Breast cancer in Western Pacific

MRI-based artificial intelligence models for post-neoadjuvant surgery personalization in breast cancer: a narrative review of evidence from Western Pacific

Yingyi Lin,[a,b,d] Minyi Cheng,[b,c,d] Cangui Wu,[b,c,d] Yuhong Huang,[b] Teng Zhu,[b] Jieqing Li,[b] Hongfei Gao,[b] and Kun Wang[a,b,*]

[a]School of Medicine, South China University of Technology, Guangzhou, Guangdong 510006, China
[b]Department of Breast Cancer, Cancer Centre, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong 510080, China
[c]Southern Medical University, Guangzhou, Guangdong 510080, China

The Lancet Regional Health - Western Pacific 2025;57: 101254

Published Online 6 December 2024
https://doi.org/10.1016/j.lanwpc.2024.101254

Summary

Breast magnetic resonance imaging (MRI) is the most sensitive imaging method for diagnosing breast cancer and assessing treatment response. Artificial intelligence (AI) and radiomics offer new opportunities to identify patterns in imaging data, supporting personalized post-neoadjuvant surgical decisions. This paper reviewed breast MRI-based AI models for predicting outcomes after neoadjuvant therapy, with a focus on evidence from the Western Pacific region, to evaluate the quality of existing models, discuss their inherent limitations, and outline potential future directions. A literature search in MEDLINE, EMBASE, and Web of Science identified 51 relevant studies in the region, with the majority conducted in China, followed by South Korea and Japan. Most studies focused on predicting pathologic complete response (pCR), with a median sample size of 152 and largely retrospective single-center designs. Model performance was commonly assessed using validation sets, with pooled sensitivity and specificity for pCR prediction showing promising results. Models incorporating multitemporal MRI features were associated with improved accuracy. While MRI-based AI models show potential for guiding surgical planning, improved methodological quality and algorithmic explainability are needed to facilitate clinical translation.

Copyright © 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Artificial intelligence; Radiomics; MRI; Breast cancer; Outcome prediction

Introduction

Neoadjuvant therapy is recommended for stage II-III, human epidermal growth factor receptor-2 (HER2) -positive, or triple-negative breast cancer to assess treatment response and enable de-escalation of surgery. Achieving pathologic complete response (pCR) after neoadjuvant therapy is crucial for predicting prognostic outcomes, determining eligibility for breast-conserving surgery (BCS),[1] identifying patients who may potentially forgo surgical intervention altogether.[2] Compared to the United States, a higher proportion of Chinese patients are diagnosed with stage II-III breast cancer, with notable prevalence of HER2-positive and triple-negative subtypes.[3,4] The proportion of patients undergoing neoadjuvant therapy in China has increased,[4] however, the prevalence of BCS in China remained relatively low, ranging from 14.6% to 22.0% in national cross-sectional surveys.[5,6] The BCS rate following neoadjuvant therapy in China was markedly lower compared to Western countries, where over 80% of hospitals performing this procedure in less than 20% of post-neoadjuvant patients.[6] Even among young breast cancer under the age of 35, the majority chose mastectomy over BCS.[7] Concerns about local recurrence, safety of BCS, and the effects of postoperative radiotherapy are major barriers to its adoption in China.[8] Breast magnetic resonance imaging (MRI)-based artificial intelligence (AI) models have shown improved diagnostic accuracy and promise in predicting treatment outcomes and prognosis in breast cancer,[9,10] aiding in preoperative patient selection and surgical planning.[11] A properly validated AI model could potentially increase the prevalence of BCS in the Chinese population by accurately predicting treatment response, enhancing confidence in choosing BCS over mastectomy.

DOIs of original articles: https://doi.org/10.1016/j.lanwpc.2025.101531, https://doi.org/10.1016/j.lanwpc.2025.101538, https://doi.org/10.1016/j.lanwpc.2025.101520, https://doi.org/10.1016/j.lanwpc.2024.101180
*Corresponding author. Department of Breast Cancer, Cancer Center, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong 510080, China.
E-mail address: wangkun@gdph.org.cn (K. Wang).
[d]These authors contributed equally to this work and shared first authorship.

Breast MRI offers extensive pathophysiological information of tumor lesions and is more sensitive for detecting breast cancer than mammography and ultrasonography.[12,13] A meta-analysis showed that MRI outperforms conventional methods in detecting cancer in women with dense breasts.[14] Breast MRI also demonstrated potential for evaluating residual disease after neoadjuvant therapy.[15,16] Although MRI accessibility varies in the Western Pacific region, the gap is gradually narrowing as economic development and healthcare demands increase in countries like China.[17,18] Given the high prevalence of dense breast tissue among Asian women,[19] breast MRI is preferred for early detection and treatment response assessment. However, discrepancies between tumor size assessed by preoperative MRI and postoperative pathology remain a concern. A prospective study found that MRI discrepancies often led to unnecessary mastectomies,[20] underscoring the need for improved diagnostic precision.

AI encompasses algorithms like machine learning and deep learning that perform tasks once achievable only by human intelligence.[21] Radiomics uses quantitative imaging features and machine learning to develop predictive models, while deep learning minimizes human input by using neural network architectures to predict outcomes without domain expertise.[22,23] There is much excitement about the potential of radiomics and deep learning to facilitate personalized medicine. Nonetheless, barriers to clinical application of AI technology remain, including the issue of reproducibility, challenges in model interpretability, and the need for rigorous validation.[24] A systematic review identified a high risk of bias in 72% of AI studies,[25] emphasizing the need for standardized data collection, evaluation criteria, and reporting guidelines. The Radiomics Quality Score (RQS) was developed to assess research quality and support clinical translation of radiomics results.[26]

Given the rising incidence of breast cancer and the unwarranted preference for mastectomy post-neoadjuvant therapy in China, this study aimed to review contemporary research on breast MRI-based AI models for personalizing post-neoadjuvant surgical strategies. The focus was on evidence from the Western Pacific region to reduce ethnicity-related heterogeneity and improve clinical relevance of the findings for Chinese patients. We examined study designs, model development methods, performance metrics, validation strategies, and assessed study quality using RQS. We also addressed the limitations inherent in current studies and proposed directions for future research.

Materials and methods

This study was conducted in accordance with the Preferred Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines. The review protocol was not prospectively registered.

Search strategy and selection criteria

Literature search was performed across three databases, MEDLINE (Ovid), EMBASE, and Web of Science from 2010 to April 22, 2024, using the following key search terms: (breast cancer) AND (neoadjuvant therapy) AND (magnetic resonance imaging OR MRI) AND (radiomics OR machine learning OR deep learning OR artificial intelligence OR texture analysis). Detailed search strategy is presented in Supplementary Table S1. The reference lists from pertinent published reviews were also screened.

Titles and abstracts of the retrieved articles were examined to determine eligibility and were included in the review if they employed radiomics with breast MRI and predicted treatment outcomes with potential implications for surgical planning following neoadjuvant therapy in breast cancer patients. Only original articles published in English were selected. Studies involving patients beyond the western pacific region were excluded. Animal studies, editorials, perspectives, systematic reviews or meta-analyses, book chapters, and conference abstracts were also excluded.

Data extraction

The following data were extracted from each eligible studies using a standardized chart: general publication information (the first author, publication year, country), study designs (study objectives, sample size, data collection strategy, model development and validation), characteristics of imaging modality (MRI sequences, imaging time points), model performance, and metrics for calculating RQS. To evaluate the predictive performance of radiomics and deep learning models, key metrics included the Area Under the Receiver Operating Characteristic Curve (AUC), accuracy, sensitivity, and specificity were extracted from both the training cohorts and the validation or test cohorts in each study. The metrics corresponding to the models that yielded the highest performance at a given level were recorded.

Data synthesis and presentation

Summary figures were generated using Graphpad Prism version 9.5.1 (Boston, MA). A bivariate analysis was performed to obtain a pooled summary of the predictive performance of the included studies if the raw diagnostic data (true positive, false positive, false negative, true negative) could be extrapolated from the studies. Subgroup analyses stratified by imaging timepoint, MRI sequences, integration of clinical features, and sample size were conducted. The "mada" package (v 0.5.11)[27] in R version 4.1.0 (RStudio, Boston, MA) was used for all statistical analysis.

Results

Identification and selection of studies

Literature search yielded 2467 studies. Among these, 938 duplicate studies were eliminated. Screening of titles and

abstracts led to the removal of 986 studies due to irrelevance to our review topic, and an additional 24 studies were excluded as they were reviews, meta-analyses, editorials, or perspectives. The remaining 519 studies were further screened for full text. 209 studies were removed for not employing radiomics analysis for neoadjuvant response prediction, and 176 studies were excluded for applying radiomics analysis to predict outcomes unrelated to neoadjuvant therapy, such as breast cancer detection, molecular subtype characterization, and metastasis prediction. 83 studies that were conducted outside the western pacific were also excluded. The study selection process is depicted in Fig. 1, and a total of 51 studies were included in the present review (Supplementary Table S2).

Study objectives

The majority of the included studies aimed to predict pCR in patients treated with neoadjuvant therapy (n = 36, 70.59%). One study predicted near-pCR,[28] defined as Miller-Payne grade 4 or 5, while another predicted either a complete or partial response following neoadjuvant therapy.[29] The remaining studies focused on determining axillary lymph node response (n = 4, 7.84%),[30–33] tumor regression patterns (n = 4, 7.84%),[34–37] disease-free survival (DFS) (n = 2, 3.92%),[38,39] recurrence rate (n = 2, 3.92%),[40,41] and residual cancer burden (RCB) (n = 1, 1.96%)[42] after neoadjuvant therapy.

Sample size

The sample size of each study ranged from 35 to 1262 participants (Fig. 2a). On average, the studies included 265.1 patients, with a median of 152 and an interquartile range (IQR) of 114–329 patients. Most of the studies had a sample size between 100 and 199 patients (n = 21, 41.18%), while nine studies (17.65%)[36,43–50] recruited less than 100 patients, and four studies (7.84%)[32,42,51,52] had a substantial sample size exceeding 800 patients.

Data collection strategy

Most of the reviewed studies were retrospective in nature (n = 48, 94.12%), with only four studies (7.84%)[31,43,53,54] using prospective datasets for model development but only one studies[43] was registered in a trial database. One study[32] included a prospective dataset for model validation. The majority of the studies recruited patients from a single local institution (n = 41, 80.39%) rather than multiple institutions (n = 10, 19.61%). Most studies were conducted in China (n = 45, 88.24%), and five studies were carried out in South Korea[40,43,55–57] and one in Japan.[49] The mean age of patients enrolled in these studies ranged from 42.9 to 54.8 years.

Imaging modality

Most radiomics analyses were conducted using dynamic contrast-enhanced (DCE) MRI (n = 28, 54.90%), and the rest collected multiparametric MRI data primarily from DCE and diffusion-weighted imaging (DWI) sequences (n = 23, 45.10%) (Fig. 2b). 26 studies (50.98%) extracted imaging features from MRIs performed prior to the initiation of neoadjuvant therapy, while three studies (5.88%) analyzed MRIs acquired after neoadjuvant but before surgery (Fig. 2c). The remaining studies utilized multitemporal MRI data obtained throughout

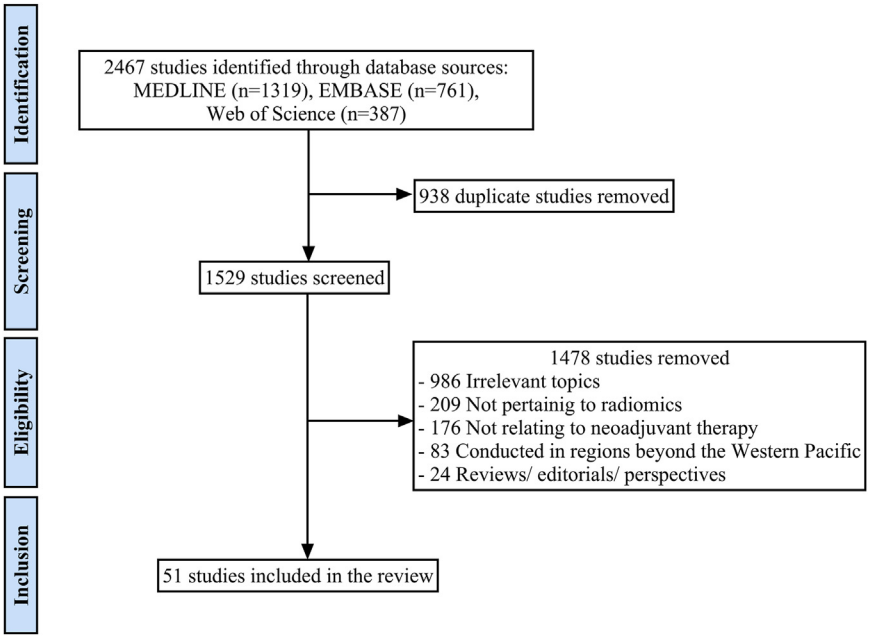

Fig. 1: Study selection process.

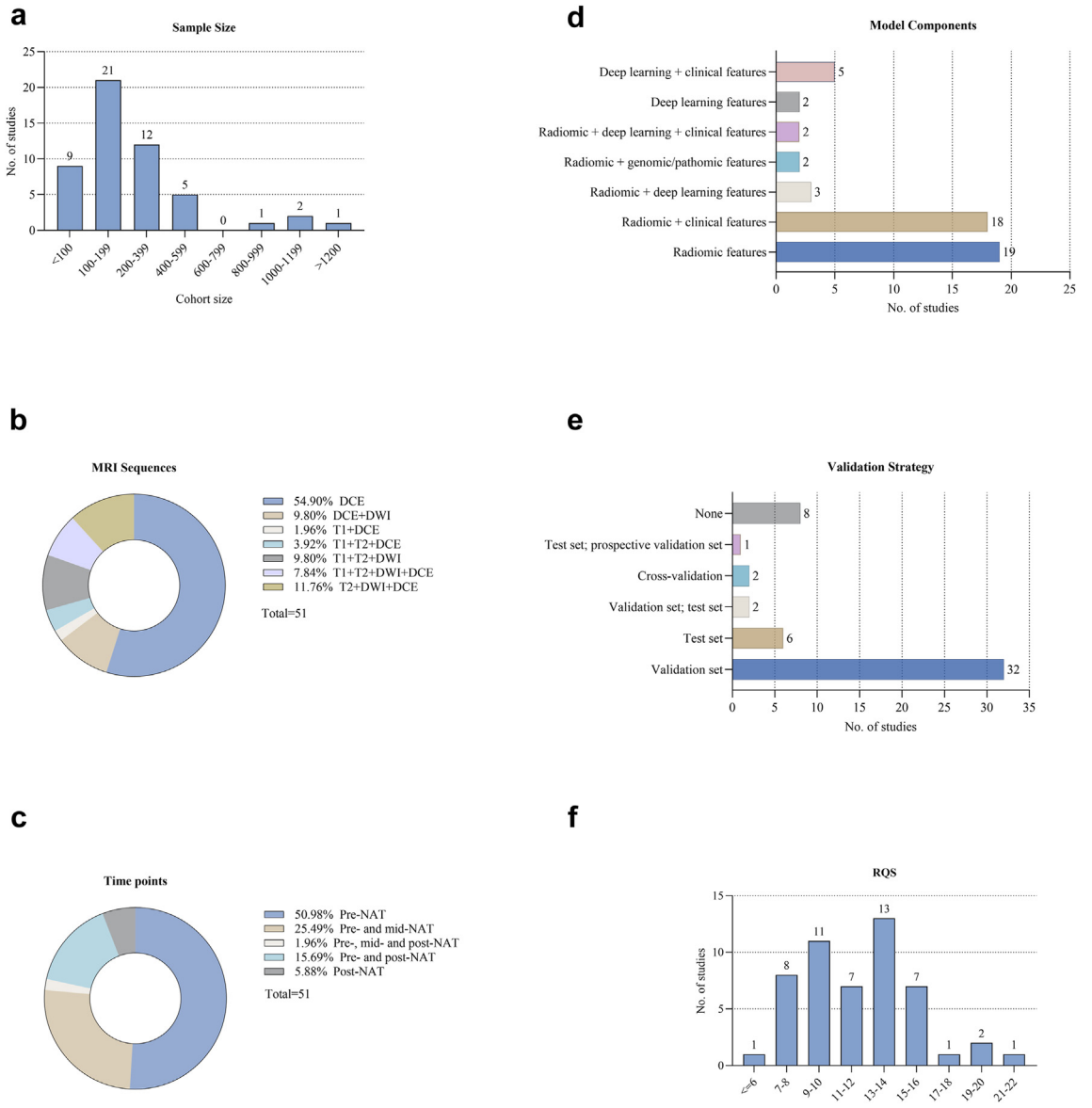

Fig. 2: (a) Distribution of study sample size; (b) MRI sequences involved; (c) MRI imaging time points; (d) Extracted features for model development; (e) Strategy for model performance validation; (f) Distribution of radiomics quality scores. MRI, magnetic resonance imaging; DCE, dynamic contrast-enhanced; DWI, diffusion-weighted imaging; NAT, neoadjuvant therapy; RQS, radiomics quality score.

neoadjuvant treatment (n = 22, 43.14%), with the majority involving MRIs conducted before and after 1–4 cycles of neoadjuvant therapy (n = 13, 25.49%).

Model development and validation

25 studies (49.01%) integrated radiomics and/or deep learning features with clinical features for model development. The most common clinical features selected were significant clinicopathologic predictors of neoadjuvant treatment outcome, including primary tumor size, hormone receptor (HR) status, HER2 status, and Ki-67 levels. Two studies (3.92%)[58,59] explored the use of multi-omics models for pCR prediction by combing radiomics signatures with genomic or pathomic data. The rest of the studies also demonstrated the potential for using exclusively radiomics and/or deep learning features to predict neoadjuvant therapy response (n = 24, 47.06%) (Fig. 2d).

Model validation was omitted in eight studies (15.69%).[40,43,44,46,47,49,55,60] The predominant strategy for assessing model performance was the use of a validation set (n = 32, 62.75%) either by splitting the original study

cohort into a training and a validation set at a certain ratio or by retrospectively recruiting a separate cohort for validation from the same institution with the same inclusion criteria. Six studies (11.76%)[28,42,51,52,61,62] used independent test sets that were enrolled from different institutions (Fig. 2e). Two studies employed a validation set followed by a test set.[63,64] Two studies applied cross-validation on the training set.[35,39] One study conducted model validation using external test sets followed by an evaluation on a prospective validation set.[32]

Performance

36 studies developed radiomics and/or deep learning signatures to predict pCR after neoadjuvant therapy, with an AUC ranging from 0.77 to 0.99 in the training cohorts and from 0.71 to 0.97 in the validation/test cohorts. The pooled sensitivity and specificity for pCR prediction were 0.84 (95% confidence interval [CI] 0.80–0.88) and 0.86 (95% CI 0.83–0.89) in the training cohorts, and 0.78 (95% CI 0.72–0.83) and 0.87 (95% CI 0.83–0.91) in the validation/test cohorts (Fig. 3). Subgroup analyses were conducted to explore relevant factors differentiating model performance. Models utilizing multitemporal MRI features were associated with numerically higher sensitivity (0.83 [95% CI 0.74–0.88] versus 0.76 [95% CI 0.65–0.84] in the validation/test cohorts) and specificity (0.88 [95% CI 0.84–0.92] versus 0.87 [95% CI 0.78–0.93] in the validation/test cohorts) for predicting pCR (Supplementary Fig. S1). Subgroup analyses examining the use of single-parametric versus multiparametric MRIs, the integration of clinical features, and sample size in predictive models did not yield consistent results (Supplementary Figs. S2–S4).

Results from the four studies aiming to predict axillary lymph node response demonstrated that radiomics and deep learning models had a sensitivity of 0.71–0.88, a specificity of 0.80–0.94, and an accuracy of 0.83–0.88 in detecting nodal metastasis after neoadjuvant therapy in the validation/test cohorts.[30–33] Identification of tumor regression patterns can also assist breast surgeons in selecting appropriate surgery for patients receiving neoadjuvant therapy. Four studies reported the capability of radiomics models in differentiating unifocal regression from multiple residual

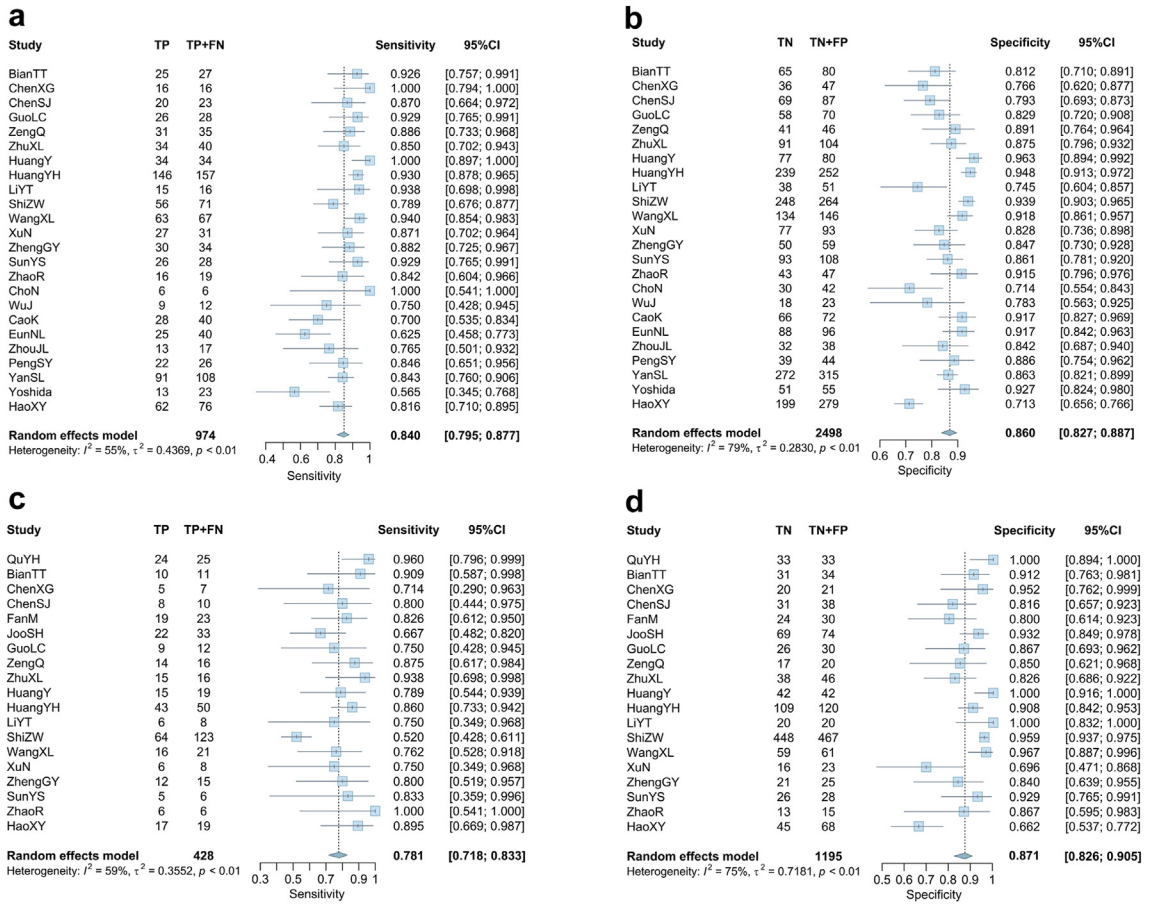

Fig. 3: Forest plot of sensitivity and specificity for pCR prediction in the training cohort (a, b), and the validation/test cohort (c, d). TP, true positive; FN, false negative; FP false positive; TN true negative.

disease, with an overall AUC ranging from 0.83 to 0.94 in the validation/test cohorts.[34–37] Radiomics signatures describing DFS had a respective concordance index of 0.92 and 0.87 in HER2-positive and triple-negative breast cancer treated with neoadjuvant therapy.[38,39] RCB is an important prognostic marker for patients undergoing neoadjuvant therapy. Notably, patients classified with RCB III frequently exhibit treatment resistance, necessitating consideration for early surgical intervention. One study established a multimodal fusion model that effectively identified RCB III cases, achieving an AUC of 0.91–0.94, a sensitivity of 0.77–0.84, and a specificity of 0.94–0.97 in external test cohorts.[42]

RQS

A 16-criterion scoring system was used to assess the quality of radiomics studies from the following five aspects: data selection, medical imaging, feature extraction, exploratory analysis, and modeling.[26] The highest possible score was 36, with higher scores indicating a more rigorous study methodology could facilitate clinical translation of radiomics models. The median RQS observed in the 51 studies was 12 (IQR 9–14). Most studies had an RQS between 9 and 16 (n = 38, 74.51%), and only two studies achieved an RQS of 20 or above (Fig. 2f). Regarding study compliance with individual RQS criteria (Supplementary Fig. S5), the majority of studies scored at least one point in terms of image protocol quality (n = 48, 94.12%), multiple segmentations (n = 43, 84.31%), feature reduction or adjustment (n = 49, 96.08%), discrimination statistics (n = 51, 100.0%), and validation (n = 43, 84.31%), while most studies failed to incorporate multivariable analysis of radiomics with non-radiomics features (n = 48, 94.12%), detect and discuss biological correlates (n = 45, 88.23%), conduct cut-off analyses (n = 41, 90.39%), report calibration statistics (n = 33, 64.71%) or potential clinical utility (n = 30, 58.82%), or implement a prospective study design registered in a trial database (n = 50, 98.04%). None of the included studies performed a phantom study or a cost-effectiveness analysis.

Discussion

Breast MRI-based AI models hold great promise in facilitating personalized surgical decisions for patients treated with neoadjuvant therapy. In the present review, several studies have successfully trained models to predict specific outcomes essential for determining appropriate surgical strategies following neoadjuvant therapy, including the prediction of pCR, axillary lymph node response, tumor regression patterns, RCB, and prognosis. Nonetheless, common methodological weaknesses need to be acknowledged and addressed to provide solid evidence for the application of AI technology in clinical settings.

Ethnicity and AI models

AI is rapidly emerging in oncology for its unparalleled ability to merge and capitalize on clinical, radiologic, pathologic, and molecular data to uncover intricate and interrelated patterns underlying the pathophysiological behaviors of cancer that were once beyond the scope of human computational capabilities. Literature reviews have highlighted the potential of radiomics and deep learning models in breast cancer diagnosis, molecular subtype classification, and treatment response prediction.[9,10,65,66] A recent study found that AI models accurately identified an individual's self-reported race from medical images with surprising precision.[67] Researchers were unable to pinpoint specific image-based factors accounting for this accuracy, as the models persistently showed robust performance even when analyzing highly degraded images that were indiscernible to human specialists, underscoring the fact that AI is not inherently impartial to race. If a model can detect a patient's race, it may utilize this information in predicting other medical outcomes and inadvertently reinforce existing biases against racial and ethnic minorities in healthcare, resulting in underdiagnosis or less frequent treatment recommendations.[68,69]

Racial and ethnic data have been significantly underreported in the development of predictive tools. Among the majority of AI products approved by the United States FDA, only a small number have disclosed the racial demographics of their study cohorts,[70] which could lead to unequal benefit across diverse populations.[71,72] A study that developed a radiomics model based on German patients, 88.6% of whom (581 of 656) were white people, using pretreatment MRI features to predict pCR found that validation in an American cohort showed performance comparable to the development cohort, while validation in a Chinese cohort revealed a significant performance drop, suggesting the presence of racial bias in radiomics-based models.[73] The collection and reporting of racial and ethnic data are essential for enabling stratified analyses that can reveal subgroup trends potentially obscured in the general population.[74] Currently, U.S. and Chinese datasets and authors are disproportionately represented in AI studies.[75] Similarly, our literature search did not identify studies outside of China, Japan, and South Korea investigating the application of AI models based on breast MRI to predict neoadjuvant response. This aligns with the observation that AI models, particularly those developed for clinical settings, are predominantly based on datasets from high-income countries.[75] To promote equitability in AI, there should be increased investment in technological infrastructure in underrepresented regions, enhanced external validation of AI models, and recalibration of models for diverse populations.[76] Standard practices should incorporate strategies to detect racial bias and develop models intentionally designed to balance outcomes across racial groups. Another solution is to shift

from the narrow emphasis on model generalizability to employing algorithms tailored for distinct subpopulations.[77] Therefore, our review selectively incorporated studies from the Western Pacific region, predominantly comprising Asian populations, to enhance the precision and clinical relevance of our findings.

MRI-based AI models in Western Pacific

Developments and implications in clinical care

Neoadjuvant therapy can downstage locally advanced inoperable breast cancer to increase the feasibility of breast-conserving surgery. However, a meta-analysis of individual patient data comparing long-term outcomes of neoadjuvant and adjuvant therapies suggested that tumors downsized by neoadjuvant therapy might be associated with higher local recurrence after breast-conserving treatment compared to similarly sized tumors in patients who underwent adjuvant therapy.[78] Careful preoperative tumor localization and treatment response assessment are essential for reducing the risk of local recurrence and ensuring the safe practice of breast-conserving therapy. Prior studies investigating the viability of breast-conserving therapy with preoperative image-guided biopsy failed to achieve prespecified outcomes and do not endorse the approach of omitting surgery for patients with excellent response following neoadjuvant therapy, underling the necessity of methodological improvements in radiological evaluation and analysis.[79,80]

In our review, breast MRI-based AI models have demonstrated preliminary success in predicting various outcomes following neoadjuvant therapy that have significant implications for surgical decision-making. These models effectively identified patients achieving pCR, with a pooled sensitivity of 0.78 (95% CI 0.72–0.83) and specificity of 0.87 (95% CI 0.83–0.91) across validation or external test cohorts. Several models also exhibited robust predictive performance across different molecular subtypes of breast cancer.[51,62] For axillary lymph node response, AI models achieved sensitivity between 0.71 and 0.88, specificity between 0.80 and 0.94, and accuracy between 0.83 and 0.88 in detecting nodal metastasis after neoadjuvant therapy within validation/test cohorts. One study further explored clinical utility by integrating AI model with standard sentinel lymph node biopsy protocols and found that AI application could significantly reduce the false-negative rate of standard biopsy from 10.12% to 4.76%.[32] A recent study introduced an innovative machine learning model that combined patient, imaging, tumor, and biopsy data, achieving a 0% (95% CI 0–13.7%) false-negative rate for identifying pCR, making patients eligible to omit breast and axilla surgery.[81] These studies highlight AI's potential to refine surgical strategies based on individual responses, paving the way for clinical trials that employ AI to identify candidates for de-escalated breast and axillary surgeries.

Application of AI has progressed beyond conventional imaging to the development of multi-omics models, showing preliminary success in predicting response to neoadjuvant therapy in breast cancer. For instance, an MRI-based radiogenomic signature developed to predict pCR in triple-negative breast cancer demonstrated higher predictive accuracy than a radiomics-only model, achieving an AUC of 0.87 during validation.[58] Similarly, a study combining multiparametric MRI-based radiomics with pathomics features found that the radiopathomics model outperformed models based on radiomics or pathomics alone in predicting pCR in breast cancer.[59] Beyond predicting treatment response and breast surgical outcomes, AI models can also play a valuable role in preoperative and postoperative management in breast cancer, supporting more individualized treatment approaches.[82] Preoperatively, AI can analyze diverse data sources to aid surgical planning by reconstructing critical anatomical structures, helping surgeons optimize techniques and minimize surgical risks. Postoperatively, AI is also valuable for monitoring recovery and predicting complications. For example, AI models can assess the risk of lymphedema after breast surgery by analyzing symptoms reported by patients,[83] enabling early interventions to mitigate this condition. Furthermore, predictive models can help identify patients at higher risk for complications like persistent postoperative pain,[84] allowing for tailored follow-up care and preventive measures. These applications exemplify how AI could enhance surgical care by providing insights that lead to personalized treatment strategies, ultimately improving recovery experiences and outcomes for patients.

Potential factors associated with model performance

Despite the promising accuracy of AI models in predicting pCR, substantial heterogeneity was observed in the pooled analysis. Subgroup analyses were conducted to identify potential factors contributing to this heterogeneity. Notably, the incorporation of multitemporal MRI features significantly reduced model heterogeneity and improved sensitivity and specificity to 0.83 (95% CI 0.74–0.88) and 0.88 (95% CI 0.84–0.92), respectively, in the validation/test cohort (Supplementary Fig. S1c and d). Neoadjuvant therapy can induce various histopathological changes in tumor cellularity and vascular density. It has been demonstrated that MRI-based tumor response patterns halfway through neoadjuvant therapy predicted pCR more accurately than those observed after the completion of therapy.[85] Delta-radiomics, which characterizes the evolution of imaging features by applying radiomics to multiple treatment time points,[86] effectively reflects the dynamic changes in tumor microstructures throughout the treatment process and has been proven to be less susceptible to scanner differences and exhibit a more robust performance compared to

analyses at a single time point.[87] Consistent with our findings, several studies have demonstrated the enhanced performance of MRI models incorporating delta radiomics features to predict pCR compared to models developed solely from single timepoint features.[51,54,63,88,89] In addition, feature importance analysis suggested that delta radiomics played a more substantial role in model development than other components,[51] underscoring the importance of integrating multitemporal MRI features to reduce potential inter-subject heterogeneity and to maintain model robustness across various validation settings.

Subgroup analyses stratified by imaging sequences and the inclusion of clinicopathologic features during model development did not identify them as primary sources of heterogeneity. Multiparametric MRI may enhance presurgical evaluation and treatment monitoring for breast cancer by combining information from different imaging sequences that assess tumor lesion from different aspects.[12] While certain studies have indicated higher predictive accuracy with radiomics derived from multiparametric MRI,[37,45,56] the benefit of using such data remains unproven in clinical practice.[9] In our review, there was no substantial difference between the performance of single parametric and multiparametric models, which could be attributed to the overriding impact of other critical elements within the radiomics workflow, such as image processing and model architecture. Additionally, combining radiomics signatures with clinicopathologic features did not consistently improve predictive accuracy for pCR in our analysis, potentially due to variations in feature reduction strategies and the application of different machine learning or deep learning algorithms. There is also a risk of overestimating the performance of combined models, as most studies lacked multivariable analyses delineate the correlations between radiomics and non-radiomics features. Heterogeneity among radiomics studies may arise from various factors, including study designs, imaging protocols, and model development approaches; our analysis only begins to explore the potential sources of this heterogeneity. Future studies should endeavor to comply with established protocols, such as those outlined by the Radiomics Quality Score,[26] to promote standardization in study procedure and improve reliability of AI models.

The accuracy of breast MRI for predicting pCR highly depends on molecular subtype due to inherent biological differences and the corresponding sensitivity to neoadjuvant regimens. HR-positive/HER2-negative breast cancer exhibits a significantly lower pCR rate compared to HER2-positive and triple-negative breast cancers, which may translate to lower sensitivity in predicting pCR for HR-positive/HER2-negative cases in conventional predictive morels.[90] However, in our review, the performance for predicting treatment response was comparable across different breast cancer subtypes. The respective AUC for pCR prediction in HER2-positive breast cancer, triple-negative breast cancer, and HR-positive/HER2-negative breast cancer was 0.81–0.93, 0.84–1.00, and 0.88–0.91 in the validation/test cohorts. These findings suggest that the application of radiomics and deep learning analysis to conventional imaging techniques could potentially reconcile the biological and treatment-related nuances of each breast cancer subtype. Nonetheless, it is premature to conclude that AI models possess robust predictive abilities across all breast cancer subtypes, given that only five of the included studies evaluated model performance for each subtype.[32,35,51,62,91] Future research should disclose detailed performance metrics across different breast cancer subtypes, or alternatively, focus on the development of AI models tailored for a specific subtype to enhance their robustness and clinical utility.

Study quality and major methodological weaknesses

Most studies did not meet the RQS criteria, with a median RQS of only 12 (IQR 9–14). This trend is widespread and not limited to studies from the Western Pacific region. A previous meta-analysis of 77 studies reported insufficient overall scientific quality and reporting in radiomics research, with a mean RQS of just 9.4 out of 36.[92] Consistent with our findings, researchers identified low scores in areas such as clinical utility demonstration, biological validation, prospective study design, and open science practices.[92] While RQS serves as a general guideline for conducting radiomics research, it highlights crucial areas in need of improvement to enhance study quality and the reliability of AI models. Notably, four studies with the highest RQS gained additional points by validating across three or more datasets,[32,42,51] demonstrating clinical utility through decision curve analysis,[32,42,51,52] identifying biological correlates,[52] and making source code openly accessible.[52]

A major challenge affecting the reliability of AI models is the reliance on small, single-center, retrospective datasets. The median sample size across the included studies was 152 (IQR 114–329), with only nine studies (17.65%) incorporating external test datasets, thereby increasing the potential risk of model overfitting. Overfitting occurs when a model learns both the underlying patterns and the random noise within the training data, resulting in poor performance on new, unseen datasets. This phenomenon is particularly prevalent in radiomics research when dealing with high-dimensional data, limited sample sizes, and highly complex models.[93] Internal validation methods, such as cross-validation and bootstrapping, should be routinely employed for preliminary performance evaluation and for fine-tuning model development to mitigate overfitting. More importantly, AI models require external validation in clinical cohorts that adequately represent the target patient population to simulate real-world

clinical settings effectively; the use of prospectively collected data is also preferable.[94] Ultimately, clinical validation of an AI model necessitates demonstrating its value through randomized controlled trials, wherein patients are randomized to receive care either with or without the AI tool under investigation.[94,95]

Discrimination and calibration represent distinct aspects of AI model performance. Our review found that while AUC metrics were consistently recorded in studies, there was a lack of calibration statistics and assessments of potential clinical utility. AUC alone may not fully capture clinical relevance; incorporating additional discrimination metrics such as sensitivity, specificity, and positive and negative predictive values is crucial for a more comprehensive evaluation of model performance.[96] Furthermore, calibration statistics and decision curve analysis should also be included to ascertain the clinical applicability of the AI models. Calibration measures the agreement between predicted probabilities and observed outcomes at an individual level and is essential for assessing the reliability of predictive models.[26] Evaluating the potential clinical utility of AI models through methods like decision curve analysis effectively bridges the gap between traditional performance metrics and clinical relevance by quantifying the net model benefit across different risk thresholds.[97] Incorporating calibration and clinical utility analysis into model performance evaluation is essential to enhance the credibility of AI models and expedite their acceptance in clinical practice.

Investigation into the biological context of the imaging features should also be included in radiomics research either as part of model development or subsequent validation. Biological validation prevents overfitting by grounding radiomics features in relevant biology, ensuring that model predictions are align with actual pathophysiologic behaviors. Integrating radiomic analysis with different biological correlates can offer insights into the biological underpinnings of radiomics, facilitating a more comprehensive evaluation of therapeutic response and cancer prognosis.[98,99] Genomic analysis is often used to correlate radiomic features with gene expression profiles, as demonstrated in study that linked radiomic markers of hypoxia with hypoxia-related genes in glioblastoma, where hypoxia-associated imaging features predicted patient survival.[100] This approach substantiated the role of hypoxia imaging phenotypes as indirect survival predictors by capturing tumor biology linked to hypoxia gene expression. Histopathology and immunohistochemistry provide direct methods for correlating radiomic features with cellular and molecular characteristics visible in tissue samples, offering insights into how specific radiomic features relate to tumor cell composition, immune infiltration, and structural organization. For instance, the texture features in computerized tomography (CT) imaging of non-small cell lung cancer have been correlated with known hypoxia markers such as carbonic anhydrase IX,[101] suggesting that radiomic features may serve as surrogates for tumor hypoxia, a crucial prognostic factor. Habitat imaging is another approach that segments a tumor into distinct "habitats" based on imaging features indicative of different microenvironments, such as necrotic, hypoxic, or proliferative areas, and utilizes multiple imaging modalities to create a spatial map of the tumor's physiological diversity.[102] In a recent breast cancer study, researchers segmented tumors into subregions based on conventional MRI radiomics features and applied a Gaussian mixture model to define different habitat features to represent intratumoral heterogeneity. By integrating habitat features with radiomics and clinicopathologic variables, the model demonstrated strong predictive performance for treatment response following neoadjuvant therapy.[52]

Feature robustness is another crucial element in the development of reliable AI models. The reproducibility of features is extremely sensitive to acquisition settings; even images of the same tissue site can differ due to variations in their acquisition parameters.[103] Most of the included studies employed multiple segmentation methods and various feature selection strategies to identify reproducible features for model construction. Another approach to enhancing feature reproducibility in radiomics involves applying image preprocessing techniques, such as standardizing voxel size, intensity normalization, and resampling, to minimize discrepancies in image acquisition across different sites and scanners. The ComBat harmonization method can effectively reduce scanner-related effects and improve feature reproducibility, making it suitable for multicenter research.[104] Researchers have applied ComBat to "phantom images" in CT scans and found that it successfully realigned radiomic feature distributions across multi-institutional datasets with varying CT protocols.[105] Additionally, other statistical approaches incorporating stability measures can also enhance the reliability of AI models. For example, researchers analyzed CT scans from multiple cohorts and selected features that were both stable across different scanners and settings and highly discriminative of recurrence.[106] A model that integrated stability measures with discriminative features demonstrated superior accuracy in predicting disease recurrence in early-stage non-small cell lung cancer compared to conventional radiomics models.

Another concern highlighted in our review is the generally poor explainability and interpretability of the AI models. Explainability pertains to the implementation of transparent and traceable statistical black-box machine learning methods, especially with deep learning, and addressed the rationale behind why predictions are made and how model parameters capture underlying biological mechanisms.[107] Interpretability offers a more generalized understanding of the model development without delving into intricate details and

can be seen as a broader component of explainability.[108] Researchers have thus developed methods to increase the interpretability of radiomics features and models. At the feature level, exploring the association between features and tumor heterogeneity can increase interpretability. At the model level, diverse technologies based on local and global interpretation can be applied to improve the interpretability of AI models, including LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive explanation), PDP (Partial Dependence Plot), and decision tree.[109] Recent advancements in interpretable AI have successfully identified patient subgroups that did not benefit from adjuvant imatinib therapy in the context of gastrointestinal stromal tumors,[110] setting a precedent for future research in developing interpretable AI models.

Comparisons with notable studies

Research in the Western Pacific region on MRI-based AI models for predicting breast cancer treatment outcomes has produced results comparable to those from other regions,[111,112] indicating that the development pipelines for radiomics and deep learning models are well-established and applicable to diverse cohorts. Incorporating insights from state-of-the-art research in other regions, such as refining study designs, creating collaborative platforms for model building, and integrating human elements into the development process, is essential for optimizing these models for clinical use.

The majority of reviewed studies have been retrospective, leaving substantial uncertainty regarding the real-world performance of AI models. A scoping review of randomized controlled trials on AI in clinical settings found that, although countries in the Western Pacific region led in the number of trials conducted, most focused on gastroenterology, with no oncology-related studies reported, contrasting with the diverse specialties covered in U.S. trials.[113] For example, the SHIELD-RT trial investigated the role of machine learning in reducing acute care visits during outpatient radiotherapy and chemoradiation.[114] In this trial, 311 high-risk treatment courses identified by the algorithm were randomized to either standard once-weekly clinical evaluations or mandatory twice-weekly evaluations. Results showed that twice-weekly evaluations significantly reduced acute care visits from 22.3% to 12.3%. Economic analysis further revealed that additional evaluations for high-risk patients, as identified by machine learning, not only reduced overall healthcare costs but also improved clinical outcomes,[115] highlighting the potential of AI to enhance patient care and reduce expenses through accurate triage and intervention. Future research could explore AI's capacity to support preoperative planning in breast cancer by prospectively stratifying patients based on treatment outcomes predicted by AI algorithms.

The prevalence of single-center study in the Western Pacific region also raises concerns about the generalizability and applicability of the resulting models. Training AI models on datasets that closely simulate the clinical settings of the target population is essential to enhance their accuracy and reliability. The OPTIMAM Mammography Image Database, a centralized and fully annotated dataset, exemplifies this approach by including mammography images and clinical data from various UK breast screening centers.[116] This database is regularly updated with images from different screening episodes, supporting the development and evaluation of AI algorithms for breast cancer detection and risk assessment.[117,118] Currently, there are few publicly available multi-institutional breast MRI datasets. In addition, public databases introduce challenges related to patient privacy, data security, and potential misuse. Federated learning (FL) is a decentralized machine learning approach that enables collaborative model training across multiple institutions without transferring sensitive data to a central server.[119] This approach is particularly advantageous in healthcare and radiomics, where patient privacy and data security are paramount. In FL, individual nodes (such as hospitals) locally train a model on their own datasets, and only the model updates—not the data itself—are shared with a central server. The server aggregates these updates to improve the model iteratively, preserving privacy while allowing access to a diverse data pool. This structure not only enhances data security but also enables research institutions to collaborate on creating robust predictive models, especially beneficial in radiomics where high-quality, multicenter data is crucial for generalizable insights. An increasing number of researchers are exploring AI models trained through FL for disease prediction while maintaining data confidentiality.[120] The GenoMed4All project exemplifies FL's utility by connecting data across European clinical sites to predict outcomes for rare diseases, enabling extensive model training while complying with data protection regulations. Such federated platforms can enhance the reliability of radiomic models, as diverse data from various sites can improve model robustness.[121] In breast cancer research, a memory-aware curriculum FL method was applied for breast cancer classification using mammography data from diverse clinical sites with varying imaging protocols. The proposed method addressed key challenges in FL, such as system heterogeneity and domain adaptation, by leveraging incorporating unsupervised domain adaptation to align feature distributions across different imaging domains and using curriculum learning to improve model consistency. The study achieved notable improvements in classification accuracy, highlighting the potential of FL to enable large-scale, privacy-preserving collaborations in breast cancer research and contribute to the development of more robust and generalizable AI models.[122]

Most of the included studies reported the performance of AI models as a standalone system. However, AI remains a supplement to, rather than a replacement for, human expertise. Prior research has shown that standalone AI performance may not surpass the expertise of human professionals.[113] A decision-referral approach was proposed to leverage the strengths of both the radiologist and AI into breast cancer screening,[123] allowing AI to automatically handle high-certainty assessments while referring uncertain cases to radiologists. This approach demonstrated improved sensitivity and specificity compared to either standalone AI or radiologists alone, suggesting it could enhance screening accuracy and reduce radiologist workload.[123] A three-phase, prospective study further evaluated an AI-assisted additional-reader approach for early breast cancer detection, revealing that the AI-enhanced workflow could identify an additional 0.7 to 1.6 cases of breast cancer per 1000 examinations, with most detected cancers being invasive (83.3%) and small (≤10 mm, 47.0%).[124] Incorporating human judgment into the development of AI models bolsters patient trust in AI-driven decisions.[125] This synergy between human expertise and AI's computational power is essential for the responsible and effective deployment of AI technologies.

Advances in data science and AI have led to the development of large AI models with sophisticated machine learning architectures characterized by immense scale, both in parameters and training datasets.[126] These models use deep learning techniques, like the Transformer architecture, to capture long-range dependencies. They are pre-trained on massive datasets in a self-supervised manner to build generalized representations, followed by fine-tuning for specific applications. Unlike conventional models, large AI models excel at generalizing across diverse tasks and synthesizing multimodal data, including text, images, and audio. This versatility makes them valuable in healthcare, where integrated, context-aware analyses are crucial.[127] For example, CheXzero, a zero-shot chest X-ray classifier, has achieved radiologist-level performance without prior exposure to specific disease labels.[128] ChatCAD combines diagnostic networks with ChatGPT for medical image analysis, where specialized networks perform initial analysis, and ChatGPT interprets and provides recommendations, and a follow-up iteration called ChatCAD + further enhanced the factual grounding of the model's outputs by incorporating a retrieval system to improve the quality of the generated diagnostic reports.[129,130] These models have shown potential to significantly advance the prediction of treatment response in oncology by analyzing imaging data, genomic information, and clinical records to identify patterns that are predictive of patient outcomes to support personalized treatment strategies and optimize therapy selection.

Limitations

There are limitations to our study. First, none of the reviewed studies included a detailed description of the ethnic distribution of participants. This lack of data prevents meaningful comparisons of AI model performance across different ethnic groups, particularly among minority populations. Given that ethnicity may impact the effectiveness and generalizability of AI models in clinical practice, future research in Western Pacific should emphasize collecting and reporting on ethnicity distribution to better understand the model's applicability across diverse demographics and ensure equitable healthcare outcomes. Second, we included only English-language publications in our review, which may have led to the exclusion of some regional studies from the analysis.

Conclusions

AI models based on breast MRI have the potential to promote personalized surgical decisions by increasing the prevalence of BCS and decreasing unnecessary mastectomies for patients undergoing neoadjuvant therapy. In the future, AI-powered clinical decision support systems could assist clinicians in distinguishing treatment responders from non-responders and in determining optimal surgical timing and strategy on an individualized basis. However, implementing AI models in clinical practice requires further evaluation, as current evidence is susceptible to overfitting and lacks explainability and interpretability. Future research should adhere to rigorous methodological standards to ensure model validity and reproducibility to facilitate the translation of theoretical AI algorithms into practical clinical tools.

Contributors

KW and YL conceptualized and designed the study. YH provided critical suggestions for the study protocol. YL, MC, and CW conducted literature search and study screening. YL, MC, CW, TZ, JL, and HG participated in data extraction. YL conducted data analysis and drafted the manuscript. KW, MC, and CW participated in the critical revision of the manuscript. All authors read and approved of the final manuscript.

Declaration of interests

All authors declare that there is no conflict of interest.

Acknowledgements

The study was supported by the National Natural Science Foundation of China (82171898); the High-level Hospital Construction Project of Guangdong Provincial People's Hospital (DFJH202109). Funding sources were not involved in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. AI-assisted technologies were used only to enhance readability and language quality of the work, under human oversight and control.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanwpc.2024.101254.

References

1 Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet*. 2014;384(9938):164–172.
2 Kuerer HM, Smith BD, Krishnamurthy S, et al. Eliminating breast surgery for invasive breast cancer in exceptional responders to neoadjuvant systemic therapy: a multicentre, single-arm, phase 2 trial. *Lancet Oncol*. 2022;23(12):1517–1524.
3 He S, Xia C, Li H, et al. Cancer profiles in China and comparisons with the USA: a comprehensive analysis in the incidence, mortality, survival, staging, and attribution to risk factors. *Sci China Life Sci*. 2024;67(1):122–131.
4 Li J, Zhou J, Wang H, et al. Trends in disparities and transitions of treatment in patients with early breast cancer in China and the US, 2011 to 2021. *JAMA Netw Open*. 2023;6(6):e2321388.
5 Yu LX, Shi P, Tian XS, Yu ZG, Chinese Society of Breast S. A multi-center investigation of breast-conserving surgery based on data from the Chinese Society of Breast Surgery (CSBrS-005). *Chin Med J (Engl)*. 2020;133(22):2660–2664.
6 Yang B, Ren G, Song E, et al. Current status and factors influencing surgical options for breast cancer in China: a nationwide cross-sectional survey of 110 hospitals. *Oncol*. 2020;25(10):e1473–e1480.
7 Wang X, Xia C, Wang Y, et al. Landscape of young breast cancer under 35 years in China over the past decades: a multicentre retrospective cohort study (YBCC-Catts study). *EClinicalMedicine*. 2023;64:102243.
8 Zhang X, Wang Y. A survey of current surgical treatment of early stage breast cancer in China. *Oncoscience*. 2018;5(7–8):239–247.
9 Adam R, Dell'Aquila K, Hodges L, Maldjian T, Duong TQ. Deep learning applications to breast cancer detection by magnetic resonance imaging: a literature review. *Breast Cancer Res*. 2023;25(1):87.
10 Campana A, Gandomkar Z, Giannotti N, Reed W. The use of radiomics in magnetic resonance imaging for the pre-treatment characterisation of breast cancers: a scoping review. *J Med Radiat Sci*. 2023;70(4):462–478.
11 Varghese C, Harrison EM, O'Grady G, Topol EJ. Artificial intelligence in surgery. *Nat Med*. 2024;30(5):1257–1268.
12 Kataoka M, Iima M, Miyake KK, Honda M. Multiparametric approach to breast cancer with emphasis on magnetic resonance imaging in the era of personalized breast cancer treatment. *Invest Radiol*. 2024;59(1):26–37.
13 Morrow M, Waters J, Morris E. MRI for breast cancer screening, diagnosis, and treatment. *Lancet*. 2011;378(9805):1804–1811.
14 Hussein H, Abbas E, Keshavarzi S, et al. Supplemental breast cancer screening in women with dense breasts and negative mammography: a systematic review and meta-analysis. *Radiology*. 2023;306(3):e221785.
15 Lobbes MBI, Prevos R, Smidt M, et al. The role of magnetic resonance imaging in assessing residual disease and pathologic complete response in breast cancer patients receiving neoadjuvant chemotherapy: a systematic review. *Insights Imaging*. 2013;4(2):163–175.
16 Marinovich ML, Houssami N, Macaskill P, et al. Meta-analysis of magnetic resonance imaging in detecting residual breast cancer after neoadjuvant therapy. *J Natl Cancer Inst*. 2013;105(5):321–333.
17 Geethanath S, Vaughan JT. Accessible magnetic resonance imaging: a review. *J Magn Reson Imag*. 2019;49(7):e65–e77.
18 He L, Yu H, Shi L, et al. Equity assessment of the distribution of CT and MRI scanners in China: a panel data analysis. *Int J Equity Health*. 2018;17(1):157.
19 del Carmen MG, Halpern EF, Kopans DB, et al. Mammographic breast density and race. *Am J Roentgenol*. 2007;188(4):1147–1150.
20 Han Y, Jung JG, Kim J-I, et al. The percentage of unnecessary mastectomy due to false size prediction using preoperative ultrasonography and MRI in breast cancer patients who underwent neoadjuvant chemotherapy: a prospective cohort study. *Int J Surg*. 2023;109(12):3993–3999.
21 Dercle L, McGale J, Sun S, et al. Artificial intelligence and radiomics: fundamentals, applications, and challenges in immunotherapy. *J Immunother Cancer*. 2022;10(9):e005292.
22 Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–577.
23 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.
24 Limkin EJ, Sun R, Dercle L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol*. 2017;28(6):1191–1206.
25 Corti C, Cobanaj M, Marian F, et al. Artificial intelligence for prediction of treatment outcomes in breast cancer: systematic review of design, reporting standards, and bias. *Cancer Treat Rev*. 2022;108:102410.
26 Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749–762.
27 Shim SR, Kim SJ, Lee J. Diagnostic test accuracy: application and practice using R software. *Epidemiol Health*. 2019;41:e2019007.
28 Fan M, Cui Y, You C, et al. Radiogenomic signatures of oncotype DX recurrence score enable prediction of survival in estrogen receptor-positive breast cancer: a multicohort study. *Radiology*. 2022;302(3):516–524.
29 Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L. Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. *Eur J Radiol*. 2017;94:140–147.
30 Gan L, Ma M, Liu Y, et al. A clinical–radiomics model for predicting axillary pathologic complete response in breast cancer with axillary lymph node metastases. *Front Oncol*. 2021;11:786346.
31 Liu S, Du S, Gao S, Teng Y, Jin F, Zhang L. A delta-radiomic lymph node model using dynamic contrast enhanced MRI for the early prediction of axillary response after neoadjuvant chemotherapy in breast cancer patients. *BMC Cancer*. 2023;23(1):15.
32 Zhu T, Huang Y-H, Li W, et al. Multifactor artificial intelligence model assists axillary lymph node surgery in breast cancer after neoadjuvant chemotherapy: multicenter retrospective cohort study. *Int J Surg*. 2023;109(11):3383–3394.
33 Zhang B, Yu Y, Mao Y, et al. Development of MRI-based deep learning signature for prediction of axillary response after NAC in breast cancer. *Acad Radiol*. 2024;31(3):800–811.
34 Zhuang X, Chen C, Liu Z, et al. Multiparametric MRI-based radiomics analysis for the prediction of breast tumor regression patterns after neoadjuvant chemotherapy. *Transl Oncol*. 2020;13(11):100831.
35 Huang Y, Chen W, Zhang X, et al. Prediction of tumor shrinkage pattern to neoadjuvant chemotherapy using a multiparametric MRI-based machine learning model in patients with breast cancer. *Front Bioeng Biotechnol*. 2021;9:662749.
36 Chen Z, Huang M, Lyu J, Qi X, He F, Li X. Machine learning for predicting breast-conserving surgery candidates after neoadjuvant chemotherapy based on DCE-MRI. *Front Oncol*. 2023;13:1174843.
37 Fan M, Wu X, Yu J, et al. Multiparametric MRI radiomics fusion for predicting the response and shrinkage pattern to neoadjuvant chemotherapy in breast cancer. *Front Oncol*. 2023;13:1057841.
38 Li Q, Xiao Q, Li J, Duan S, Wang H, Gu Y. MRI-based radiomic signature as a prognostic biomarker for HER2-positive invasive breast cancer treated with NAC. *Cancer Manag Res*. 2020;12:10603–10613.
39 Xia B, Wang H, Wang Z, et al. A combined nomogram model to predict disease-free survival in triple-negative breast cancer patients with neoadjuvant chemotherapy. *Front Genet*. 2021;12:783513.
40 Eun NL, Kang D, Son EJ, Youk JH, Kim J-A, Gweon HM. Texture analysis using machine learning-based 3-T magnetic resonance imaging for predicting recurrence in breast cancer patients treated with neoadjuvant chemotherapy. *Eur Radiol*. 2021;31(9):6916–6928.
41 Ma M, Gan L, Liu Y, et al. Radiomics features based on automatic segmented MRI images: prognostic biomarkers for triple-negative breast cancer treated with neoadjuvant chemotherapy. *Eur J Radiol*. 2022;146:110095.
42 Li W, Huang Y-H, Zhu T, et al. Noninvasive artificial intelligence system for early predicting residual cancer burden during neoadjuvant chemotherapy in breast cancer. *Ann Surg*. 2024. https://doi.org/10.1097/SLA.0000000000006279.
43 Cho N, Im S-A, Park I-A, et al. Breast cancer: early prediction of response to neoadjuvant chemotherapy using parametric response maps for MR imaging. *Radiology*. 2014;272(2):385–396.
44 Wu J, Gong G, Cui Y, Li R. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. *J Magn Reson Imag*. 2016;44(5):1107–1115.
45 Chen X, Chen X, Yang J, Li Y, Fan W, Yang Z. Combining dynamic contrast-enhanced magnetic resonance imaging and apparent diffusion coefficient maps for a radiomics nomogram to predict pathological complete response to neoadjuvant chemotherapy in breast cancer patients. *J Comput Assist Tomogr*. 2020;44(2):275–283.
46 Zhou J, Lu J, Gao C, et al. Predicting the response to neoadjuvant chemotherapy for breast cancer: wavelet transforming radiomics in MRI. *BMC Cancer*. 2020;20(1):100.

47 Peng S, Chen L, Tao J, et al. Radiomics analysis of multi-phase DCE-MRI in predicting tumor response to neoadjuvant therapy in breast cancer. *Diagnostics*. 2021;11(11):2086.
48 Zhao R, Lu H, Li YB, Shao ZZ, Ma WJ, Liu PF. Nomogram for early prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using dynamic contrast-enhanced and diffusion-weighted MRI. *Acad Radiol*. 2022;29:S155–S163.
49 Yoshida K, Kawashima H, Kannon T, et al. Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using radiomics of pretreatment dynamic contrast-enhanced MRI. *Magn Reson Imag*. 2022;92:19–25.
50 Li Y, Fan Y, Xu D, et al. Deep learning radiomic analysis of DCE-MRI combined with clinical characteristics predicts pathological complete response to neoadjuvant chemotherapy in breast cancer. *Front Oncol*. 2023;12:1041142.
51 Huang Y, Zhu T, Zhang X, et al. Longitudinal MRI-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: a multicenter, retrospective study. *EClinicalMedicine*. 2023;58:101899.
52 Shi Z, Huang X, Cheng Z, et al. MRI-Based quantification of intratumoral heterogeneity for predicting treatment response to neoadjuvant chemotherapy in breast cancer. *Radiology*. 2023;308(1):e222830.
53 Sun YS, He YJ, Li J, et al. Predictive value of DCE-MRI for early evaluation of pathological complete response to neoadjuvant chemotherapy in resectable primary breast cancer: a single-center prospective study. *Breast*. 2016;30:80–86.
54 Guo L, Du S, Gao S, et al. Delta-radiomics based on dynamic contrast-enhanced MRI predicts pathologic complete response in breast cancer patients treated with neoadjuvant chemotherapy. *Cancers*. 2022;14(14):3515.
55 Eun NL, Kang D, Son EJ, et al. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. *Radiology*. 2020;294(1):31–41.
56 Joo S, Ko ES, Kwon S, et al. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. *Sci Rep*. 2021;11(1):18800.
57 Park J, Kim MJ, Yoon JH, et al. Machine learning predicts pathologic complete response to neoadjuvant chemotherapy for ER+HER2- breast cancer: integrating tumoral and peritumoral MRI radiomic features. *Diagnostics*. 2023;13(19):3031.
58 Zhang Y, You C, Pei Y, et al. Integration of radiogenomic features for early prediction of pathological complete response in patients with triple-negative breast cancer and identification of potential therapeutic targets. *J Transl Med*. 2022;20(1):256.
59 Xu N, Guo X, Ouyang Z, et al. Multiparametric MRI-based radiomics combined with pathomics features for prediction of the efficacy of neoadjuvant chemotherapy in breast cancer. *Heliyon*. 2024;10(2):e24371.
60 Yan S, Peng H, Yu Q, et al. Computer-aided classification of MRI for pathological complete response to neoadjuvant chemotherapy in breast cancer. *Future Oncol*. 2022;18(8):991–1001.
61 Liu Z, Li Z, Qu J, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. *Clin Cancer Res*. 2019;25(12):3538–3547.
62 Li C, Lu N, He Z, et al. A noninvasive tool based on magnetic resonance imaging radiomics for the preoperative prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. *Ann Surg Oncol*. 2022;29(12):7685–7693.
63 Huang Y, Cao Y, Hu X, et al. Early identification of pathologic complete response to neoadjuvant chemotherapy using multiphase DCE-MRI by siamese network in breast cancer: a longitudinal multicenter study. *J Magn Reson Imaging*. 2024;60(4):1325–1337.
64 Wang X, Hua H, Han J, Zhong X, Liu J, Chen J. Evaluation of multiparametric MRI radiomics-based nomogram in prediction of response to neoadjuvant chemotherapy in breast cancer: a two-center study. *Clin Breast Cancer*. 2023;23(6):e331–e344.
65 Lin JY, Ye JY, Chen JG, Lin ST, Lin S, Cai SQ. Prediction of receptor status in radiomics: recent advances in breast cancer research. *Acad Radiol*. 2024;31(7):3004–3014.
66 Khan N, Adam R, Huang P, Maldjian T, Duong TQ. Deep learning prediction of pathologic complete response in breast cancer using MRI and other clinical data: a systematic review. *Tomography*. 2022;8(6):2784–2795.
67 Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. *Lancet Digit Health*. 2022;4(6):e406–e414.
68 Jabbour S, Fouhey D, Kazerooni E, Sjoding MW, Wiens J. *Deep learning applied to chest X-rays: exploiting and preventing shortcuts*. PMLR; 2020:750–782.
69 Wiens J, Creary M, Sjoding MW. AI models in health care are not colour blind and we should not be either. *Lancet Digit Health*. 2022;4(6):e399–e400.
70 Ebrahimian S, Kalra MK, Agarwal S, et al. FDA-Regulated AI algorithms: trends, strengths, and gaps of validation studies. *Acad Radiol*. 2022;29(4):559–566.
71 Swami N, Corti C, Curigliano G, Celi LA, Dee EC. Exploring biases in predictive modelling across diverse populations. *Lancet Healthy Longev*. 2022;3(2):e88.
72 The Lancet Digital H. Race representation matters in cancer care. *Lancet Digit Health*. 2021;3(7):e408.
73 Pfob A, Liu J, Sidey-Gibbons C, et al. 147P Racial bias in pretreatment MRI radiomics features to predict response to neoadjuvant systemic treatment in breast cancer: a multicenter study in China, Germany, and the US. *ESMO Open*. 2024;9.
74 Knight HE, Deeny SR, Dreyer K, et al. Challenging racism in the use of health data. *Lancet Digit Health*. 2021;3(3):e144–e146.
75 Celi LA, Cellini J, Charpignon ML, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities-A global review. *PLoS Digit Health*. 2022;1(3):e0000022.
76 Viswanathan VS, Parmar V, Madabhushi A. Towards equitable AI in oncology. *Nat Rev Clin Oncol*. 2024;21(8):628–637.
77 Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit Health*. 2020;2(9):e489–e492.
78 Early Breast Cancer Trialists' Collaborative G. Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: meta-analysis of individual patient data from ten randomised trials. *Lancet Oncol*. 2018;19(1):27–39.
79 van Loevezijn AA, van der Noordaa MEM, van Werkhoven ED, et al. Minimally invasive complete response assessment of the breast after neoadjuvant systemic therapy for early breast cancer (MICRA trial): interim analysis of a multicenter observational cohort study. *Ann Surg Oncol*. 2021;28(6):3243–3253.
80 Basik M, Cecchini RS, Santos JFDL, et al. Abstract GS5-05: primary analysis of NRG-BR005, a phase II trial assessing accuracy of tumor bed biopsies in predicting pathologic complete response (pCR) in patients with clinical/radiological complete response after neoadjuvant chemotherapy (NCT) to explore the feasibility of breast-conserving treatment without surgery. *Cancer Res*. 2020;80(4_Supplement):GS5-05.
81 Pfob A, Sidey-Gibbons C, Rauch G, et al. Intelligent vacuum-assisted biopsy to identify breast cancer patients with pathologic complete response (ypT0 and ypN0) after neoadjuvant systemic treatment for omission of breast and axillary surgery. *J Clin Oncol*. 2022;40(17):1903–1915.
82 Seth IA-O, Lim B, Joseph K, et al. Use of artificial intelligence in breast surgery: a narrative review. *Gland Surg*. 2024;13(3):395–411.
83 Fu MR, Wang Y, Li C, et al. Machine learning for detection of lymphedema among breast cancer survivors. *mHealth*. 2018;4:17.
84 Juwara L, Arora N, Gornitsky M, Saha-Chaudhuri P, Velly AM. Identifying predictive factors for neuropathic pain after breast cancer surgery using machine learning. *Int J Med Inform*. 2020;141:104170.
85 Goorts B, Dreuning KMA, Houwers JB, et al. MRI-based response patterns during neoadjuvant chemotherapy can predict pathological (complete) response in patients with breast cancer. *Breast Cancer Res*. 2018;20(1):34.
86 Shur JD, Doran SJ, Kumar S, et al. Radiomics in oncology: a practical guide. *Radiographics*. 2021;41(6):1717–1732.
87 Nardone V, Reginelli A, Guida C, et al. Delta-radiomics increases multicentre reproducibility: a phantom study. *Med Oncol*. 2020;37(5):38.
88 Zeng Q, Ke M, Zhong L, et al. Radiomics based on dynamic contrast-enhanced MRI to early predict pathologic complete response in breast cancer patients treated with neoadjuvant therapy. *Acad Radiol*. 2023;30(8):1638–1647.
89 Fan M, Chen H, You C, et al. Radiomics of tumor heterogeneity in longitudinal dynamic contrast-enhanced magnetic resonance imaging for predicting response to neoadjuvant chemotherapy in breast cancer. *Front Mol Biosci*. 2021;8:622219.

90 Janssen LM, den Dekker BM, Gilhuijs KGA, van Diest PJ, van der Wall E, Elias SG. MRI to assess response after neoadjuvant chemotherapy in breast cancer subtypes: a systematic review and meta-analysis. *NPJ breast cancer*. 2022;8(1):107.
91 Duan J, Zhao Y, Sun Q, et al. Imaging-proteomic analysis for prediction of neoadjuvant chemotherapy responses in patients with breast cancer. *Cancer Med*. 2023;12(23):21256–21269.
92 Park JE, Kim D, Kim HS, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol*. 2020;30(1):523–536.
93 Aliferis C, Simon G. Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. In: Simon GJ, Aliferis C, eds. *Artificial intelligence and machine learning in health care and medical sciences: best practices and pitfalls*. Cham: Springer International Publishing; 2024:477–524.
94 Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018;286(3):800–809.
95 Group IC. Computerised interpretation of fetal heart rate during labour (INFANT): a randomised controlled trial. *Lancet*. 2017;389(10080):1719–1729.
96 Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195.
97 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18.
98 Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. *Radiology*. 2021;298(3):505–516.
99 Kang W, Qiu X, Luo Y, et al. Application of radiomics-based multiomics combinations in the tumor microenvironment and cancer prognosis. *J Transl Med*. 2023;21(1):598.
100 Beig N, Patel J, Prasanna P, et al. Radiogenomic analysis of hypoxia pathway is predictive of overall survival in Glioblastoma. *Sci Rep*. 2018;8(1):7.
101 Tunali I, Tan Y, Gray JE, et al. Hypoxia-related radiomics and immunotherapy response: a multicohort study of non-small cell lung cancer. *JNCI Cancer Spectr*. 2021;5(4):pkab048.
102 Li S, Dai Y, Chen J, Yan F, Yang Y. MRI-based habitat imaging in cancer treatment: current technology, applications, and challenges. *Cancer Imag*. 2024;24(1):107.
103 Zhang Y-P, Zhang X-Y, Cheng Y-T, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. *Military Med Res*. 2023;10(1):22.
104 Orlhac F, Eertink JJ, Cottereau A-S, et al. A guide to ComBat harmonization of imaging biomarkers in multicenter studies. *J Nucl Med*. 2022;63(2):172.
105 Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of A Method to compensate multicenter effects affecting CT radiomics. *Radiology*. 2019;291(1):53–59.
106 Khorrami M, Bera K, Leo P, et al. Stable and discriminating radiomic predictor of recurrence in early stage non-small cell lung cancer: multi-site study. *Lung Cancer*. 2020;142:90–97.
107 Holzinger A, Langs G, Denk H, Zatloukal K, Muller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2019;9(4):e1312.
108 Frasca M, La Torre D, Pravettoni G, Cutica I. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. *Discover Artificial Intelligence*. 2024;4(1):15.
109 Gupta J, Seeja KR. A comparative study and systematic analysis of XAI models and their applications in healthcare. *Arch Computat Methods Eng*. 2024;31:3977–4002.
110 Bertsimas D, Antonios Margonis G, Sujichantararat S, et al. Interpretable artificial intelligence to optimise use of imatinib after resection in patients with localised gastrointestinal stromal tumours: an observational cohort study. *Lancet Oncol*. 2024;25(8):1025–1037.
111 O'Donnell JPM, Gasior SA, Davey MG, et al. The accuracy of breast MRI radiomic methodologies in predicting pathological complete response to neoadjuvant chemotherapy: a systematic review and network meta-analysis. *Eur J Radiol*. 2022;157:110561.
112 Choudhery S, Gomez-Cardona D, Favazza CP, et al. MRI radiomics for assessment of molecular subtype, pathological complete response, and residual cancer burden in breast cancer patients treated with neoadjuvant chemotherapy. *Acad Radiol*. 2022;29(Suppl 1):S145–S154.
113 Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. *Lancet Digit Health*. 2024;6(5):e367–e373.
114 Hong JC, Eclov NCW, Dalal NH, et al. System for high-intensity evaluation during radiation therapy (SHIELD-RT): a prospective randomized study of machine learning-directed clinical evaluations during radiation and chemoradiation. *J Clin Oncol*. 2020;38(31):3652–3661.
115 Natesan D, Eisenstein EL, Thomas SM, et al. Health care cost reductions with machine learning-directed evaluations during radiation therapy - an economic analysis of a randomized controlled study. *NEJM AI*. 2024;1(4).
116 Halling-Brown MD, Warren LM, Ward D, et al. OPTIMAM mammography image database: a large-scale resource of mammography images and clinical data. *Radiol Artif Intell*. 2021;3(1):e200103.
117 Ellis S, Gomes S, Trumble M, et al. Deep learning for breast cancer risk prediction: application to a large representative UK screening cohort. *Radiol Artif Intell*. 2024;6(4):e230431.
118 Pedemonte S, Tsue T, Mombourquette B, et al. A semiautonomous deep learning system to reduce false positives in screening mammography. *Radiol Artif Intell*. 2024;6(3):e230033.
119 Kaur H, Rani V, Kumar M, Sachdeva M, Mittal A, Kumar K. Federated learning: a comprehensive review of recent advances and applications. *Multimed Tool Appl*. 2024;83(18):54165–54188.
120 Sharma S, Guleria K. A comprehensive review on federated learning based models for healthcare applications. *Artif Intell Med*. 2023;146:102691.
121 Cremonesi F, Planat V, Kalokyri V, et al. The need for multimodal health data modeling: a practical approach for a federated-learning healthcare platform. *J Biomed Inform*. 2023;141:104338.
122 Jimenez-Sanchez A, Tardy M, Gonzalez Ballester MA, Mateus D, Piella G. Memory-aware curriculum federated learning for breast cancer classification. *Comput Methods Programs Biomed*. 2023;229:107318.
123 Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *Lancet Digit Health*. 2022;4(7):e507–e519.
124 Ng AY, Oberije CJG, Ambrózay É, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. *Nat Med*. 2023;29(12):3044–3049.
125 Taylor-Phillips S, Freeman K. Artificial intelligence to complement rather than replace radiologists in breast screening. *Lancet Digit Health*. 2022;4(7):e478–e479.
126 Qiu J, Li L, Sun J, et al. Large AI models in health informatics: applications, challenges, and the future. *IEEE Journal of Biomedical and Health Informatics*. 2023;27(12):6074–6087.
127 AlSaad R, Abd-alrazaq A, Boughorbel S, et al. Multimodal large language models in health care: applications, challenges, and future outlook. *J Med Internet Res*. 2024;26:e59505.
128 Tiu E, Talius E, Patel P, Langlotz CP, Ng AY, Rajpurkar P. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. *Nat Biomed Eng*. 2022;6(12):1399–1406.
129 Wang S, Zhao Z, Ouyang X, Liu T, Wang Q, Shen D. Interactive computer-aided diagnosis on medical image using large language models. *Commun Eng*. 2024;3(1):133.
130 Zhao Z, Wang S, Gu J, et al. ChatCAD+: towards a universal and reliable interactive CAD using LLMs. *IEEE Trans Med Imaging*. 2024;43(11):3755–3766.