibility using digital image analysis: an inter-platform and inter-operator study. *Lab Invest* 2019;99:107-17.
[PUBMED](#) | [CROSSREF](#)
141. Debien V, De Caluwé A, Wang X, Piccart-Gebhart M, Tuohy VK, Romano E, et al. Immunotherapy in breast cancer: an overview of current strategies and perspectives. *NPJ Breast Cancer* 2023;9:7.
[PUBMED](#) | [CROSSREF](#)
142. Humphries MP, Hynes S, Bingham V, Cougot D, James J, Patel-Socha F, et al. Automated tumour recognition and digital pathology scoring unravels new role for PD-L1 in predicting good outcome in ER-/HER2+ breast cancer. *J Oncol* 2018;2018:2937012.
[PUBMED](#) | [CROSSREF](#)
143. Wang X, Wang L, Bu H, Zhang N, Yue M, Jia Z, et al. How can artificial intelligence models assist PD-L1 expression scoring in breast cancer: results of multi-institutional ring studies. *NPJ Breast Cancer* 2021;7:61.
[PUBMED](#) | [CROSSREF](#)
144. Shamaï G, Livne A, Polónia A, Sabo E, Cretu A, Bar-Sela G, et al. Deep learning-based image analysis predicts PD-L1 status from H&E-stained histopathology images in breast cancer. *Nat Commun* 2022;13:6753.
[PUBMED](#) | [CROSSREF](#)
145. Denkert C, Loibl S, Noske A, Roller M, Müller BM, Komor M, et al. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* 2010;28:105-13.
[PUBMED](#) | [CROSSREF](#)
146. de Jong VM, Wang Y, Ter Hoeve ND, Opdam M, Stathonikos N, Józwiak K, et al. Prognostic value of stromal tumor-infiltrating lymphocytes in young, node-negative, triple-negative breast cancer patients who did not receive (neo)adjuvant systemic therapy. *J Clin Oncol* 2022;40:2361-74.
[PUBMED](#) | [CROSSREF](#)
147. Swisher SK, Wu Y, Castaneda CA, Lyons GR, Yang F, Tapia C, et al. Interobserver agreement between pathologists assessing tumor-infiltrating lymphocytes (TILs) in breast cancer using methodology proposed by the international TILs working group. *Ann Surg Oncol* 2016;23:2242-8.
[PUBMED](#) | [CROSSREF](#)
148. Kos Z, Roblin E, Kim RS, Michiels S, Gallas BD, Chen W, et al. Pitfalls in assessing stromal tumor infiltrating lymphocytes (sTILs) in breast cancer. *NPJ Breast Cancer* 2020;6:17.
[PUBMED](#) | [CROSSREF](#)
149. Amgad M, Stovgaard ES, Balslev E, Thagaard J, Chen W, Dudgeon S, et al. Report on computational assessment of tumor infiltrating lymphocytes from the international immuno-oncology biomarker working group. *NPJ Breast Cancer* 2020;6:16.
[PUBMED](#) | [CROSSREF](#)
150. Rasmusson A, Zilenaite D, Nestarenkaite A, Augulis R, Laurinaviciene A, Ostapenko V, et al. Immunogradient indicators for antitumor response assessment by automated tumor-stroma interface zone detection. *Am J Pathol* 2020;190:1309-22.
[PUBMED](#) | [CROSSREF](#)
151. Choi S, Cho SI, Jung W, Lee T, Choi SJ, Song S, et al. Deep learning model improves tumor-infiltrating lymphocyte evaluation and therapeutic response prediction in breast cancer. *NPJ Breast Cancer* 2023;9:71.
[PUBMED](#) | [CROSSREF](#)
152. Shin S, Cho SY, Cho EY, Kim SW, Jung M, Song SG, et al. Artificial intelligence-powered tumor-infiltrating lymphocytes analyzer to reveal distinct immune landscapes in breast cancer by molecular subtype and HER2 score. *J Clin Oncol* 2023;41:1049.
[CROSSREF](#)
153. Albusayli R, Graham JD, Pathmanathan N, Shaban M, Raza SE, Minhas F, et al. Artificial intelligence-based digital scores of stromal tumour-infiltrating lymphocytes and tumour-associated stroma predict disease-specific survival in triple-negative breast cancer. *J Pathol* 2023;260:32-42.
[PUBMED](#) | [CROSSREF](#)

154. Kim S, Kim S, Lim Y, Park G, Song S, Song H, et al. Spatial analysis of tumor-infiltrating lymphocytes (TILs) based on HER2 expression across multiple cancer types. *J Immunother Cancer* 2022;10:A14603.
155. Green J, Reeves GK, Floud S, Barnes I, Cairns BJ, Gathani T, et al. Cohort profile: the million women study. *Int J Epidemiol* 2019;48:28-29e.
[PUBMED](#) | [CROSSREF](#)
156. Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PH, Monninkhof EM, et al. Supplemental MRI screening for women with extremely dense breast tissue. *N Engl J Med* 2019;381:2091-102.
[PUBMED](#) | [CROSSREF](#)
157. Lee HN, Sohn YM, Han KH. Comparison of mammographic density estimation by Volpara software with radiologists' visual assessment: analysis of clinical-radiologic factors affecting discrepancy between them. *Acta Radiol* 2015;56:1061-8.
[PUBMED](#) | [CROSSREF](#)
158. Kim J, Han W, Moon HG, Ahn S, Shin HC, You JM, et al. Breast density change as a predictive surrogate for response to adjuvant endocrine therapy in hormone receptor positive breast cancer. *Breast Cancer Res* 2012;14:R102.
[PUBMED](#) | [CROSSREF](#)
159. Kim G, Bahl M. Assessing risk of breast cancer: a review of risk prediction models. *J Breast Imaging* 2021;3:144-55.
[PUBMED](#) | [CROSSREF](#)
160. Vilmun BM, Vejborg I, Lyng E, Lillholm M, Nielsen M, Nielsen MB, et al. Impact of adding breast density to breast cancer risk models: a systematic review. *Eur J Radiol* 2020;127:109019.
[PUBMED](#) | [CROSSREF](#)
161. Arasu VA, Habel LA, Achacoso NS, Buist DS, Cord JB, Esserman LJ, et al. Comparison of mammography AI algorithms with a clinical risk model for 5-year breast cancer risk prediction: an observational study. *Radiology* 2023;307:e222733.
[PUBMED](#) | [CROSSREF](#)
162. Kim KH, Nam H, Lim E, Ock CY. Development of AI-powered imaging biomarker for breast cancer risk assessment on the basis of mammography alone. *J Clin Oncol* 2021;39:10519.
[CROSSREF](#)
163. Lee H, Kim J, Park E, Kim M, Kim T, Kooi T. Enhancing breast cancer risk prediction by incorporating prior images. 26th International Conference on Medical Image Computing and Computer-Assisted Intervention 2023; 2023 Oct 8-12; Vancouver, BC. Cham: Springer Nature Switzerland; 2023.
[CROSSREF](#)
164. Yala A, Mikhael PG, Strand F, Lin G, Smith K, Wan YL, et al. Toward robust mammography-based models for breast cancer risk. *Sci Transl Med* 2021;13:eaba4373.
[PUBMED](#) | [CROSSREF](#)
165. Wang X, Zou C, Zhang Y, Li X, Wang C, Ke F, et al. Prediction of BRCA gene mutation in breast cancer based on deep learning and histopathology images. *Front Genet* 2021;12:661109.
[PUBMED](#) | [CROSSREF](#)
166. Xu Z, Verma A, Naveed U, Bakhoun SF, Khosravi P, Elemento O. Deep learning predicts chromosomal instability from histopathology images. *iScience* 2021;24:102394.
[PUBMED](#) | [CROSSREF](#)
167. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;24:3726-34.
[PUBMED](#) | [CROSSREF](#)
168. Cho SY, Lee JH, Ryu JM, Lee JE, Cho EY, Ahn CH, et al. Deep learning from HE slides predicts the clinical benefit from adjuvant chemotherapy in hormone receptor-positive breast cancer patients. *Sci Rep* 2021;11:17363.
[PUBMED](#) | [CROSSREF](#)
169. Huynh BQ, Antropova N, Giger ML. Comparison of breast DCE-MRI contrast time points for predicting response to neoadjuvant chemotherapy using deep convolutional neural network features with transfer learning. *Proceedings of the SPIE*; 2017 Feb 11-16; Orlando, FL. Bellingham, WA: SPIE; 2017.
170. Wang H, Li X, Yuan Y, Tong Y, Zhu S, Huang R, et al. Association of machine learning ultrasound radiomics and disease outcome in triple negative breast cancer. *Am J Cancer Res* 2022;12:152-64.
[PUBMED](#)
171. Qian L, Lv Z, Zhang K, Wang K, Zhu Q, Zhou S, et al. Application of deep learning to predict underestimation in ductal carcinoma in situ of the breast with ultrasound. *Ann Transl Med* 2021;9:295.
[PUBMED](#) | [CROSSREF](#)

172. Qu YH, Zhu HT, Cao K, Li XT, Ye M, Sun YS. Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning (DL) method. *Thorac Cancer* 2020;11:651-8.
[PUBMED](#) | [CROSSREF](#)
173. Ha R, Chang P, Mutasa S, Karcich J, Goodman S, Blum E, et al. Convolutional neural network using a breast MRI tumor dataset can predict oncotype Dx recurrence score. *J Magn Reson Imaging* 2019;49:518-24.
[PUBMED](#) | [CROSSREF](#)
174. Cho SY, Cho EY, Paeng K, Jung G, Lee S, Song SY. Deep learning-based predictive biomarker for adjuvant chemotherapy in early-stage hormone receptor-positive breast cancer. *Cancer Res* 2019;79:3144.
[CROSSREF](#)
175. Meti N, Saednia K, Lagree A, Tabbarah S, Mohebpour M, Kiss A, et al. Machine learning frameworks to predict neoadjuvant chemotherapy response in breast cancer using clinical and pathological features. *JCO Clin Cancer Inform* 2021;5:66-80.
[PUBMED](#) | [CROSSREF](#)
176. Krishnamurthy S, Jain P, Tripathy D, Basset R, Randhawa R, Muhammad H, et al. Predicting response of triple-negative breast cancer to neoadjuvant chemotherapy using a deep convolutional neural network-based artificial intelligence tool. *JCO Clin Cancer Inform* 2023;7:e2200181.
[PUBMED](#) | [CROSSREF](#)
177. Huang Z, Shao W, Han Z, Alkashash AM, De la Sancha C, Parwani AV, et al. Artificial intelligence reveals features associated with breast cancer neoadjuvant chemotherapy responses from multi-stain histopathologic images. *NPJ Precis Oncol* 2023;7:14.
[PUBMED](#) | [CROSSREF](#)
178. Sammut SJ, Crispin-Ortuzar M, Chin SE, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature* 2022;601:623-9.
[PUBMED](#) | [CROSSREF](#)
179. U.S. Food and Drug Administration (FDA). Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>. Accessed August 18th, 2023.
180. Srinidhi CL, Ciga O, Martel AL. Deep neural network models for computational histopathology: a survey. *Med Image Anal* 2021;67:101813.
[PUBMED](#) | [CROSSREF](#)
181. Ren J, Hacihaliloglu I, Singer EA, Foran DJ, Qi X. Unsupervised domain adaptation for classification of histopathology whole-slide images. *Front Bioeng Biotechnol* 2019;7:102.
[PUBMED](#) | [CROSSREF](#)
182. Jose L, Liu S, Russo C, Nadort A, Di Ieva A. Generative adversarial networks in digital pathology and histopathological image processing: a review. *J Pathol Inform* 2021;12:43.
[PUBMED](#) | [CROSSREF](#)
183. Seyyed-Kalantari L, Zhang H, McDermott MB, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med* 2021;27:2176-82.
[PUBMED](#) | [CROSSREF](#)
184. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447-53.
[PUBMED](#) | [CROSSREF](#)
185. Gastouniotti A, Kontos D. Is it time to get rid of black boxes and cultivate trust in AI? *Radiol Artif Intell* 2020;2:e200088.
[PUBMED](#) | [CROSSREF](#)
186. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021;3:e745-50.
[PUBMED](#) | [CROSSREF](#)
187. Abràmoff MD, Roehrenbeck C, Trujillo S, Goldstein J, Graves AS, Repka MX, et al. A reimbursement framework for artificial intelligence in healthcare. *NPJ Digit Med* 2022;5:72.
[PUBMED](#) | [CROSSREF](#)