



OPEN ACCESS

EDITED BY
Claudia Mello-Thoms,
The University of Iowa, United States

REVIEWED BY
Robert Nishikawa,
University of Pittsburgh, United States
Ziba Gandomkar,
The University of Sydney, Australia

*CORRESPONDENCE
Meredith A. Jones
Meredith.jones@ou.edu

SPECIALTY SECTION
This article was submitted to
Breast Cancer,
a section of the journal
Frontiers in Oncology

RECEIVED 28 June 2022
ACCEPTED 04 August 2022
PUBLISHED 31 August 2022

CITATION
Jones MA, Islam W, Faiz R, Chen X and
Zheng B (2022) Applying artificial
intelligence technology to assist
with breast cancer diagnosis and
prognosis prediction.
Front. Oncol. 12:980793.
doi: 10.3389/fonc.2022.980793

COPYRIGHT
© 2022 Jones, Islam, Faiz, Chen and
Zheng. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which
does not comply with these terms.

Applying artificial intelligence technology to assist with breast cancer diagnosis and prognosis prediction

Meredith A. Jones^{1*}, Warid Islam², Rozwat Faiz²,
Xuxin Chen² and Bin Zheng²

¹School of Biomedical Engineering, University of Oklahoma, Norman, OK, United States, ²School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, United States

Breast cancer remains the most diagnosed cancer in women. Advances in medical imaging modalities and technologies have greatly aided in the early detection of breast cancer and the decline of patient mortality rates. However, reading and interpreting breast images remains difficult due to the high heterogeneity of breast tumors and fibro-glandular tissue, which results in lower cancer detection sensitivity and specificity and large inter-reader variability. In order to help overcome these clinical challenges, researchers have made great efforts to develop computer-aided detection and/or diagnosis (CAD) schemes of breast images to provide radiologists with decision-making support tools. Recent rapid advances in high throughput data analysis methods and artificial intelligence (AI) technologies, particularly radiomics and deep learning techniques, have led to an exponential increase in the development of new AI-based models of breast images that cover a broad range of application topics. In this review paper, we focus on reviewing recent advances in better understanding the association between radiomics features and tumor microenvironment and the progress in developing new AI-based quantitative image feature analysis models in three realms of breast cancer: predicting breast cancer risk, the likelihood of tumor malignancy, and tumor response to treatment. The outlook and three major challenges of applying new AI-based models of breast images to clinical practice are also discussed. Through this review we conclude that although developing new AI-based models of breast images has achieved significant progress and promising results, several obstacles to applying these new AI-based models to clinical practice remain. Therefore, more research effort is needed in future studies.

KEYWORDS

breast cancer, machine learning, deep learning, computer aided detection, computer aided diagnosis, mammography

Introduction

The latest cancer statistics data for the USA estimates that in 2022, 31% of cancer cases detected in women are breast cancer with 43,250 cases resulting in death. This accounts for 15% of total cancer-related deaths (1). Thus, breast cancer remains the most diagnosed cancer among women with the second highest mortality rate. Over the past three decades, population-based breast cancer screening has played an important role in helping detect breast cancer in the early stage and reduce the mortality rate. From 1989 to 2017, the mortality rate of breast cancer dropped 40% which translates to 375,900 breast cancer deaths averted (2). Even though the mortality rate continues to decline, the rate of decline has slowed from 1.9% per year from 1998-2011 to 1.3% per year from 2011-2017 (2). However, the efficacy of population-based breast cancer screening is a controversial topic due to the low cancer prevalence ($\leq 0.3\%$) in annual breast cancer screening resulting in a low cancer detection yield and high false-positive rate (3). This high false positive rate is indicative of a high rate of unnecessary biopsies which is not only an economic burden but also leads to unnecessary patient anxieties which often result in women being less likely to continue with routine breast cancer screening (4). Conversations pertaining to the benefits and harms of screening mammography as well as its efficacy in decreasing breast cancer mortality as screening exams do not reduce the incidence of advanced/aggressive cancers are now common (5). For example, detection of ductal carcinoma *in situ* (DCIS) or early invasive cancers that will never progress or be of risk to the patient are occurring at a disproportionately higher rate than aggressive cancers. This is referred to as overdiagnosis and often results in unnecessary treatment that may cause more harm than the cancer itself (6). Thus, improving the efficacy of breast cancer detection and/or diagnosis remains an extremely pressing global health issue (7).

While advances in medical imaging technology and progress towards better understanding the complex biological and chemical nature of breast cancer have greatly contributed to the large decline in breast cancer mortality, breast cancer is a complex and dynamic process, making cancer management a difficult journey with many hurdles along the way. The cancer detection and management pipeline has many steps, including detecting suspicious tumors, diagnosing said tumors as malignant or benign, staging the subtype and histological grade of a cancer, developing an optimal treatment plan, identifying tumor margins for surgical resections, evaluating and predicting response to chemo or radiation therapies, or predicting risk of future occurrence or reoccurrence. In this clinical pipeline, medical imaging plays a crucial role in the decision-making process for each of these tasks. Traditionally, radiologists will rely on qualitative or semi-quantitative information visually extracted from medical images to detect suspicious tumors, predict the likelihood of malignancy, and

evaluate cancer prognosis. The clinically relevant information may include enhancement patterns, presence or absence of necrosis or blood, density and size of suspicious tumors, tumor boundary margin spiculation, or location of the suspicious tumor. However, interpreting and integrating information visually detected from medical images to make a final diagnostic decision is not an easy task.

Although mammography is the most frequently employed imaging modality in breast cancer screening, its performance is often unsatisfactory with lower sensitivity (i.e., missing 1 in 8 cancers during interpretation) and very high false positive rates (i.e., $<30\%$ of biopsies are malignant) (8). Thus, the downfalls of mammography have led to an increase in the use of other adjunct imaging modalities in clinical practice including ultrasound (US) and dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) (9, 10). Digital breast tomosynthesis (DBT) is a newer modality that is commonly used in which X-ray images are taken over multiple angles in a limited range (i.e., $\pm 15^\circ$ and the acquired scanning data is reconstructed into quasi-3D breast images to reduce the impact of dense breast tissue overlap in 2D mammograms (11). Additionally, several other new imaging modalities including contrast enhanced spectral mammography (CESM) (9, 10), phase contrast breast imaging (12), breast computed tomography (13), thermography and electrical impedance tomography of breast imaging (14), and molecular breast imaging (15), have also been investigated and tested in many prospective studies or clinical trials. However, using more imaging modalities for breast cancer detection and diagnosis increases the workload of radiologists in busy clinical practice. Over the last three decades, computer-aided detection and diagnosis (CAD) schemes are being rapidly developed to optimize the busy clinical workflow by assisting radiologists in more accurately and efficiently reading and interpreting multiple images from multiple sources (16, 17).

In the literature, CAD is often differentiated as computer-aided detection (CADe) or computer-aided diagnosis (CADx). The goal of CADe schemes is to reduce observational oversight by drawing the attention of radiologists to suspicious regions in an image. Commercialized CADe schemes of mammograms have been in clinical use since 1998 (18). One study reported that in 2016 CADe was used in about 92% of screening mammograms read in the United States (18, 19). Despite the wide scale clinical adoption, the utility of CADe schemes for breast cancer screening is often questioned (20–22). On the other hand, the goal of computer-aided diagnosis (CADx) schemes is to characterize a suspicious area and assign it to a specific class. US FDA approved the first CADx scheme of breast MR images, QuantX by Qlarity Imaging (Chicago, IL) in 2017 (23). The goal of QuantX is to assist radiologists in deciding if a lesion is malignant or benign by providing a probability estimation of malignancy. This software has yet to be extensively adopted and requires much more clinical testing.

Despite great research efforts and the availability of commercialized CAD tools, the added clinical value of CAD schemes and ML-based prediction models for breast images is limited. Thus, more novel research efforts are needed to explore new approaches (24). While using radiological features from medical images to infer phenotypic information has been done for many years, recent rapid advances in bioinformatics coupled with the advent of high performing computers has led to the field of radiomics. Radiomics involves the computation of quantitative image-based features that can be mined and used to predict clinical outcomes (25). In medical imaging, radiomic techniques are used to extract a large number of features from a set of medical images to quantify and characterize the size, shape, density, heterogeneity, and texture of the targeted tumors (26). Then, a statistics-based feature analysis tool such as Lasso regression or a machine learning (ML) based pipeline is applied to identify small sets of features that are more clinically relevant to the specific application. One method to ensure the extracted features contain some clinical relevance is to segment the tumor region and extract features from there. Despite the relative simplicity of extracting relevant radiomics features, automated tumor segmentation remains a major challenge. Thus, many radiomics-based schemes use manual or semi-automated tumor segmentation. Additionally, recent enthusiasm for deep learning based artificial intelligence (AI) technology has led to new approaches for developing CAD schemes which are being rapidly explored and reported in the literature (27). Several studies have compared CAD schemes using conventional radiomics and deep learning methods to investigate their advantages and limitations (28, 29). Deep learning (DL) based CAD schemes are appealing as majority of such CAD schemes eliminate the need for tedious error prone segmentation steps and no longer need to compute and select optimal radiomic features since deep learning models can extract features directly from the medical images (30). However, despite the challenge of how to achieve high scientific rigor when developing AI-based deep learning models (31), applying AI technology to develop CAD schemes has become the mainstream technique of the CAD research community. Additionally, new AI-based models are being expanded to include broad clinical applications in realms beyond cancer detection and diagnosis, such as prediction of short-term cancer risk and prognosis or clinical outcome.

In order to help researchers better understand state-of-the-art research progress and existing technical challenges, several review articles have recently been published with a variety of goals, such as a review of deep learning (DL) models developed for breast lesion detection, segmentation, and classification (27), radiomics models developed to classify breast lesions and monitor treatment efficacy (32), and how to optimally apply DL models to three commonly used breast imaging modalities (mammograms, ultrasound, and MRI) (33). The focus of this review paper is different from the previously published review

articles for the following reasons. First, our paper details the recent advances in both radiomics and DL-based AI technologies to develop new prediction models. Second, this review paper does not review and discuss CADe (lesion detection or segmentation) schemes. It focuses on three more challenged application realms namely, prediction of breast cancer risk, tumor classification (diagnosis) and cancer prognosis (treatment response). Third, to help readers better understand the scientific rationales of applying new AI-based models of medical image to predict breast cancer risk, classify breast lesions, and predict cancer prognosis, this paper reviews recent studies that demonstrate the important relationship between medical image features and the tumor environment (genomic biomarkers), which supports the physiological relevance of radiomics based studies. Last, based on this review process, we are able to summarize several important conclusions that may benefit future research efforts in medical imaging of breast cancer. For this purpose, the rest of this paper is organized as follows. Section two briefly discusses the correlation of extracted medical image features and the tumor environment, followed by section three that surveys recent studies, which detail novel image-based applications of both radiomics and DL-based new AI-supported CAD schemes in three application fields. Lastly, section four discusses and summarizes key points that can be learned or observed from this review paper and future perspectives in developing CAD schemes of breast images.

Relationship between medical image features and tumor environment

A major focus of breast cancer research in the medical imaging field is uncovering the relationships between medical image features and the tumor microenvironment to better predict clinical outcomes (Table 1). Since traditional CAD schemes involve handcrafting a set of features, it is important to understand what kind of descriptors correlate with cancer specific genomic biomarkers, based on radiomic concepts (25), so that optimal and descriptive handcrafted feature sets can be chosen. Additionally, if an image-based marker is widely established as a biomarker for a specific hallmark of cancer such as sustaining proliferative signaling, evading growth suppressors, invasion and metastasis, angiogenesis, or resisting cell death, then monitoring changes in that image-based marker overtime will have high degree of predictive power in many aspects of the cancer management pipeline (32).

For example, many studies investigated the correlation between image-based biomarkers and tumor mechanisms of angiogenesis. As tumors grow and metastasize, there is a decrease in the amount of available oxygen due an increase in demand, resulting in a hypoxic environment (33, 48–51). To adapt to the newly hypoxic environment, the tumor will enter an

TABLE 1 Studies of correlating image-based features with tumor physiology.

Year	Author	Imaging Modality	Image Based Features Extracted	Physiological Features	Relevant Results
2015	Li et al. (34)	DCE-MRI	Quantitative Kinetic Features: K^{trans} , K_{ep} , V_e , ADC	MVD and Proliferation	K^{trans} , K_{ep} , and ADC closely correlate with MVD and Proliferation
2021	Xiao et al. (35)	DCE-MRI	Shape, intensity, and texture features Semi-Quantitative Kinetic Features: PE, SER, FTV, WF	MVD	MVD associates with SER, WF, and radiomic features
2019	Mori et al. (36)	DCE-MRI	Semi-Quantitative Kinetic Features: IER, SER, TIE Quantitative Kinetic Features: EMM derived metrics: A, α , $A\alpha$, AUC30	MVD	A, α , $A\alpha$, AUC30, and TIE significantly correlate with MVD
2016	Kim et al. (37)	DCE-MRI	Quantitative Kinetic Features: K^{trans} , K_{ep} , V_e	MVD and VEGF	MVD correlates with V_e and there is significant association between K^{trans} , tumor size, and MVD
2014	Li et al. (38)	DCE-MRI	Semi-Quantitative Kinetic Features: longest dimension, tumor volume, SER, initialAUC Quantitative Kinetic Features: K^{trans} , K_{ep} , V_e , v_p , and τ^i	pathological response to chemotherapy	SER and K_{ep} are significantly different between responders and non-responders ($p < 0.05$) and can be used to predict breast cancer response to NACT
2007	Yu et al. (39)	DCE-MRI	Quantitative Kinetic Features: K^{trans} , K_{ep} Tumor size	response to chemotherapy based on RECIST	Tumor size significantly correlates with K^{trans} and K_{ep} and change in tumor size is a better response predictor than both K^{trans} or K_{ep}
2020	Kang et al. (40)	DCE-MRI	Quantitative Kinetic Features: K^{trans} , k_{ep} , v_e , and v_p	ER, PR, HER2, Ki67, p53, EGFR, CK5/6 and lymphovascular space invasion	High K^{trans} and k_{ep} associate with poor prognostic histopathologic factors
2019	Braman et al. (41)	DCE-MRI	Texture and statistical features	HER2+	DCE-MRI texture and statistical features can identify molecular subtype of HER2+ breast cancer from HER2- breast cancers
2016	da Rocha et al. (42)	Mammography	Texture features from the local binary pattern of images	Malignant or benign lesion	GLCM features derived from the Local Binary Pattern have the best results for lesion classification ACC: 88.31% SEN: 85% SPE: 91.89%
2015	Zhu et al. (43)	DCE-MRI	Size, shape, morphological, enhancement texture, kinetic curves, enhancement-variance	miRNA expression, protein expression, gene mutations, transcriptional activities, and gene copy number variation	Transcriptional activities of various genetic pathways positively associate with tumor size, blurred tumor margin and irregular tumor shape, The miRNA expressions associates with the tumor size and enhancement texture
2018	Drukker et al. (44)	DCE-MRI	Semi-Quantitative Kinetic Features: Most enhancing tumor volume (METV)	recurrence free survival based on clinical examination after surgery	METV from pre-NACT and early treatment scans associate with recurrence-free survival
2006	Varela et al. (45)	Mammography	Texture features to characterize contrast and spiculations from the interior, border, and outer area of the mass	Malignant or benign lesions	Features from the mass border and outer regions contain the most information for distinguishing lesions.
2020	La Forgia et al. (46)	CESM	Statistical features	ER, PR, HER2, Ki67, Grade, Triple-negative	Statistical radiomic features extracted from CESM can predict histological outcomes
2017	Wu et al. (47)	DCE-MRI	Semi-Quantitative Kinetic features: FTV features, BPE features Morphological and texture features	molecular subtypes based on IHC	DCE-MRI based features may be able to non-invasively determine the subtype of a breast cancer

SEN, sensitivity; SPE, Specificity; ACC, Overall accuracy.

angiogenic state which changes the microvasculature. In this state the tumor will switch on angiogenic growth factors such as vascular endothelial growth factor (VEGF) and fibroblast growth factors (FGF) to stimulate the formation of new capillaries so that oxygen and nutrients can adequately feed the tumor (48). This process is known as angiogenesis, which is a

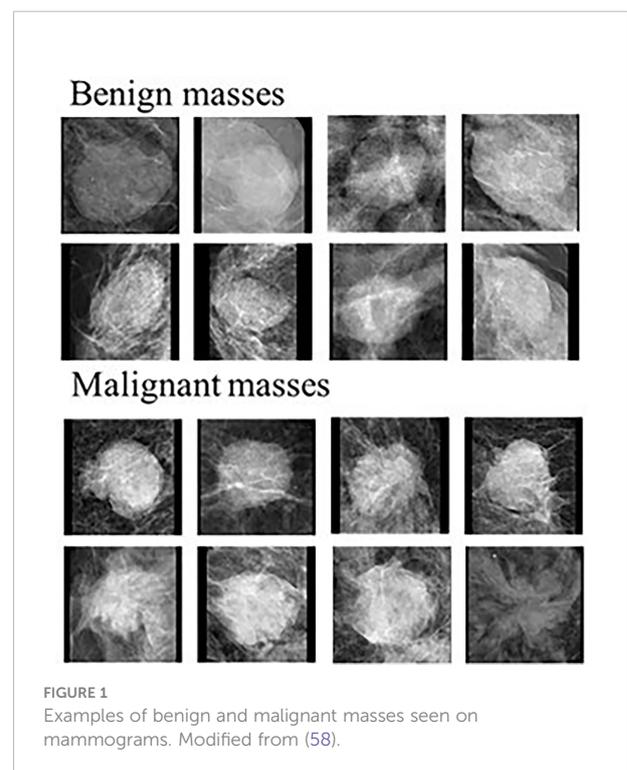
hallmark of most cancers that can be characterized by non-hierarchical, immature, and highly permeable vasculature that looks obviously different from normal vasculature (52). Traditionally, angiogenesis is indirectly quantified as microvessel density (MVD) after immunohistochemical staining of tumor tissue. While high MVD has been established as a

biomarker of poor prognosis and correlated with increased levels of angiogenesis, quantification of MVD is subject to inter- and intra-reader variability, making MVD a non-reproducible and non-standardized marker (53). Thus, development of a quick and non-invasive biomarker that can differentiate between highly immature angiogenic vasculature and normal vasculature has been a hot research topic over the past decade (48, 54).

DCE-MRI is a non-invasive method to detect and characterize the tumor microenvironment. Specifically, dynamic/kinetic image features computed from DCE-MRI characterize the permeability and perfusion kinetics of the tumor microvasculature which can reflect tumor angiogenesis. Many studies have been conducted to correlate quantitative and semi-quantitative DCE-MRI based kinetic features with MVD to demonstrate the relationship between DCE-MRI and tumor angiogenesis (34–37). Peak signal enhancement ratio (peak SER) and washout fraction (WF) are two semi-quantitative metrics extracted from the contrast enhancement curve that reflect the clearance of a contrast agent from the tumor. These metrics directly relate to a highly angiogenic state as rapid washout will occur with a large number of immature and leaky vessels (35). Extracting quantitative features from DCE-MRI requires a pharmacokinetic analysis which requires at high temporal resolution, often resulting in a poor spatial resolution. Clinical DCE-MRI scans prioritize spatial resolution as opposed to temporal resolution, which makes it difficult to do a fully quantitative analysis of clinical DCE-MRI scans. Most studies that have a goal of quantitative analysis of DCE imaging may not be appropriate for clinical use. However, studies have shown that quantitative DCE-MRI parameters such as, K^{trans} and K_{ep} , correlated well with angiogenesis markers and can be used to predict response to treatment or risk of recurrence (34). Physiologically, K_{ep} is a marker of the efflux of contrast agent. High K_{ep} values indicate two observations of tumor microenvironment. The first indicates a strong blood flow with highly permeable vessels which represents existence of an irregular and highly vascularized space associated with tumor angiogenesis. The second indicates the smaller extravascular extracellular space, meaning large quantities of the contrast agent cannot accumulate here; this is expected as there will be an increase in cell density in the tumor environment (38). Technical details pertaining to the extraction of semi-quantitative and fully quantitative kinetic features is beyond the scope of this review, interested readers should explore the following manuscripts for more information (55, 56). While there are many studies exploring the correlations between K^{trans} and K_{ep} and cancer prognosis, there are inconsistent conclusions of the biological relevance of these markers which make studies using kinetic DCE-MRI features non-reproducible (39, 40).

Recent studies demonstrated that radiomics features are thought to be more robust and reproducible than kinetic features computed from breast MRI for different prediction tasks (i.e., classification between malignant and benign tumors,

prediction of axillary lymph node metastasis, molecular subtypes of breast cancer, tumor response to chemotherapies and overall survival of patients) (57). For example, malignant tumors as seen on mammograms are typically irregular in shape with spiculated margins and architectural distortions while benign tumors are typically rounded with well-defined margins (Figure 1) (58–60). Quantification of these features can help train robust ML classifiers to better differentiate between benign and malignant masses. Features that describe the shape of the tumor may include eccentricity, diameter, convex area, orientation, and more. Shape based features may help differentiate between traditionally round benign tumors and spiculated malignant tumors. While shape features are important, breast compression during mammography makes extraction of these features difficult (60). Features can also be extracted to quantify the spiculations of the tumors which will be particularly helpful for detecting malignant breast tumors (45). First order statistical features are basic metrics that describe the distribution of intensities within an image, this includes mean, standard deviation, variance, entropy, uniformity, and others. For example, entropy quantifies the image histogram randomness which can quantify heterogeneity of the image patterns (61). Texture features belong to the biggest group of radiomics features, which are extremely useful for image recognition and image classification tasks (62, 63). Gray-level cooccurrence matrix (GLCM) based features and gray-level run length matrix (GLRLM) based features are two examples of common texture features that characterize the heterogeneity of intensities within a neighborhood of pixels. Quantification of the heterogeneity of



tumors is one of the advantages of radiomics-generated imaging markers as heterogeneity is often very difficult for radiologists to visually capture and quantify in clinical practice.

While identification of physical or biological reasoning for the correlations between image-based markers and cancer specific traits is lacking, there are some studies that do correlate radiomics based features with cancer specific markers that have been obtained from IHC analysis or genomic assays (35, 41). For example, Xiao et al. assessed the correlation between radiomic based DCE-MRI features with MVD in order to identify angiogenesis in breast cancer using DCE-MRI (35). GLCM and GLRLM derived textural features extracted from 3D segmented tumor regions were found to significantly correlate with MVD, therefore, correlate with angiogenesis levels. GLCM derived features from ROIs represented by the local binary patterns were also shown to be extremely useful for distinguishing malignant and benign masses detected on mammograms (42). Radiogenomics is the field that incorporates radiomics based features with patient specific genomic information. Correlation of the image-based features that characterize cancer through genetic information pertaining to tumor hormone receptors and genetic mutations can be very helpful for predicting risk of cancer recurrence and thus help develop optimal personalized treatment plans. Quantitative MRI-based features of tumor size, shape, and blood flow kinetics have been mapped to cancer specific genomic markers (Figure 2) (43, 44, 64). This is a great step forward in development of non-invasive techniques for understanding cancer on a molecular level.

Although DCE-MRI is an important imaging modality used to study the tumor microenvironment and predict tumor staging and/or response to therapies, other modalities have also been investigated for this purpose. For example, contrast enhanced spectral mammography (CESM) has been attracting broad clinical research interest as an alternative to DCE-MRI due to its advantages of low cost, high image resolution, and fast imaging acquisition times. Like DCE-MRI, injection of an intravenous contrast agent in CESM imaging allows for the visualization of contrast enhancement patterns which give insight into the vascular arrangement in the breast tissue. One recent paper reviewed 23 studies that investigated CESM and demonstrated that textural features and/or enhancement patterns obtained from CESM can differentiate between malignant and benign breast lesions as benign lesions often display weak and uniform contrast uptake with enhancing wash-out patterns, while malignant lesions tend to display quick decreasing wash-out patterns (65). As a result, many research studies have recently been conducted and published that compare CESM and DCE-MRI. These studies have demonstrated that CESM could achieve quite comparable performance as DCE-MRI in breast tumor diagnosis (i.e., classifying between malignant and benign tumors) (66), staging or characterizing suspicious breast lesions (46, 67), and

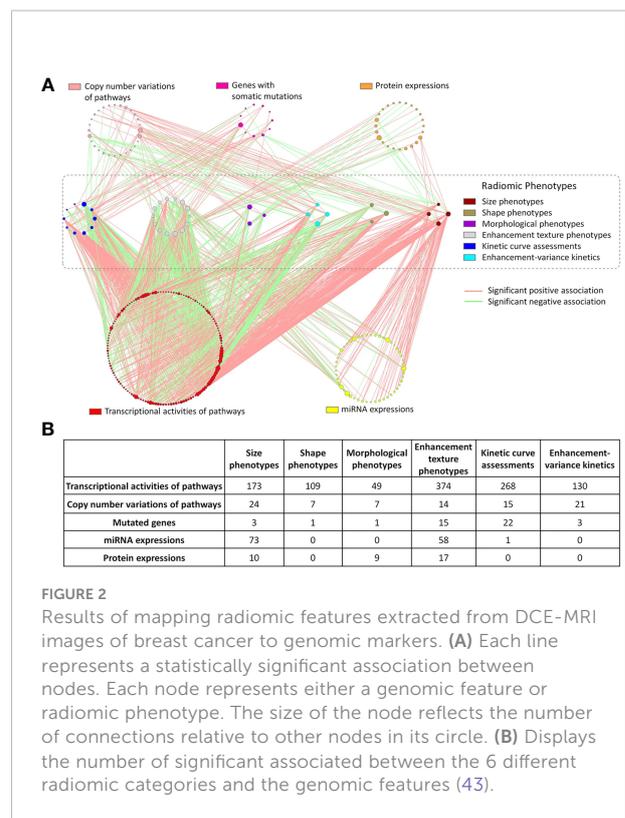


FIGURE 2 Results of mapping radiomic features extracted from DCE-MRI images of breast cancer to genomic markers. **(A)** Each line represents a statistically significant association between nodes. Each node represents either a genomic feature or radiomic phenotype. The size of the node reflects the number of connections relative to other nodes in its circle. **(B)** Displays the number of significant associated between the 6 different radiomic categories and the genomic features (43).

predicting or evaluating breast tumor response to neoadjuvant therapy (68). Thus, in the last several years, exploring and extracting image features from CESM also attracts research interest in developing new quantitative image markers or CAD schemes in breast cancer research field (69).

In previous studies, radiomics features are often only extracted from the segmented tumor regions, meaning potentially valuable information of the environment surrounding the tumor and background regions is ignored. To overcome this issue and improve the accuracy of prediction models, several studies report the importance of extracting features from the targeted or global breast parenchyma as these regions may also contain important information relating to cancer state (45, 47). While there has been a wide variety of radiomics features extracted from many different locations for different cancer applications, there is no consensus on what features make up an optimal feature set. Deciding what features should be extracted remains dependent on the goal of the individual study.

Applications of AI-based quantitative image analysis and prediction models

Rapid advances in AI technologies have promoted the development of new quantitative image feature analysis-based

prediction models in breast cancer research. In addition to the conventional CADe and CADx applications, novel AI-based models have also been expanded to new applications. In this section, we review the development and applications of AI-based prediction models in three applications namely, cancer risk prediction, tumor diagnosis or classification, and cancer prognosis prediction or response to treatment (Tables 2–4). There exists an extremely large number of studies pertaining to AI in breast cancer in the three realms mentioned. We apply the following criteria and steps to select the most relevant studies. The titles and abstracts of potentially relevant papers in the literature database (i.e., PubMed and Google Scholar) were first analyzed for terms related to either breast cancer risk (Table 2), breast cancer diagnosis/classification or computer aided diagnosis of breast cancer (Table 3), and breast cancer treatment response or prognosis prediction (Table 4). Papers were then selected if a ML or a DL method was used for predictive modeling and breast image derived features or breast images were used as model inputs. Thus, all studies also use predominantly imaging data as an input to the model. Studies were omitted if there was no explicit methodology of how the model was trained and tested or if the study lacked novelty. Studies that use solely statistical methods or do not report AUC values to make predictions were also omitted from this review. All papers listed in Tables 2–4 are published in the last 8 years. It should be noted that some studies

investigate and report performance values for multiple combinations of features or multiple classifiers, we report only the performance results of the best model.

Prediction of breast cancer risk

Women at a high risk for developing breast cancer should undergo supplemental screening exams as early detection is necessary to ensure the best prognosis (97). However, the existing risk models are mainly built based on epidemiological studies that integrate risk factors based on groups of sampled women such as: family history, hormonal and reproductive factors, breast density, obesity, smoking history, and alcohol intake, and output a breast cancer risk estimate (98, 99). By reporting odds ratios or relative risks, these risk models typically do not have discriminatory power applying to individual women. Thus, cancer detection yield in currently defined high risk groups of women remains quite low (< 3%) using mammography plus MRI screening (100). Meanwhile, up to 60% of women diagnosed with breast cancer are not considered high risk patients (101). This coupled with the increased attention to establish a new paradigm of personalized breast cancer screening highlights the need for identifying a non-invasive biomarker or developing AI-based prediction models

TABLE 2 Studies of developing AI-based image feature analysis models to predict breast cancer risk.

Year	Author	Imaging Modality	# of Images	Feature Information	ML Model	Evaluation Metrics
2018	Heidari et al. (70)	Mammography	570	43 features from the discrete cosine transform of the ROI and the spatial domain	SVM	AUC: 0.70 ± 0.04
2015	Sun et al. (71)	Mammography	340	765 texture features from multiscale subregions	SVM RBF Kernel	AUC: 0.729 ± 0.021 PPV: 0.657 (94/140) NPV: 0.755 (151/200)
2018	Mirniaharikandehi et al. (72)	Mammography	1044	8 existing CADe based features	Logistic Regression	MLO based AUC: 0.65 ± 0.017 CC based AUC: 0.586 ± 0.018
2015	Tan et al. (73)	Mammography	870	79 texture and density features	two stage ANN	AUC: 0.725 ± 0.026
2014	Gierach et al. (74)	Mammography	237	38 texture features	Bayesian ANN (BANN)	AUC: 0.72 ± 0.08
2017	Li et al. (75)	Mammography	456	4096 features from last fully connected layer of AlexNet pretrained on ImageNet	SVM	AUC: 0.83
2018	Saha et al. (76)	MRI	133	8 BPE features	multivariate logistic regression	AUC: 0.700
2019	Portnoi et al. (77)	MRI	1656	–	ResNet18 pretrained imageNet and fine tuned	AUC: 0.638 ± 0.094
2019	Yala et al. (78)	Mammography	88994	–	ResNet18	AUC: 0.70 (95% CI: 0.64, 0.73)
2021	Yala et al. (79)	Mammography	275,674	–	MIRAI	AUC: 0.76–0.79 SEN: 26.0%–41.5% SPE: 85.2%–93.1%

AUC, area under ROC curve; SEN, sensitivity; SPE, Specificity; PPV, Positive predictive value; NPV, Negative predictive value.

TABLE 3 Studies of developing new CADx models to classify between malignant and benign breast tumors.

Year	Author	Imaging Modality	# of images	Feature Information	Model	Evaluation Metrics
2020	El-Sokkary et al. (80)	Mammography	322	20 Shape and Texture Features	SVM RBF Kernel	PSO Segmentation ACC: 89.5% GMM Segmentation ACC: 87.5%
2016	Dalmis et al. (81)	MRI	395	23 Shape and Kinetic Features	Random Forest	AUC: 0.8543
2017	Qiu et al. (82)	Mammography	560	–	8 Layer CNN	AUC: 0.790 ± 0.019
2020	Yurttakal et al. (83)	MRI	200	–	multilayer CNN	ACC: 98.33% SEN: 1.0 SPE: 0.9688
2020	Hassan et al. (84)	Mammography	600	–	AlexNet pretrained on ImageNet and fine tuned	ACC: 98.29% SEN: 0.9782 SPE: 0.9876
				–	GoogleNet pretrained on ImageNet and fine tuned	Acc: 95.63% SEN: 0.9047 SPE: 0.9822
2019	Mendel et al. (85)	Mammography and DBT	78	VGG19 pretrained on ImageNet as a Feature Extractor	SVM	Mammography AUC: 0.810 ± 0.05 2D DBT AUC: 0.86 ± 0.04 Key DBT AUC: 0.89 ± 0.04
2021	Caballo et al. (86)	breast CT	284	1354 radiomic features	fusion of radiomic features and CNN based features through MLP	AUC: 0.947
2017	Antropova et al. (87)	Mammography Ultrasound MRI	739 2393 690	VGG19 pretrained on ImageNet as a Feature Extractor and radiomic features	fusion of radiomic features and CNN based features to a SVM RBF Kernel	AUC:0.86 AUC:0.90 AUC:0.89
2015	Tan et al. (88)	Mammography	1896	96 radiomic features	Multistage ANN	AUC: 0.779 ± 0.025
2019	Li et al. (89)	Mammography	182	32 lesion-based features 45 parenchymal features from contralateral breast	Bayesian ANN	AUC: 0.84 ± 0.03
2020	Heidari et al. (90)	Mammography	1000	12 Structural Similarity Index Features	SVM	AUC: 0.84 ± 0.016 ACC: 79.00%
2020	Moon et al. (91)	Ultrasound	1687 697	–	Ensemble of VGGNet, ResNet, and DenseNet	ACC: 91.10% SEN: 85.14% SPE: 95.77% Precision: 94.03% F1: 89.36% AUC: 0.9697 ACC: 94.62% SEN: 92.31% SPE: 95.60% Precision: 90% F1: 91.14% AUC: 0.9711

AUC, area under ROC curve; SEN, sensitivity; SPE, Specificity; ACC, Overall accuracy; F1, F1 index.

that can better stratify women with high or low risk of developing breast cancer in the short term based on individual testing.

Since previous studies have found that women with dense breast have a higher risk of developing breast cancer (102–106), it then leads that many studies aim to quantify breast density

TABLE 4 Studies of developing new AI-based models to predict tumor response to chemotherapy.

Year	Author	Imaging Modality	# Of Images	Feature Information	ML Model	Evaluation Metrics
2017	Giannini et al. (92)	DCE-MRI	44	27 textural features	Bayesian Classifier	ACC: 70% SPE: 0.72
2015	Michoux et al. (93)	DCE-MRI	69	3 kinetic features, 2 BI-RADS based features, 21 texture-based features	Logistic Regression	ACC: 74% SEN: 0.74 SPE: 0.74
					K-means clustering	ACC: 68% SEN: 0.84 SPE: 0.62
2015	Aghaei et al. (94)	DCE-MRI	68	39 contrast enhanced features from both segmented malignant tumor and background parenchymal enhancement regions	ANN	AUC: 0.96 ± 0.03 ACC: 94% SEN: 0.88 SPE: 0.98
2016	Aghaei et al. (95)	DCE-MRI	151	10 global kinetic features	ANN	AUC: 0.83 ± 0.03
2018	Ravichandran et al. (96)	DCE-MRI	166	–	CNN	AUC: 0.85 ACC: 82%

AUC, area under ROC curve; SEN, sensitivity; SPE, Specificity; ACC, Overall accuracy.

from screening mammograms so that patients can be informed if they have dense breast therefore are at a higher risk. It is the hope that informing women of their breast density and the risks associated with dense breast will encourage supplemental and more frequent screening exams. The American College of Radiology developed the Breast Imaging Reporting and Data System (BI-RADS) to group mammographic density into one of four categories. While BI-RADS has been used extensively, it is often unreliable as the categorization varies between observers. Machine learning and deep learning techniques have been developed that quantify breast density using computerized schemes to make this a more robust metric (107–110). While many studies have shown a correlation between breast density and breast cancer risk (111–113), this metric alone is often not enough to create robust risk assessment models (102, 114). Recent studies indicate that texture-based features may have a higher discriminatory power in stratifying women based on breast cancer risk (107, 115, 116). MRI images from The Cancer Genome Atlas (TCGA) project of the National Cancer Institute (NCI) were used to demonstrate that quantitative radiomic features extracted from breast MRI images can replicate observer-rated breast density based on BI-RADS guideline (117).

In addition to the measured breast density from mammograms, other types of medical images have been explored to develop new imaging markers or AI-based prediction models to predict breast cancer risk in individual women, particularly the short-term risk, which can help better stratify women into different breast cancer screening groups (Table 2). Heidari et al. developed a AI-based prediction scheme

to predict the risk of developing breast cancer in the short term (less than 2 years) based on features extracted from negative screening mammograms that had enhanced breast density tissue (70). The dataset used in this study included craniocaudal (CC) views of 570 negative screening mammograms that had a follow up screening exam within 2 years where 285 of these cases were then cancer positive as confirmed by tissue biopsy and 285 cases remained screening negative. The breast area was segmented from each initial negative screening mammogram and enhanced to better visualize the dense tissue as opposed to the fatty tissues. Forty-three global features were computed from the spatial domain and discrete cosine transform domain of both the left and right CC view images. This study takes advantage of the bilateral asymmetry between two breasts when creating the final feature vector that is then used to train a support vector machine (SVM) model which produces a likelihood score that the next sequential screening exam is positive. The results of this scheme were significantly better than the same scheme that does not include the segmentation and dense tissue enhancement step, emphasizing that there is important textural information in the dense tissue of negative screening mammograms that can be used to predict if there is a short-term risk of developing breast cancer.

Like conventional CADe schemes, integrating all four views of screening mammograms enables development of new cancer risk prediction models with increased performance. Mirniaharikandehei et al. investigated the hypothesis that CADe-generated false-positive lesions contain valuable information that can help predict short-term breast cancer risk (72). The motivation for this study is driven by the fact that some

early abnormalities picked up on CADe schemes may have a higher risk of developing into detectable cancers in the short-term (118, 119). All cases used in this study were negative screening exams where some of these cases contained early suspicious tumors that were only considered detectable in a retrospective review of the images. A CADe scheme was applied to right and left CC and mediolateral oblique (MLO) view images and then a feature vector was created that describes the number of initial detection seeds, the number of final false positives, the average, and the sum of all detection scores. To quantify the bilateral asymmetry, the features from the left and right CC or MLO views were summed to create one CC and one MLO view feature vector with four features in each vector. Two independent multinomial logistic regression classifiers were trained, one using the CC view feature vector and another using the MLO view feature vector. The results indicated that using the MLO view model achieved higher prediction accuracy, which indicates image features computed from CC and MLO views are different since mammograms are 2D projection images and fibroglandular tissue may appear quite different along the two projection directions. Since CADe schemes are routinely used in the clinic, this study provides a unique and cost-effective approach for developing CADe generated biomarkers from negative screening exams to help predict short term breast cancer risk. Tan et al. also took advantage of all four views of the breast and the bilateral asymmetry between breasts to predict short term breast cancer risk (73). In this study, eight groups of features were extracted from either the whole breast region or the dense tissue region of the breast to train a two-stage artificial neural network (ANN). Each feature set was used independently and in combination to train the model. The best performing model was developed when the model was trained using GLRLM based texture features computed from the dense breast regions. Both studies demonstrate that using bilateral asymmetry features computed from CC and MLO views is advantageous in that overlapping dense fibroglandular tissue can be visualized in two different configurations, providing more information about the dense tissue which is a known risk factor for breast cancer development. Clinical adoption of computerized models that can predict short-term breast cancer risk will be extremely valuable to stratify women and decide optimal intervals and methods of breast cancer screening (i.e., whether need to add breast MRI to mammography).

Genetic risk factors are also measured and used by epidemiological studies to indicate the lifetime risk of developing breast cancer. One of these genetic risk factors is an autosomal dominant mutation in the BRCA1 or BRCA2 gene. Up to 72% of women who inherit the BRCA1 mutation and 69% of women who inherit the BRCA2 mutation will develop breast cancer in their lifetime (120). Many women are unaware of their BRCA1/2 status when going in for a screening mammogram. Identification of BRCA1/2 status from routine mammographic images will be clinically useful for determining high-risk

individuals. Gierach et al. conducted a texture analysis study of breast cancer negative mammograms to differentiate individuals with BRCA1/2 mutations from those without a BRCA1/2 mutation based on 38 texture features extracted from the breast parenchyma on CC view mammograms (74). After performing feature selection, five features were used to train a Bayesian artificial neural network (BANN) model that outputs a likelihood of having a BRCA1/2 mutation which would classify the individual as high risk. Individuals with BRCA1/2 mutations used in this study were on average 10 years younger than the group without BRCA1/2 mutations. When an age-matched testing dataset was used to evaluate the performance of the BANN model and an AUC of 0.72 ± 0.08 was observed. Results of this study demonstrate that radiomic based texture features extracted from negative screening mammograms can help identify women who have BRCA1/2 mutations. The significance of this study highlights that image analysis of screening mammograms can be expanded to include risk stratification in addition to detection of suspicious tumors.

Breast parenchymal patterns are another biomarker that has been established as a tool for cancer risk prediction (104, 105, 116, 121). Extracting texture features from the breast parenchyma provides local descriptors that can characterize the physiological conditions of the breast tissue which may give more insight into breast cancer risk than breast density or BRCA mutation status. Li et al. used deep transfer learning with pre-trained CNNs to extract features directly from the breast parenchyma depicted on the CC view of FFDM images to differentiate between high-risk patients with a BRCA mutation and the low-risk patients and to differentiate between high-risk patients with unilateral cancer and the low-risk patients (75). In this study, regions of interest (ROIs) were selected from the central region directly behind the nipple as this region has been shown to give best results for describing breast parenchyma (116). ROIs were then input to a pretrained CNN and features were extracted from the last fully connected layer. In addition, texture-based features were also extracted from the ROIs so that the results of deep transfer learning-based classifier and traditional radiomic based classifier can be analyzed. A fusion classifier was created that used features extracted from the pretrained deep CNN and traditional texture features. The fusion classifier was able to differentiate BRCA mutation carriers from low-risk women and unilateral cancer patients from low-risk women with an AUC of 0.86 and 0.84, respectively. Additionally, the pre-trained CNN extracted features were able to differentiate between unilateral breast cancer patients and low risk patients significantly better than using traditional texture features, where AUC = 0.82 and AUC = 0.73, respectively. This study demonstrates the advantages of exploring deep learning techniques independently and in combination with conventional machine learning techniques to better stratify patients on breast cancer risk. In addition to extracting one ROI from one mammogram, other studies

investigate the effect of using either multiple ROIs or global features to develop breast cancer risk assessment models. For example, Sun et al. extracted texture features from multiple subregions within the mammogram that had relatively homogeneous densities and fused the features to train an SVM with a radial basis function (RBF) kernel to predict short-term breast cancer risk (71). The classifier trained using multiscale fusion of features extracted from different density subregions showed superior performance to the classifier trained using features extracted from the whole breast. Zheng et al. developed a fully automated scheme that captures the texture of the entire breast parenchyma using a lattice-based approach (122). Using smaller local windows to extract features provided the best performance when compared to single ROI and may lead to improved model performance in predicting breast cancer risk.

Besides analyzing negative mammograms, the level of background parenchymal enhancement (BPE) on breast MRI has also demonstrated power in predicting breast cancer risk (123–125). BPE refers to the volume and intensity enhancement of normal fibroglandular tissue after intravenous contrast is injected. The hypothesis is that high levels of BPE is associated with a high risk of developing breast cancer, hence why radiologists may group women into risk groups based on BPE (126). However, there is high inter-reader variability in radiologist interpretation of BPE suggesting that developing computerized schemes to quantify BPE has the potential to produce a more robust marker to predict breast cancer risk. Saha et al. automatically quantified the BPE from screening MR exams to predict future breast cancer risk within two years using a logistic regression classifier (76). In the study, eight BPE features were extracted from the fibroglandular tissue mask from both the first post-contrast fat-saturated sequence and the T1 nonfat-saturated sequence. Five breast radiologists also reviewed MR images and categorized each case as either minimal, mild, moderate, or marked BPE according to the BI-RADS guideline. The predictive performance of the multivariate logistic regression model trained using quantitative BPE features yielded higher performance than that of the qualitative BPE assessment of the five radiologists, suggesting that computerized quantification of BPE is a more accurate predictor of breast cancer risk.

Several studies have compared new image feature analysis models with pre-existing epidemiology-based statistical models in predicting cancer risk. For example, Portnoi et al. developed a deep learning breast cancer risk prediction model using DCE-MRI taken from a high-risk population (77). The 3D MR images were converted to 2D projection images using the axial view of the maximum intensity projection (MIP) and then used to fine tune a ResNet-19 CNN that had been pretrained on the ImageNet dataset. Results from the MRI-based deep learning model were compared with the Tyrer-Cuzick model and a logistic regression model that used all risk factors from the

Tyrer-Cuzick model in addition to the qualitative BPE assessment made by an expert radiologist based on the BI-RADS guidelines. The AUC of the MRI-based deep learning model, Tyrer-Cuzick model, and logistic regression model were reported as, 0.638 ± 0.094 , 0.493 ± 0.092 , and 0.558 ± 0.108 , respectively. Study results demonstrate that new MRI-based deep learning model has higher discriminatory power to predict breast cancer risk than the existing epidemiology-based risk prediction models.

Finally, based on the hypothesis that new imaging markers and the existing epidemiology-based risk factors may contain complementary information, Yala et al. sought to combine traditional risk factors and image-based risk factors extracted from mammograms using deep learning to investigate whether fusion of the two would yield a superior 5-year risk prediction model (78). In this study, ResNet18 was trained, validated, and tested using 71,689, 8,554 and 8,869 images acquired from 31,806, 3,804 and 3,978 patients, respectively. Four different risk prediction models were compared, namely: the Tyrer-Cuzick Model, a logistic regression model using standard clinical risk factors, the deep learning model, and a hybrid model using traditional clinical risk factors and the deep learning model (AUC = 0.62, 0.67, 0.68, 0.70, respectively). This work laid the foundation for the development of the MIRAI model in 2021 (79), which predicts the risk of developing breast cancer for each year within the next 5 years. All four mammograms acquired in routine screening (LCC, LML, RCC, RML view) are passed as an input to this model which first go through an image encoder, next to an image aggregator, then to a risk factor predictor, followed by an additive-hazard layer. MIRAI model was first trained and validated using 210,819 and 25,644 screening mammography exams from 56,786 and 7,020 patients from Massachusetts General Hospital (MGH), respectively. MIRAI model was then tested on three different testing sets, one acquired from MGH that contained 25,855 exams from 7,005 patients, the second acquired from Karolinska University Hospital in Sweden that contained 19,328 exams from 19,328 patients, and the third acquired from Chang Gung Memorial Hospital in Taiwan that contained 13,356 exams from 13,356 patients, respectively. AUCs obtained from MIRAI model was significantly higher than those yielded by Tyrer-Cuzick model and both the hybrid deep learning model and image based deep learning model developed in 2019 foundational study (81). Thus, MIRAI model is unique for a few reasons, the first being that traditional clinical risk factors are incorporated into the imaging feature analysis model as the previous Yala et al. study (78) demonstrated that addition of this information will improve performance. If traditional risk information is not provided, MIRAI model is still able to predict cancer risk from mammographic image features. This increases its potential clinical utility in clinics that may not record many risk factors used in Tyrer-Cuzick models. Second, MIRAI model focuses

directly on clinical implementation by training the model on a large dataset and validating this model on different datasets.

In summary, the above studies demonstrate that imaging markers computed from breast density distribution, textural features of parenchymal patterns, and parenchymal enhancement patterns are promising to build AI-based models to predict breast cancer risk. Study results have demonstrated that using image-based risk prediction models can perform superiorly to existing cancer risk prediction models that use epidemiological study data only. However, a majority of these state-of-the-art image-based risk models have not been tested or used in clinical practice due to lack of diversity in the training set leading to a model with poor generalizability on data from different locations and different scanners. Thus, these new image-based prediction models need to undergo vigorous and widespread prospective testing in future studies.

Tumor Classification or Diagnosis

Due to the high rates of false-positive recalls and high number of benign biopsy results in current clinical practice using the existing imaging modalities, it is important to investigate new methods to help decrease the false positive recall and benign biopsy rates so that women are more willing to continue participating routine breast cancer screening. Over the past few decades, a variety of AI-based CADx schemes of different types of medical images have been developed aiming to differentiate between malignant and benign tumors more accurately to help radiologists decrease the false-positive recall rates in future clinical practice (Table 3).

In order to classify a detected tumor, many CADx schemes first segment the tumor or a ROI surrounding the suspicious area before computing image features. Some studies rely on semi-automated segmentation using prior knowledge of the tumor location marked by a radiologist as an initial seed, and other studies focus on fully automated segmentation. Dalmis et al. developed an AI-based CADx scheme for DCE-MRI that uses a semi-automated tumor segmentation technique prior to feature extraction. This is done by a multi-seed smart opening algorithm that first has the user identify a seed point and then a region growing algorithm is conducted followed by a morphological opening to segment out the tumor (81). El-Sokkary et al. recently investigate two new methods for the fully automated segmentation of the ROI from the whole breast mammogram prior to feature computation and classification. The first method segments the ROI using a Gaussian Mixture Model (GMM) and the second method uses a particle swarm optimization (PSO) algorithm. Twenty texture and shape features were then extracted from each ROI independently and used to train a non-linear SVM implemented with an RBF kernel. The accuracy of classifying malignant vs benign tumors using PSO-based segmentation and GMM-based segmentation prior to feature extraction was 89.5% and 87.5%, respectively (80).

To mirror the cognitive process of a radiologist in reading and interpreting bilateral and ipsilateral CC and MLO view mammograms of the left and right breasts simultaneously, researchers have developed and tested CAD schemes that integrate tumor image features with the corresponding features computed from the matched ROIs in other mammograms. For example, Li et al. conducted and reported a study in which image features were extracted from the segmented tumor region and the contralateral breast parenchyma; when these two feature sets were combined and used to train a Bayesian artificial neural network (BANN), there significantly improved tumor classification over the BANN trained using just features from the segmented tumor region (AUC = 0.84 vs 0.79, $p=0.047$) (89).

Identifying matched ROIs from different breasts is a difficult process. To avoid errors in tumor segmentation and image registration when identifying the matched ROIs in different images, researchers have investigated the feasibility of developing CAD schemes based on global image feature analysis of multiple images. For example, Tan et al. developed a CADx scheme using bilateral mammograms to classify screening mammography cases as malignant or benign. Ninety-two handcrafted features were extracted from each of the four view images and then concatenated into separate CC and MLO feature vectors, each containing the features from the left and right breast of the respective views. A multistage ANN was then trained where the first stage had two ANNs that were trained on either the CC feature vector or the MLO feature vector, and the second stage had a singular ANN that combine the classification scores output from both the prior ANNs and outputs a final score that estimates the likelihood of the case being malignant (88). To overcome the potential limitation of losing classification sensitivity from using the whole breast image, Heidari et al. developed a novel case-based CADx scheme that quantified the bilateral asymmetry between breasts using a tree structure-based analysis of the structural similarity index (SSIM). The left and right images are equally divided into four sub-blocks, the SSIM of each pair of two matched regions is calculated and a pair of the matched sub-blocks with the lowest SSIM among the original four pairs of sub-blocks is selected. The selected sub-blocks (one from left image and one from right image) are then divided into four small sub-blocks again to search for a new pair of matched sub-blocks with the smallest SSIM. This process is repeated six times. As a result, the six smallest SSIM features are extracted for each bilateral CC and MLO view images for each case. Then, three SVMs are trained and tested using a 5-fold cross validation method using the six SSIM features computed from the bilateral CC and MLO view images separately and the combined 12 SSIM features. Each SVM produces an outcome score indicating the likelihood of the case being malignant (90). The study demonstrates that using two bilateral images of MLO view yield significantly higher performance than using two bilateral CC view images (AUC = 0.75 ± 0.021 vs. 0.53 ± 0.026). However, fusion of SSIM features computed from both CC

and MLO view images, SVM yields further increased classification accuracy with $AUC = 0.84 \pm 0.016$.

Another popular method to eliminate the tumor segmentation step in CADx schemes is by using convolutional neural networks (CNN). CNNs can automatically learn hierarchical representations of the images directly from the image, eliminating the need for semi-automated or fully automated tumor segmentation and handcrafted feature selection. Due to the limitation of image dataset sizes in the medical imaging field, researchers have developed and trained shallow CNN models (127), which do not require as much training data as a deep CNN models. However, developing an architecture and training a CNN from scratch is still an extremely time-consuming process. Additionally, the robustness of studies using shallow CNNs is often questionable as they are trained on smaller dataset. Qiu et al. trained an eight-layer CNN to predict the likelihood of a mass being malignant, demonstrating that shallow CNNs can be trained fully on medical images (82). Yurttakal et al. trained a CNN with six convolutional blocks followed by five max pooling layers, a dropout layer, one fully connected layer, and a softmax layer to output a probability of malignancy of tumors detected on MR images. The accuracy of this system is 98.33% which outperformed many other studies of similar goals (83). The deeper a model is, the more complex representations can be learned, so the question of how deep a CNN must be to sufficiently capture features for a large classification task remains (128). However, training a deep CNN from scratch is not possible without a large diverse dataset which are not readily available in the medical imaging field.

By recognizing the limitation of shallow CNN models, transfer learning has emerged as a solution to lack of big data in medical imaging. In transfer learning, a CNN is trained in one domain and applied in a new target domain (129). This involves taking advantage of existing CNNs that have been pretrained on a large data set like ImageNet and repurposing them for a new task (130). There are two approaches to transfer learning (Figure 3), one is fine tuning where some layers of a pre-trained model are frozen while other layers will be trained using the target task dataset (131). The other is using a pre-trained network exactly as is to extract feature maps that will be used to train a separate ML model or classifier. The former is beneficial in that it will train the network to have some target specific features, but the latter is advantageous in that it is computationally inexpensive as it does not require any deep CNN training. In one study, Hassan et al. fine-tuned two existing deep CNNs, AlexNet and GoogleNet, that had been pretrained on the ImageNet database to classify tumors as malignant or benign using mammograms (84). The lower layers of each deep CNN were kept frozen, and the last layers of both networks were replaced to accommodate the two-class classification task and trained using the mammograms. Many different experiments were conducted to determine the most optimal hyperparameters for each deep CNN. The mammograms used in this study were a combination of images from four databases including the Curated Breast Imaging Subset of DDSM (CBIS-DDSM), the Mammographic Image Analysis Society (MIAS), INbreast, and mammogram images from the Egyptian National Cancer Institute (NCI), demonstrating the robustness of this fully automated CADx system. In another study, Mendel et al.

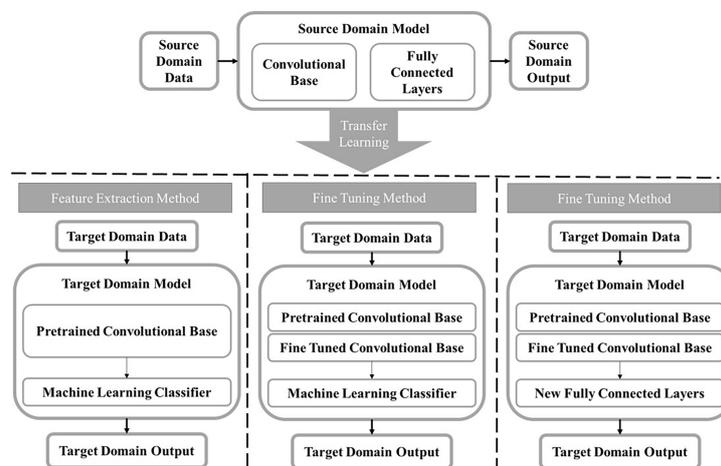


FIGURE 3

A block diagram displaying the transfer learning process. A model is trained in the source domain using a large diverse dataset. The information learned by the model is transferred to the target domain and used on a new task. The two main methods for transfer learning are feature extraction and fine tuning. For the feature extraction method, a feature map is extracted from the convolutional base taken from the source model and used to train a separate machine learning classifier. There are two ways to use transfer learning by fine tuning. The first is freezing the initial layers in the convolutional base from the source model and fine tuning the final layers using the target domain dataset then training a separate classifier. The second method does the same, except instead of training a new machine learning classifier, new fully connected layers will be added and trained using the target domain data.

used transfer learning as a feature extractor to compare the performance of a CADx model trained using DBT images and mammography images, independently. A radiologist placed a ROI around the tumor in corresponding the mammogram, DBT synthesized 2D image, and DBT key image which were then used as an input to the pre-trained VGG19 network. Features were extracted after each max-pooling layer. A stepwise feature selection method was used, and the most frequently selected features were used to train SVM models to predict the likelihood of malignancy. SVM model using DBT images yielded significantly higher classification accuracy than SVM model trained using mammograms, demonstrating that the features extracted from the DBT images may carry more clinically relevant tumor classification information than mammograms (85).

While deep CNN based models have seen tremendous success, traditional ML-based models that use handcrafted radiomic features benefit from prior knowledge of useful feature extraction methods making the handcrafted features more interpretable than automated features produced by deep learning models. Recently, fusion of traditional handcrafted features and deep learning-based features has been a hot topic and several studies report superior performance of the fusion approach over using either method alone. For example, Caballo et al. developed a CADx scheme for 3D breast computed tomography (bCT) images. The 3D mass classification problem was collapsed into a 2D classification problem by extracting nine 2D square boxes from each mass that mirror one of the nine symmetry planes of a 3D cube. The developed CADx scheme was then designed to take nine-2D images as an input. A U-Net based CNN model was used to segment the tumor from each of the nine 2D images. Then, 1,354 radiomic features were extracted from each image patch. The architecture of the rest of the proposed CADx scheme had two branches that work in parallel. The first arm of the system was a multilayer perceptron (MLP) composed of four fully connected layers that takes the radiomic features as an input. The second arm of the system was a CNN that processes the 2D image patch as is, meaning without the U-Net segmentation of the mass. The results of the last fully connected layer of both arms of the system were concatenated and processed by two more fully connected layers before tumor classification result is produced. The proposed model yielded AUC = 0.947 that outperforms three radiologists with AUC ranging from 0.814 – 0.902. This study demonstrates the utility of combining handcrafted features and CNN generated features in a singular CADx scheme (86).

Last, since original deep learning (CNN) models have been pretrained on a natural image data set like ImageNet, the models have three input channels to accept color images, yet medical images are typically gray scale images that only occupy a single input channel of the deep learning model. Thus, some studies directly copy the original grayscale image into three channels,

while other studies added additional images into the other two input channels (28). Antropova et al. conducted a study that developed a classification model that fuses radiomics and deep transfer learning generated image features using a mammogram dataset, a DCE-MRI dataset, and an US dataset (87). The mammograms and ultrasound images were stacked in three input channels and fed to a pretrained VGG19 model, while the DCE-MRI pre-contrast (t0), first time-point (t1), and post-contrast (t2) were stacked in three input channels to form the input image of another VGG19 model. The deep CNN based features were extracted after each max pooling layer, average pooled in the spatial dimension, and concatenated into a final CNN feature vector. A semi-automated tumor segmentation method was used to segment the suspicious tumors before radiomic feature extraction. The radiomic and deep CNN feature set were used to train non-linear SVM with an RBF kernel using 5-fold cross validation. To build the fusion classifier the outputs of each SVM were averaged. Classifiers trained using the fusion of the two types of features outperformed all classifiers that used either feature set alone, demonstrating that traditional radiomic features and features extracted from transfer learning may provide complimentary information that can increase the performance of CADx schemes to help radiologist better make decisions. In addition to developing this CADx scheme for three independent imaging modalities, this study also demonstrated that features extracted from each max pooling layer of a pretrained CNN outperformed features extracted from the fully connected layers. This is significant as authors claim this is the first study using a hierarchical deep feature extraction technique for CADx of breast tumor classification. Similarly, Moon et al. developed a CADx scheme using multiple US image representations to train multiple CNNs which were then combined using an ensemble method (91). Four different US image representations were used: an ROI surrounding the whole tumor and tumor boundary that was manually annotated by an expert, the segmented tumor region, the tumor shape image which is a binary mask of the segmented tumor region, and a fused RGB image of the three prior image types. Multiple CNNs were then trained on each of the four image types and the best models were combined *via* an ensemble method. All models were evaluated using one private and one public dataset involving 1,687 and 697 tumors, respectively. Results of this study further demonstrate that the more information used in the input image, the better the model performs. Future work to automate the segmentation steps will improve the robustness of this model.

The above studies demonstrate that tumor segmentation remains one of the most difficult challenges that traditional ML based CADx schemes encounter and a major hurdle to clinical implementation. The shift from manual to semi-automated to fully automated lesion segmentation has decreased the inherent

bias associated with human intervention, but elimination of the segmentation step in its entirety through either feature extraction from whole breast images or CNNs will be more generalizable than models involving a segmentation step when a large and diverse image database is available. Additionally, there remains no consensus on whether conventional ML models or new CNN-based DL models are better for breast lesion diagnosis as both methods have unique strengths and limitations. However, fusion of the two types of models has been shown to produce the best results as meaning these models may provide complementary information.

Prediction of tumor response to treatment

Monitoring response to treatment is one of the most crucial aspects of breast cancer treatment and management. This must be done continuously through a combination of physical examinations, imaging techniques, surgical interventions, and pathological analyses. Molecular subtyping of each cancer based on histopathology into either luminal A, luminal B, human epidermal growth factor 2 (HER2) enriched, and basal-like subtypes is an important first step before deciding on the optimal treatment plan as each group has shown different responses to treatments and has varying survival outcomes (132, 133). Discovery of additional molecular signatures such as presence or absence of Ki67, expression of estrogen receptors (ER) and progesterone receptor (PR), cyclin-dependent kinases (CDKs), PIK3CA mutation, and others has opened the door for new targeted therapies that aim to inhibit cancer growth rather than shrink solid tumors (134, 135).

Neoadjuvant chemotherapy (NACT) is often used as a first line treatment with the goal of decreasing the size of the tumor. Evaluation of the efficacy of NACT is traditionally done through clinical evaluation using the Response Evaluation Criteria in Solid Tumors (RECIST), a size-based guideline (136, 137). The goal of the RECIST criteria is to categorize the response as either complete response (CR), partial response (PR), progressive disease (PD), or stable disease (SD). However, changes in the size of tumors will often not be detectable until 6-8 weeks in the treatment course therefore patients may continue experiencing the toxicity affects from chemotherapy or radiation therapy while not actually treating the cancer (138). In addition, the invention of many molecularly targeted therapies may be successful without showing a decrease in the size of the tumors, other factors such as change in vasculature or molecular composition may be better indicators of treatment response (139). Immunohistochemical (IHC) analysis can also be conducted before and after therapies to uncover molecular signatures and information about the vascular density of the

tumor microenvironment (140–142). However, IHC analysis is an invasive procedure that is limited by the heterogeneity of the tumor since the biopsy sample is not necessarily reflective of the entire tumor (140, 143). The heterogeneity of tumors is a major hallmark of cancer, yet it is difficult to capture in a clinical setting making it difficult to predict response to therapy without knowing the entire molecular composition of the tumor. The need for non-invasive imaging markers that can quickly and accurately predict response to therapies has never been greater.

In current clinical practice, breast MRI is the most accurate imaging modality for monitoring tumor response to treatment as confirmed by The American College of Radiology Imaging Network (ACRIN) 6657 study performed in combination with the multi-institutional Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging And molecular Analysis (I-SPY TRIAL) (144). In these clinical trials, radiologists read MR images and predict tumor response to treatment based on RECIST guidelines. In order to predict tumor response or cancer prognosis more accurately and effectively, many researchers have tried to develop AI-based prediction models of breast MR images acquired before, during or post therapy to predict tumor response to chemotherapy at an early stage.

In one study, Giannini et al. extracted 27 texture features from pre-NACT MRI and trained a Bayesian classifier to predict pathological complete response (pCR) post-NACT (92). In another study, Michoux et al. extracted texture, kinetic, and BI-RADS features from pre-NACT MRI to try and differentiate between individuals who would have no response (NR) and those who had either a partial response (PR) or complete response (CR) (93). Predictive capabilities of the features were analyzed independently and in combination through supervised and unsupervised ML models. Results showed that texture and kinetic features helped differentiate responders vs. non-responders, but BI-RADS features did not significantly contribute to the differentiation.

Aghaei et al. reported two studies that identified two new imaging markers by training two ANN models using kinetic image features extracted from DCE-MRI acquired prior to NACT to predict complete response (CR) to NACT (94). In the first study, an existing CAD scheme was applied to segment tumors depicting on DCE-MRI. Thirty-nine contrast enhanced kinetic features were then extracted from five groups: the whole tumor area, the contrast-enhanced tumor area, the necrotic tumor area, the entire background parenchymal region of both breasts, and the absolute value of bilateral BPE between the left and right breast. Using a leave-one-case-out cross validation method embedded with a feature selection algorithm, the trained ANN yielded prediction performance with an AUC = 0.96 ± 0.03 when 10 kinetic features were used. When comparing some of the common MRI features between the CR and NR groups using

DeLong's Method, no significant differences were seen between the two groups which demonstrates that conventional MR features alone may not have enough discriminatory power to predict whether a patient will respond to NACT or not. This study demonstrates that extracting more complex MRI features will yield greater performance in predicting the likelihood of a patient responding to NACT. As with many CAD studies, inclusion of the segmentation step often limits the robustness of the scheme. Thus, Aghaei et al. conducted a follow-up study using an increased image dataset and a new scheme that only computes 10 global kinetic features from the whole breast volume including average enhancement value (EV), standard deviation (STD) of EV, skewness of EV, maximum EV, average EV of top 10%, average EV of 5%, bilateral average EV difference, bilateral STD EV difference, bilateral difference of average EV of top 10%, and bilateral difference of average EV of top 5% without tumor segmentation. Then, by using the same ANN training and testing method, the ANN trained using 4 features yielded an AUC = 0.83 ± 0.04 . Three of these four features were computed to characterize the bilateral asymmetry between left and right breasts, highlighting the key role that breast asymmetry may play in predicting whether a patient will respond well to chemotherapy (95).

CNNs provide another tool that can overcome the limitations intrinsic to tumor segmentation steps. Ravichandran et al. used a CNN with six convolutional blocks trained over 30 epochs to extract features from pre-NACT DCE-MRI to predict the likelihood of a pathological CR (pCR) (96). This study looked at the pre-contrast and post-contrast images separately and together and found that the CNN performed best when using 3-channel images that contained the pre-contrast images in the red and green channel and the post-contrast images in the blue channel. The addition of clinical variables such as age, largest diameter, and hormone receptor status increased the AUC values from 0.77 to 0.85, demonstrating how the addition of AI can streamline

imaging and clinical data into a single workflow for the increased prediction accuracy. Additionally, regions in the images that contain the most valuable information for predicting response to NACT can often be displayed in a heatmap (Figure 4). This may be an important step to reveal rationale of DL model prediction as few existing DL models are very interpretable which hinders their clinical translation.

Traditionally, pathological assessment of a representative tissue sample from the original tumor mass is used to identify the molecular subtype and develop a treatment plan. This is a sub-optimal technique as this representative tissue sample cannot capture the molecular composition of the whole tumor as cancer is often extremely heterogenous. Imaging modalities have the unique advantage of being able to capture information relating to an entire tumor which can help to overcome the limitations intrinsic to tissue biopsies. Additionally, the mechanism of many therapies is dependent on tumor vasculature which is not often probed before deciding on a treatment plan. Modalities that can image tumor vasculature such as DCE-MRI continue to be the most accurate and useful modalities in AI-based models for predicting response to treatment as valuable information pertaining to treatment response is contained in the tumor vasculature. Despite pre-clinical research progress, there are currently no image-based markers clinically used to predict response to any cancer therapies. Thus, more research efforts are needed to continue making progress to identify and validate robust image-based biomarkers that can predict response to therapy before the therapy is administered.

Discussion – outlook and challenges

Breast cancer remains an extremely deadly disease with incidence on the rise. Early detection through routine

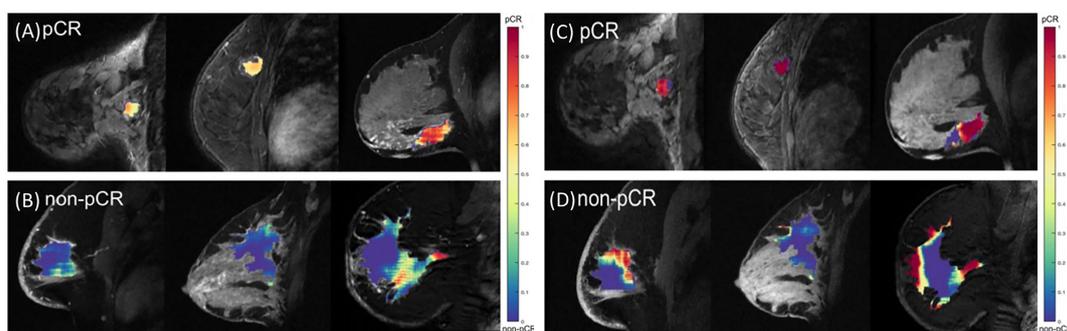


FIGURE 4

Illustration of heatmaps displaying the regions within a tumor that were used to predict the probability of pathological complete response. (A, B) show the results when using the CNNs trained on only the pre-contrast images. (C, D) show the results when using the CNN trained using a combination of pre-contrast and post-contrast images. (A, C) display cases that were correctly identified as pCR, while (B, D) are cases that were correctly identified as non-pCR. Modified from (96).

screening exams remains the best method for reducing the mortality associated with the disease. However, the efficacy including both sensitivity and specificity of current breast screening must be improved. The increase in the number of breast imaging modalities coupled with a large amount of clinical, pathological, and genetic information has made it more difficult and time consuming for clinicians to digest all available information and make an accurate diagnosis and appropriate personalized treatment plan. Recent advances in radiomics and DL-based AI technology provide promising opportunities to extract more clinically relevant image features as well as to streamline many different types of diagnostic information to build novel decision-making support tools that aim to help clinicians make more accurate and robust cancer diagnosis and treatment decisions. In this review paper, we reviewed recent studies of developing AI-based models of breast images in three application realms.

In recent years, many “omics” topics including genomics, transcriptomics, proteomics, metabolomics, and others have attracted broad research interest in order to improve early diagnosis of breast cancer, better characterize the molecular biology of tumors, and establish an optimal personalized cancer treatment paradigm. However, these “omics” studies often require additionally invasive procedures and expensive tests generating high-throughput data that is difficult to do robust data analysis. Radiomics is advantageous in that it is non-invasive and low cost (because it only uses existing image data and does not require additional tests). Thus, the reported studies that directly apply radiomics concept and software to medical images has grown exponentially in recent years. In breast imaging, a large number of radiomics features can be extracted and computed such as from mammograms and DCE-MRI. Despite great research effort and progress, the association between radiomics and other “omics” is still not very clear and more in-depth research is needed. Thus, in this paper, we review several recent studies that investigated the relationship between radiomics features and the tumor microenvironment or tumor subtypes, which may provide researchers valuable references to continue in-depth research.

In addition, AI-based prediction models have expanded from the traditional task of detecting and diagnosing suspicious breast lesions in CAD schemes to much broader applications in breast cancer research. In this paper, we select and review application of AI-based prediction models to predict risk of having or developing breast cancer, the likelihood of the detected lesion being malignant, and cancer prognosis or response to treatment. These studies demonstrate that by applying either radiomics concepts through ML methods or deep transfer learning methods, clinically relevant image features can be extracted to build new quantitative image markers or prediction models for different breast cancer

research tasks. If successful, the role of AI in breast cancer is paving the way for developing personalized medicine as detecting and diagnosing cancer are no longer driven by generic qualitative markers but now driven by quantitative patient specific data.

Despite the extensive research efforts dedicated to the development and testing of new AI-based models in the laboratory environment, very few of these studies or models have made into clinical practice. This can be attributed to several obstacles or challenges. First, currently, most of the studies reported in the literature trained AI-based models using small datasets (i.e., <500 images). Training a model using a small dataset often results in poor generalizability and poor performance due to unavoidable bias and model overfitting. Thus, one important obstacle is lack of large and high-quality image databases for many different application tasks. Although several breast image databases are publicly available including DDSM, INbreast, MIAS, and BCDR (87), these databases mainly contain easy cases and lack subtle cases, which substantially reduces the diversity and heterogeneity of these image databases. Many existing databases reported in previous research papers are also either obsolete (i.e., DDSM and MIAS used the digitized screen-film based mammograms) or have a lack of biopsy-approved ground-truth (i.e., INbreast). Thus, AI-models developed using these “easy” databases have lower performance in applying to real diverse images acquired in clinical practice. By recognizing such limitations or challenges, more research efforts continue to build better public image databases. For example, The Cancer Imaging Archive (TCIA) was created in 2011 with the aim of developing a large, de-identified, open-access archive of medical images from a wide variety of cancers and imaging modalities (145). New significant progress is expected in future studies to build this important infrastructure in help develop robust AI-based models in medical imaging field.

Second, medical images acquired using different machines made by different companies and different image acquisition or scanning protocols in different medical centers or hospitals may have different image characteristics (i.e., image contrast or contrast-to-noise ratio). CAD schemes or AI-models are often quite sensitive to the small variations of image characteristics due to the risk of overtraining. Thus, AI-models developed in this manner are not easily translatable to independent test images acquired by different imaging machines at different clinical sites. Compared to mammograms and MRI, developing AI-models of ultrasound images faces additional challenges because the quality of US images (particularly US images acquired using handheld US devices) heavily depends on the operators. Thus, establishment of TCIA allows researchers to train and validate their prediction models on imaging data acquired from other clinical sites to help researchers develop

more accurate and robust models that can eventually be translated to the clinic. Additionally, developing and implementing image pre-processing algorithms to effectively standardize or normalize images acquired from different machines or clinic sites (146, 147) have also attracted research interest and effort, which may also need before AI-based models can be adopted on a widescale clinical level.

Third, another common limitation of traditional ML or radiomics based AI-based models is that they often require a lesion segmentation step prior to feature extraction. Whether lesion segmentation is done semi-automatically based on an initial seed or automatically without human intervention, accurate and robust segmentation of breast lesions from the highly heterogeneous background tissue remains difficult (148). The lesion segmentation error introduces uncertainty or bias to the model due to the variation of the computed image features and hinders the translation of the AI-based models to clinical applications. Recent attention to DL technology provides a way to overcome this limitation as the deep CNNs will extract features directly from the images themselves, bypassing the need for a lesion segmentation step. However, the lack of big and diverse datasets is a major challenge in developing robust DL-based AI models. Although transfer learning has emerged as a mainstream in the medical imaging field, its advantages and limitations are still under investigation. While there is a huge focus on using pre-trained CNNs as feature extractors as it is computationally inexpensive and generalizable since these models avoid having to train or re-train the CNN at different centers with different imaging parameters, fine tuning the models has showed better results (129). Additionally, no CNN-based transfer learning models have made it to clinical use since the models are still not robust as investigated in a recent comprehensive AI-model evaluation study (31). Therefore, more development and validation studies are needed to address and overcome this challenge.

Fourth, currently most AI-based models use a “black-box” type approach and lack explainability. As a result, it reduces the confidence or willingness of clinicians to consider or accept AI-generated prediction results (149). Understanding how an AI-based CAD scheme or prediction model can make reliable prediction is non-trivial to most individuals because it is very difficult to explain the clinical or physical meanings of the features automatically extracted by a CNN-based deep transfer learning model. Thus, developing explainable AI models in medical image analysis has emerged as a hot research topic (150). Among these efforts, visualization tools with interactive capability or functions have been developed that aim to show the user what regions in an image or image patterns (i.e., “heat maps”) contribute the most to the decision made by AI models (151, 152). In general, new explainable AI models must be able to provide sound interpretation of how the features extracted result in the output produced. Ideally this should be done in ways that directly tie to the medical condition in question. Since

this is an emerging research field and important research direction, more research efforts should dedicate to extensive development of new technologies to make AI-based CAD schemes and/or prediction models more transparent, interpretable, and explainable before AI-based models or decision-making supporting tools can be fully accepted by the clinicians and then integrated into the clinical workflow.

Fifth, performance of AI-based models reported in the literature based on laboratory studies may not be directly applicable to clinical practice. For example, researchers have found that higher sensitivity of AI-based models may not help radiologists in reading and interpreting images in clinical practice. One previous observer performance study reported that radiologists failed to recognize correct prompts of CADe scheme in 71% of missed cancer cases due to higher false-positive prompts (153). By retrospectively analyzing a large cohort of clinical data before and after implementing CADe schemes in multiple community hospitals, one study reported that the current method of using CADe schemes in mammography reduced radiologists’ performance as seen by decreased specificity and positive predictive values (21). In order to overcome this issue, researchers have investigated several new approaches of using CADe schemes. One study reported that using an interactive prompt method to replace a conventional “second reader” prompt method significantly improves radiologists’ performance in detecting malignant masses from mammograms (154). However, this interactive prompting method has not been accepted in clinical practice. Thus, the lessons learned from CADe schemes used in clinical practice indicate that more research efforts are needed to investigate and develop new methods, including FDA clearance processes, to evaluate the potential clinical utility of all new AI-based models for many different clinical medical imaging applications (155).

In conclusion, besides CADe schemes that have been commercially available, advances in new technologies including data analysis of high throughput radiomics features and AI-based deep transfer learning have led to the development of large number of new CAD schemes or prediction models for different research tasks in breast cancer including prediction of cancer risk, likelihood of tumor being malignant, tumor subtypes or staging, tumor response to chemotherapies or radiation therapies, and patient progression-free survival (PFS) or overall survival (OS). However, before each of the new AI-based CAD schemes can be accepted in clinic practice, more work still needs to be done to overcome the remaining obstacles and validate its scientific rigor using large and diverse image databases acquired from multiple clinical sites. The overarching goal of this review paper is to provide readers with a better understanding of state-of-the-art status of developing new AI-based prediction models of breast images and the promising potential of using these models to help improve efficacy of breast cancer screening, diagnosis, and treatment. Additionally, by better understanding the remaining obstacles or challenges, we

expect more progress and future breakthroughs will be made by continuing research efforts in the future.

Author contributions

MJ writing of original manuscript preparation, revisions, and editing. WI, RF, XC writing, revisions, and editing. BZ. writing, revisions, editing, and funding acquisition All authors contributed to the article and approved the submitted version.

Funding

This work was funded in part by the National Institutes of Health, USA, under grant number P20GM135009.

References

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA: A Cancer J Clin* (2022) 72(1):7–33. doi: 10.3322/caac.21708
- DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. *CA: A Cancer J Clin* (2019) 69(6):438–51. doi: 10.3322/caac.21583
- Berlin L, Hall FM. More mammography muddle: emotions, politics, science, costs, and polarization. *Radiology* (2010) 255(2):311–6. doi: 10.1148/radiol.10100056
- McCann J, Stockton D, Godward S. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res* (2002) 4(5):1–9. doi: 10.1186/bcr455
- Goetzsch PC. Mammography screening is harmful and should be abandoned. *J R Soc Med* (2015) 108(9):341–5. doi: 10.1177/0141076815602452
- Brennan M, Houssami N. Discussing the benefits and harms of screening mammography. *Maturitas* (2016) 92:150–3. doi: 10.1016/j.maturitas.2016.08.003
- Wilkinson L, Gathani T. Understanding breast cancer as a global health concern. *Br J Radiol* (2022) 95(1130):20211033–. doi: 10.1259/bjr.20211033
- Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw Open* (2020) 3(3):e200265. doi: 10.1001/jamanetworkopen.2020.0265
- Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *Jama* (2012) 307(13):1394–404. doi: 10.1001/jama.2012.388
- Patel BK, Lobbes M, Lewin J. Contrast enhanced spectral mammography: a review. *Semin Ultrasound CT MRI* (2018) 39(1):70–79. doi: 10.1053/j.sult.2017.08.005
- Vedantham S, Karellas A, Vijayaraghavan GR, Kopans DB. Digital breast tomosynthesis: state of the art. *Radiology* (2015) 277(3):663. doi: 10.1148/radiol.2015141303
- Taba ST, Gureyev TE, Alakhras M, Lewis S, Lockie D, Brennan PC. X-Ray phase-contrast technology in breast imaging: principles, options, and clinical application. *Am J Roentgenology* (2018) 211(1):133–45. doi: 10.2214/AJR.17.19179
- Berger N, Marcon M, Saltybaeva N, Kalender WA, Alkadi H, Frauenfelder T, et al. Dedicated breast computed tomography with a photon-counting detector: initial results of clinical *in vivo* imaging. *Invest Radiology* (2019) 54(7):409–18. doi: 10.1097/RLI.0000000000000552
- Zuluaga-Gomez J, Zerhouni N, Al Masry Z, Devalland C, Varnier C. A survey of breast cancer screening techniques: thermography and electrical impedance tomography. *J Med Eng Technol* (2019) 43(5):305–22. doi: 10.1080/03091902.2019.1664672
- Covington MF, Parent EE, Dibble EH, Rauch GM, Fowler AM. Advances and future directions in molecular breast imaging. *J Nucl Med* (2022) 63(1):17–21. doi: 10.2967/jnumed.121.261988
- Katzen J, Dodelzon K. A review of computer aided detection in mammography. *Clin Imaging* (2018) 52:305–9. doi: 10.1016/j.clinimag.2018.08.014
- Dorrius MD, der Weide MC, van Ooijen P, Pijnappel RM, Oudkerk M. Computer-aided detection in breast MRI: a systematic review and meta-analysis. *Eur Radiol* (2011) 21(8):1600–8. doi: 10.1007/s00330-011-2091-9
- Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiol* (2001) 220(3):781–6. doi: 10.1148/radiol.2203001282
- Keen JD, Keen JM, Keen JE. Utilization of computer-aided detection for digital screening mammography in the united states, 2008 to 2016. *J Am Coll Radiology* (2018) 15(1):44–8. doi: 10.1016/j.jacr.2017.08.033
- Rodríguez-Ruiz A, Krupinski E, Mordang J-J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of breast cancer with mammography: Effect of an artificial intelligence support system. *Radiology* (2018) 290(2):305–14. doi: 10.1148/radiol.2018181371
- 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(14):1399–409. doi: 10.1056/NEJMoa066099
- Henriksen EL, Carlsen JF, Vejborg IM, Nielsen MB, Lauridsen CA. The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review. *Acta Radiol* (2019) 60(1):13–8. doi: 10.1177/0284185118770917
- Jiang Y, Edwards AV, Newstead GM. Artificial intelligence applied to breast MRI for improved diagnosis. *Radiology* (2021) 298(1):38–46. doi: 10.1148/radiol.2020200292
- Nishikawa RM, Gur D. CADe for early detection of breast cancer—current status and why we need to continue to explore new approaches. *Acad Radiol* (2014) 21(10):1320–1. doi: 10.1016/j.acra.2014.05.018
- Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* (2018) 2(1):1–8. doi: 10.1186/s41747-018-0068-z
- Lambin P, Leijenaar RT, Deist TM, Peerlings J, De Jong EE, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62. doi: 10.1038/nrclinonc.2017.141
- Chan H-P, Samala RK, Hadjiiski LM. CAD And AI for breast cancer—recent development and challenges. *Br J Radiol* (2019) 93(1108):20190580. doi: 10.1259/bjr.20190580

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

28. Jones MA, Faiz R, Qiu Y, Zheng B. Improving mammography lesion classification by optimal fusion of handcrafted and deep transfer learning features. *Phys Med Biol* (2022) 67(5):054001. doi: 10.1088/1361-6560/ac5297
29. Danala G, Maryada SK, Islam W, Faiz R, Jones M, Qiu Y, et al. Comparison of computer-aided diagnosis schemes optimized using radiomics and deep transfer learning methods. *Bioengineering (Basel)* (2022) 9(6):256. doi: 10.3390/bioengineering9060256
30. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* (2021) 13(1):152. doi: 10.1186/s13073-021-00968-x
31. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intelligence* (2021) 3(3):199–217. doi: 10.1038/s42256-021-00307-0
32. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013
33. Li T, Kang G, Wang T, Huang H. Tumor angiogenesis and anti-angiogenic gene therapy for cancer. *Oncol Lett* (2018) 16(1):687–702. doi: 10.3892/ol.2018.8733
34. Li L, Wang K, Sun X, Wang K, Sun Y, Zhang G, et al. Parameters of dynamic contrast-enhanced MRI as imaging markers for angiogenesis and proliferation in human breast cancer. *Med Sci Monit* (2015) 21:376–82. doi: 10.12659/MSM.892534
35. Xiao J, Rahbar H, Hippe DS, Rendi MH, Parker EU, Shekar N, et al. Dynamic contrast-enhanced breast MRI features correlate with invasive breast cancer angiogenesis. *NPJ Breast Cancer* (2021) 7(1):42. doi: 10.1038/s41523-021-00247-3
36. Mori N, Abe H, Mugikura S, Takasawa C, Sato S, Miyashita M, et al. Ultrafast dynamic contrast-enhanced breast MRI: Kinetic curve assessment using empirical mathematical model validated with histological microvessel density. *Acad Radiol* (2019) 26(7):e141–e9. doi: 10.1016/j.acra.2018.08.016
37. Kim SH, Lee HS, Kang BJ, Song BJ, Kim H-B, Lee H, et al. Dynamic contrast-enhanced MRI perfusion parameters as imaging biomarkers of angiogenesis. *PLoS One* (2016) 11(12):e0168632–e. doi: 10.1371/journal.pone.0168632
38. Li X, Arlinghaus LR, Ayers GD, Chakravarthy AB, Abramson RG, Abramson VG, et al. DCE-MRI analysis methods for predicting the response of breast cancer to neoadjuvant chemotherapy: Pilot study findings. *Magnetic Resonance Med* (2014) 71(4):1592–602. doi: 10.1002/mrm.24782
39. Yu HJ, Chen J-H, Mehta RS, Nalcioglu O, Su M-Y. MRI Measurements of tumor size and pharmacokinetic parameters as early predictors of response in breast cancer patients undergoing neoadjuvant anthracycline chemotherapy. *J Magnetic Resonance Imaging* (2007) 26(3):615–23. doi: 10.1002/jmri.21060
40. Kang SR, Kim HW, Kim HS. Evaluating the relationship between dynamic contrast-enhanced MRI (DCE-MRI) parameters and pathological characteristics in breast cancer. *J Magnetic Resonance Imaging* (2020) 52(5):1360–73. doi: 10.1002/jmri.27241
41. Braman N, Prasanna P, Whitney J, Singh S, Beig N, Etesami M, et al. Association of peritumoral radiomics with tumor biology and pathologic response to preoperative targeted therapy for HER2 (ERBB2)-positive breast cancer. *JAMA Netw Open* (2019) 2(4):e192561–e. doi: 10.1001/jamanetworkopen.2019.2561
42. da Rocha SV, Braz Junior G, Silva AC, de Paiva AC, Gattass M. Texture analysis of masses malignant in mammograms images using a combined approach of diversity index and local binary patterns distribution. *Expert Syst Applications* (2016) 66:7–19. doi: 10.1016/j.eswa.2016.08.070
43. Zhu Y, Li H, Guo W, Drukker K, Lan L, Giger ML, et al. Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma. *Sci Rep* (2015) 5(1):17787. doi: 10.1038/srep17787
44. Drukker K, Li H, Antropova N, Edwards A, Papaioannou J, Giger ML. Most-enhancing tumor volume by MRI radiomics predicts recurrence-free survival "early on" in neoadjuvant treatment of breast cancer. *Cancer Imaging* (2018) 18(1):12–. doi: 10.1186/s40644-018-0145-9
45. Varela C, Timp S, Karssemeijer N. Use of border information in the classification of mammographic masses. *Phys Med Biol* (2006) 51(2):425–41. doi: 10.1088/0031-9155/51/2/016
46. La Forgia D, Fanizzi A, Campobasso F, Bellotti R, Didonna V, Lorusso V, et al. Radiomic analysis in contrast-enhanced spectral mammography for predicting breast cancer histological outcome. *Diagnostics* (2020) 10(9):708. doi: 10.3390/diagnostics10090708
47. Wu J, Sun X, Wang J, Cui Y, Kato F, Shirato H, et al. Identifying relations between imaging phenotypes and molecular subtypes of breast cancer: model discovery and external validation. *J Magnetic Resonance Imaging* (2017) 46(4):1017–27. doi: 10.1002/jmri.25661
48. Madu CO, Wang S, Madu CO, Lu Y. Angiogenesis in breast cancer progression, diagnosis, and treatment. *J Cancer* (2020) 11(15):4474–94. doi: 10.7150/jca.44313
49. Horak ER, Klenk N, Leek R, LeJeune S, Smith K, Stuart N, et al. Angiogenesis, assessed by platelet/endothelial cell adhesion molecule antibodies, as indicator of node metastases and survival in breast cancer. *Lancet* (1992) 340(8828):1120–4. doi: 10.1016/0140-6736(92)93150-L
50. Weidner N, Semple JP, Welch WR, Folkman J. Tumor angiogenesis and metastasis—correlation in invasive breast carcinoma. *New Engl J Med* (1991) 324(1):1–8. doi: 10.1056/NEJM199101033240101
51. Shrivastav S, Bal A, Singh G, Joshi K. Tumor angiogenesis in breast cancer: Pericytes and maturation does not correlate with lymph node metastasis and molecular subtypes. *Clin Breast Cancer* (2016) 16(2):131–8. doi: 10.1016/j.clbc.2015.09.002
52. Gelao L, Criscitiello C, Fumagalli L, Locatelli M, Manunta S, Esposito A, et al. Tumour dormancy and clinical implications in breast cancer. *Ecancermedicalscience* (2013) 7:320. doi: 10.3332/ecancer.2013.320
53. Uzzan B, Nicolas P, Cucherat M, Perret GY. Microvessel density as a prognostic factor in women with breast cancer: a systematic review of the literature and meta-analysis. *Cancer Res* (2004) 64(9):2941–55. doi: 10.1158/0008-5472.CAN-03-1957
54. Schneider BP, Miller KD. Angiogenesis of breast cancer. *J Clin Oncol* (2005) 23(8):1782–90. doi: 10.1200/JCO.2005.12.017
55. Moon M, Cornfeld D, Weinreb J. Dynamic contrast-enhanced breast MR imaging. *Magn Reson Imaging Clin N Am* (2009) 17(2):351–62. doi: 10.1016/j.mric.2009.01.010
56. Paldino MJ, Barboriak DP. Fundamentals of quantitative dynamic contrast-enhanced MR imaging. *Magn Reson Imaging Clin N Am* (2009) 17(2):277–89. doi: 10.1016/j.mric.2009.01.007
57. Ye D-M, Wang H-T, Yu T. The application of radiomics in breast MRI: a review. *Technol Cancer Res Treat* (2020) 19:1533033820916191. doi: 10.1177/1533033820916191
58. Cui Y, Li Y, Xing D, Bai T, Dong J, Zhu J. Improving the prediction of benign or malignant breast masses using a combination of image biomarkers and clinical parameters. *Front Oncol* (2021) 11:629321–. doi: 10.3389/fonc.2021.629321
59. Goto M, Ito H, Akazawa K, Kubota T, Kizu O, Yamada K, et al. Diagnosis of breast tumors by contrast-enhanced MR imaging: comparison between the diagnostic performance of dynamic enhancement patterns and morphologic features. *J Magn Reson Imaging* (2007) 25(1):104–12. doi: 10.1002/jmri.20812
60. Rezaei Z. A review on image-based approaches for breast cancer detection, segmentation, and classification. *Expert Syst Appl* (2021) 182:115204. doi: 10.1016/j.eswa.2021.115204
61. Wang T, Gong J, Duan HH, Wang LJ, Ye XD, Nie SD. Correlation between CT based radiomics features and gene expression data in non-small cell lung cancer. *J Xray Sci Technol* (2019) 27(5):773–803. doi: 10.3233/XST-190526
62. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Trans Systems Man Cybernetics* (1973) 3(6):610–21. doi: 10.1109/TSMC.1973.4309314
63. Nailon WH. Texture analysis methods for medical image characterisation. *Biomed Imaging* (2010) 75:100. doi: 10.5772/8912
64. Ashraf AB, Daye D, Gavenonis S, Mies C, Feldman M, Rosen M, et al. Identification of intrinsic imaging phenotypes for breast cancer tumors: preliminary associations with gene expression profiles. *Radiology* (2014) 272(2):374–84. doi: 10.1148/radiol.14131375
65. Savidas SL, Tennant SL. Quantifying lesion enhancement on contrast-enhanced mammography: a review of published data. *Clin Radiology* (2022) 77(4):e313–e20. doi: 10.1016/j.crad.2021.12.010
66. Xiang W, Rao H, Zhou L. A meta-analysis of contrast-enhanced spectral mammography versus MRI in the diagnosis of breast cancer. *Thorac Cancer* (2020) 11(6):1423–32. doi: 10.1111/1759-7714.13400
67. Lobbes MBI, Heuts EM, Moosdorff M, van Nijntzen TJA. Contrast enhanced mammography (CEM) versus magnetic resonance imaging (MRI) for staging of breast cancer: The pro CEM perspective. *Eur J Radiol* (2021) 142:109883. doi: 10.1016/j.ejrad.2021.109883
68. Patel BK, Hilal T, Covington M, Zhang N, Kosiorek HE, Lobbes M, et al. Contrast-enhanced spectral mammography is comparable to MRI in the assessment of residual breast cancer following neoadjuvant systemic therapy. *Ann Surg Oncol* (2018) 25(5):1350–6. doi: 10.1245/s10434-018-6413-x
69. Patel BK, Ranjbar S, Wu T, Pockaj BA, Li J, Zhang N, et al. Computer-aided diagnosis of contrast-enhanced spectral mammography: A feasibility study. *Eur J Radiol* (2018) 98:207–13. doi: 10.1016/j.ejrad.2017.11.024
70. Heidari M, Khuzani AZ, Danala G, Qiu Y, Zheng B. Improving performance of breast cancer risk prediction using a new CAD-based region segmentation scheme. In: *Medical Imaging 2018: Computer-Aided Diagnosis*. SPIE (2018) 10575:166–171.

71. Sun W, Tseng T-LB, Qian W, Zhang J, Saltzstein EC, Zheng B, et al. Using multiscale texture and density features for near-term breast cancer risk analysis. *Med Physics* (2015) 42(6):2853–62. doi: 10.1118/1.4919772
72. Mirniaharikandehi S, Hollingsworth AB, Patel B, Heidari M, Liu H, Zheng B. Applying a new computer-aided detection scheme generated imaging marker to predict short-term breast cancer risk. *Phys Med Biol* (2018) 63(10):105005–. doi: 10.1088/1361-6560/aabefe
73. Tan M, Pu J, Cheng S, Liu H, Zheng B. Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk. *Ann Biomed Engineering* (2015) 43(10):2416–28. doi: 10.1007/s10439-015-1316-5
74. Gierach GL, Li H, Loud JT, Greene MH, Chow CK, Lan L, et al. Relationships between computer-extracted mammographic texture pattern features and BRCA1/2 mutation status: a cross-sectional study. *Breast Cancer Res* (2014) 16(4):424. doi: 10.1186/s13058-014-0424-8
75. Li H, Giger ML, Huynh BQ, Antropova NO. Deep learning in breast cancer risk assessment: evaluation of convolutional neural networks on a clinical dataset of full-field digital mammograms. *J Med Imaging (Bellingham)* (2017) 4(4):041304. doi: 10.1117/1.JMI.4.4.041304
76. Saha A, Grimm LJ, Ghate SV, Kim CE, Soo MS, Yoon SC, et al. Machine learning-based prediction of future breast cancer using algorithmically measured background parenchymal enhancement on high-risk screening MRI. *J Magn Reson Imaging* (2019) 50(2):456–64. doi: 10.1002/jmri.26636
77. Portnoi T, Yala A, Schuster T, Barzilay R, Dontchos B, Lamb L, et al. Deep learning model to assess cancer risk on the basis of a breast MR image alone. *Am J Roentgenology* (2019) 213(1):227–33. doi: 10.2214/AJR.18.20813
78. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology* (2019) 292(1):60–6. doi: 10.1148/radiol.2019182716
79. 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(578). doi: 10.1126/scitranslmed.aba4373
80. El-Sokkary N, Arafa AA, Asad AH, Hefny HA. (2019). Machine learning algorithms for breast cancer CADx system in the mammography, 2019 15th International Computer Engineering Conference (ICENCO), (2019) 2019:210–215.
81. Dalmiş MU, Gubern-Mérida A, Vreemann S, Karssemeijer N, Mann R, Platel B. A computer-aided diagnosis system for breast DCE-MRI at high spatiotemporal resolution. *Med Phys* (2016) 43(1):84–94. doi: 10.1118/1.4937787
82. Qiu Y, Yan S, Gundreddy RR, Wang Y, Cheng S, Liu H, et al. A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. *J X-ray Sci Technology* (2017) 25(5):751–63. doi: 10.3233/XST-16226
83. Yurttakal AH, Erbay H, İkizceli T, Karavaş S. Detection of breast cancer via deep convolution neural networks using MRI images. *Multimedia Tools Applications* (2020) 79(21):15555–73. doi: 10.1007/s11042-019-7479-6
84. Hassan S, Sayed MS, Abdalla MI, Rashwan MA. Breast cancer masses classification using deep convolutional neural networks and transfer learning. *Multimedia Tools Applications* (2020) 79(41):30735–68. doi: 10.1007/s11042-020-09518-w
85. Mendel K, Li H, Sheth D, Giger M. Transfer learning from convolutional neural networks for computer-aided diagnosis: a comparison of digital breast tomosynthesis and full-field digital mammography. *Acad Radiol* (2019) 26(6):735–43. doi: 10.1016/j.acra.2018.06.019
86. Caballo M, Hernandez AM, Lyu SH, Teuwen J, Mann RM, van Ginneken B, et al. Computer-aided diagnosis of masses in breast computed tomography imaging: deep learning model with combined handcrafted and convolutional radiomic features. *J Med Imaging (Bellingham)* (2021) 8(2):024501. doi: 10.1117/1.JMI.8.2.024501
87. Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys* (2017) 44(10):5162–71. doi: 10.1002/mp.12453
88. Tan M, Qian W, Pu J, Liu H, Zheng B. A new approach to develop computer-aided detection schemes of digital mammograms. *Phys Med Biol* (2015) 60(11):4413. doi: 10.1088/0031-9155/60/11/4413
89. Li H, Mendel KR, Lan L, Sheth D, Giger ML. Digital mammography in breast cancer: additive value of radiomics of breast parenchyma. *Radiology* (2019) 291(1):15–20. doi: 10.1148/radiol.2019181113
90. Heidari M, Mirniaharikandehi S, Danala G, Qiu Y, Zheng B. A new case-based CAD scheme using a hierarchical SSIM feature extraction method to classify between malignant and benign cases, in: *SPIE Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*; (2020) doi: 10.1117/12.2549130.
91. Moon WK, Lee YW, Ke HH, Lee SH, Huang CS, Chang RF. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput Methods Programs Biomed* (2020) 190:105361. doi: 10.1016/j.cmpb.2020.105361
92. Giannini V, Mazzetti S, Marmo A, Montemurro F, Regge D, Martincich L. A computer-aided diagnosis (CAD) scheme for pretreatment prediction of pathological response to neoadjuvant therapy using dynamic contrast-enhanced MRI texture features. *Br J Radiol* (2017) 90(1077):20170269. doi: 10.1259/bjr.20170269
93. Michoux N, Van den Broeck S, Lacoste L, Fellah L, Galant C, Berlière M, et al. Texture analysis on MR images helps predicting non-response to NAC in breast cancer. *BMC Cancer* (2015) 15:574–. doi: 10.1186/s12885-015-1563-8
94. Aghaei F, Tan M, Hollingsworth AB, Qian W, Liu H, Zheng B. Computer-aided breast MR image feature analysis for prediction of tumor response to chemotherapy. *Med Phys* (2015) 42(11):6520–8. doi: 10.1118/1.4933198
95. Aghaei F, Tan M, Hollingsworth AB, Zheng B. Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy. *J Magn Reson Imaging* (2016) 44(5):1099–106. doi: 10.1002/jmri.25276
96. 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. In: *Medical imaging 2018: computer-aided diagnosis*. SPIE (2018) 10575:79–88.
97. Wang L. Early diagnosis of breast cancer. *Sensors (Basel)* (2017) 17(7). doi: 10.3390/s17071572
98. Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. *J Natl Cancer Inst* (2010) 102(10):680–91. doi: 10.1093/jnci/djq088
99. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat* (2005) 94(2):115–22. doi: 10.1007/s10549-005-5152-4
100. Hollingsworth AB, Stough RG. An alternative approach to selecting patients for high-risk screening with breast MRI. *Breast J* (2014) 20(2):192–7. doi: 10.1111/tbj.12242
101. Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the united states explained by well-established risk factors. *JNCI* (1995) 87(22):1681–5. doi: 10.1093/jnci/87.22.1681
102. Harvey JA, Bovbjerg VE. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. *Radiology* (2004) 230(1):29–41. doi: 10.1148/radiol.2301020870
103. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* (2002) 225(1):165–75. doi: 10.1148/radiol.2251011667
104. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Prev Biomarkers* (2006) 15(6):1159–69. doi: 10.1158/1055-9965.EPI-06-0034
105. Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* (1976) 37(5):2486–92. doi: 10.1002/1097-0142(197605)37:5<2486::AID-CNCR2820370542>3.0.CO;2-8
106. 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(3):227–36. doi: 10.1056/NEJMoa062790
107. Manduca A, Carston MJ, Heine JJ, Scott CG, Pankratz VS, Brandt KR, et al. Texture features from mammographic images and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* (2009) 18(3):837–45. doi: 10.1158/1055-9965.EPI-08-0631
108. Vachon CM, Brandt KR, Ghosh K, Scott CG, Maloney SD, Carston MJ, et al. Mammographic breast density as a general marker of breast cancer risk. *Cancer Epidemiol Prev Biomarkers* (2007) 16(1):43–9. doi: 10.1158/1055-9965.EPI-06-0738
109. Tan M, Zheng B, Ramalingam P, Gur D. Prediction of near-term breast cancer risk based on bilateral mammographic feature asymmetry. *Acad Radiology* (2013) 20(12):1542–50. doi: 10.1016/j.acra.2013.08.020
110. Mohamed AA, Berg WA, Peng H, Luo Y, Jankowitz RC, Wu S. A deep learning method for classifying mammographic breast density categories. *Med Phys* (2018) 45(1):314–21. doi: 10.1002/mp.12683
111. Chang Y-H, Wang X-H, Hardesty LA, Chang TS, Poller WR, Good WF, et al. Computerized assessment of tissue composition on digitized mammograms. *Acad Radiol* (2002) 9(8):899–905. doi: 10.1016/S1076-6332(03)80459-2
112. Byng JW, Yaffe MJ, Lockwood GA, Little LE, Tritchler DL, Boyd NF. Automated analysis of mammographic densities and breast carcinoma risk. *Cancer* (1997) 80(1):66–74. doi: 10.1002/(SICI)1097-0142(19970701)80:1<66::AID-CNCR9>3.0.CO;2-D
113. Glide-Hurst CK, Duric N, Littrup P. A new method for quantitative analysis of mammographic density. *Med Phys* (2007) 34(11):4491–8. doi: 10.1118/1.2789407

114. Van Gils CH, Otten JD, Verbeek AL, Hendriks JH. Mammographic breast density and risk of breast cancer: masking bias or causality? *Eur J Epidemiol* (1998) 14(4):315–20. doi: 10.1023/a:1007423824675
115. Nielsen M, Karemore G, Loog M, Raundahl J, Karssemeijer N, Otten JD, et al. A novel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer. *Cancer Epidemiol* (2011) 35(4):381–7. doi: 10.1016/j.canep.2010.10.011
116. Li H, Giger ML, Huo Z, Olopade OI, Lan L, Weber BL, et al. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location. *Med Phys* (2004) 31(3):549–55. doi: 10.1118/1.1644514
117. Sutton EJ, Huang EP, Drukker K, Burnside ES, Li H, Net JM, et al. Breast MRI radiomics: comparison of computer- and human-extracted imaging phenotypes. *Eur Radiol Exp* (2017) 1(1):22. doi: 10.1186/s41747-017-0025-2
118. Birdwell RL, Ikeda DM, O'Shaughnessy KF, Sickles EA. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology* (2001) 219(1):192–202. doi: 10.1148/radiology.219.1.r01ap16192
119. Zheng B, Good WF, Armfield DR, Cohen C, Hertzberg T, Sumkin JH, et al. Performance change of mammographic CAD schemes optimized with most-recent and prior image databases. *Acad Radiol* (2003) 10(3):283–8. doi: 10.1016/S1076-6332(03)80102-2
120. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Bloom MJ, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *Jama* (2017) 317(23):2402–16. doi: 10.1001/jama.2017.7112
121. Wei J, Chan HP, Wu YT, Zhou C, Helvie MA, Tsodikov A, et al. Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study. *Radiology* (2011) 260(1):42–9. doi: 10.1148/radiol.11101266
122. Zheng Y, Keller BM, Ray S, Wang Y, Conant EF, Gee JC, et al. Parenchymal texture analysis in digital mammography: A fully automated pipeline for breast cancer risk assessment. *Med Phys* (2015) 42(7):4149–60. doi: 10.1118/1.4921996
123. Arasu VA, Miglioretti DL, Sprague BL, Alsheik NH, Buist DSM, Henderson LM, et al. Population-based assessment of the association between magnetic resonance imaging background parenchymal enhancement and future primary breast cancer risk. *J Clin Oncol* (2019) 37(12):954–63. doi: 10.1200/JCO.18.00378
124. Bauer E, Levy MS, Domachevsky L, Anaby D, Nissan N. Background parenchymal enhancement and uptake as breast cancer imaging biomarkers: A state-of-the-art review. *Clin Imaging* (2022) 83:41–50. doi: 10.1016/j.clinimag.2021.11.021
125. Dontchos BN, Rahbar H, Partridge SC, Korde LA, Lam DL, Scheel JR, et al. Are qualitative assessments of background parenchymal enhancement, amount of fibroglandular tissue on MR images, and mammographic density associated with breast cancer risk? *Radiology* (2015) 276(2):371–80. doi: 10.1148/radiol.2015142304
126. Niell BL, Abdalah M, Stringfield O, Raghunand N, Ataya D, Gillies R, et al. Quantitative measures of background parenchymal enhancement predict breast cancer risk. *AJR Am J Roentgenol* (2021) 217(1):64–75. doi: 10.2214/AJR.20.23804
127. Gao F, Wu T, Li J, Zheng B, Ruan L, Shang D, et al. SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. *Computerized Med Imaging Graphics* (2018) 70:53–62. doi: 10.1016/j.compmedimag.2018.09.004
128. Alzubaidi L, Fadhel MA, Al-Shamma O, Zhang J, Santamaria J, Duan Y, et al. Towards a better understanding of transfer learning for medical imaging: A case study. *Appl Sci* (2020) 10(13):4523. doi: 10.3390/app10134523
129. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* (2016) 35(5):1285–98. doi: 10.1109/TMI.2016.2528162
130. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. (2009). Imagenet: A large-scale hierarchical image database, in: *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255)
131. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. *BMC Med Imaging* (2022) 22(1):1–13. doi: 10.1186/s12880-022-00793-7
132. Omranipour R, Jalili R, Yazdankhahkenary A, Assarian A, Mirzania M, Eslami B. Evaluation of pathologic complete response (pCR) to neoadjuvant chemotherapy in Iranian breast cancer patients with estrogen receptor positive and HER2 negative and impact of predicting variables on pCR. *Eur J Breast Health* (2020) 16(3):213–8. doi: 10.5152/ejbh.2020.5487
133. Haque W, Verma V, Hatch S, Suzanne Klimberg V, Brian Butler E, Teh BS. Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. *Breast Cancer Res Treat* (2018) 170(3):559–67. doi: 10.1007/s10549-018-4801-3
134. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* (2012) 490(7418):61–70. doi: 10.1038/nature11412
135. Nwabo Kamdje AH, Seke Etet PF, Vecchio L, Muller JM, Krampera M, Lukong KE. Signaling pathways in breast cancer: therapeutic targeting of the microenvironment. *Cell Signal* (2014) 26(12):2843–56. doi: 10.1016/j.cellsig.2014.07.034
136. Wang H, Mao X. Evaluation of the efficacy of neoadjuvant chemotherapy for breast cancer. *Drug Des Devel Ther* (2020) 14:2423–33. doi: 10.2147/DDDT.S253961
137. Graham LJ, Shupe MP, Schneble EJ, Flynt FL, Clemenshaw MN, Kirkpatrick AD, et al. Current approaches and challenges in monitoring treatment responses in breast cancer. *J Cancer* (2014) 5(1):58–68. doi: 10.7150/jca.7047
138. Thoeny HC, Ross BD. Predicting and monitoring cancer treatment response with diffusion-weighted MRI. *J Magn Reson* (2010) 32(1):2–16. doi: 10.1002/jmri.22167
139. Gerwing M, Herrmann K, Helfen A, Schliemann C, Berdel WE, Eisenblätter M, et al. The beginning of the end for conventional RECIST — novel therapies require novel imaging approaches. *Nat Rev Clin Oncol* (2019) 16(7):442–58. doi: 10.1038/s41571-019-0169-5
140. Choi M, Park YH, Ahn JS, Im Y-H, Nam SJ, Cho SY, et al. Evaluation of pathologic complete response in breast cancer patients treated with neoadjuvant chemotherapy: Experience in a single institution over a 10-year period. *J Pathol Transl Med* (2017) 51(1):69–78. doi: 10.4132/jptm.2016.10.05
141. Zaha DC. Significance of immunohistochemistry in breast cancer. *World J Clin Oncol* (2014) 5(3):382–92. doi: 10.5306/wjco.v5.i3.382
142. Bergin A, Loi S. Triple-negative breast cancer: recent treatment advances [version 1; peer review: 2 approved]. *F1000Res*. (2019) 8(F1000 Faculty Rev-1342). doi: 10.12688/f1000research.18888.1
143. Arunachalam HB, Mishra R, Daescu O, Cederberg K, Rakheja D, Sengupta A, et al. Viable and necrotic tumor assessment from whole slide images of osteosarcoma using machine-learning and deep-learning models. *PLoS One* (2019) 14(4):e0210706–e. doi: 10.1371/journal.pone.0210706
144. Hylton NM, Blume JD, Bernreuter WK, Pisano ED, Rosen MA, Morris EA, et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy—results from ACRIN 6657/I-SPY TRIAL. *Radiology* (2012) 263(3):663. doi: 10.1148/radiol.12110748
145. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging* (2013) 26(6):1045–57. doi: 10.1007/s10278-013-9622-7
146. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology: Opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol* (2018) 15(3 Pt B):504–8. doi: 10.1016/j.jacr.2017.12.026
147. Li XT, Huang RY. Standardization of imaging methods for machine learning in neuro-oncology. *Neurooncol Adv* (2020) 2(Suppl 4):iv49–55. doi: 10.1093/noonj/vdaa054
148. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin Radiol* (2017) 72(1):3–10. doi: 10.1016/j.crad.2016.09.013
149. 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. doi: 10.1186/s12916-019-1426-2
150. van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal* (2022) 79:102470. doi: 10.1016/j.media.2022.102470
151. Linares-Datos P, Papastefanopoulos V, Kotsiantis S. Explainable ai: A review of machine learning interpretability methods. *Entropy* (2020) 23(1):18. doi: 10.3390/e23010018
152. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? *arXiv* (2017) arXiv:1712.09923.
153. Nishikawa RM, Schmidt RA, Linver MN, Edwards AV, Papaioannou J, Stull MA. Clinically missed cancer: how effectively can radiologists use computer-aided detection? *AJR Am J Roentgenol* (2012) 198(3):708–16. doi: 10.2214/AJR.11.6423
154. Hupse R, Samulski M, Lobbes MB, Mann RM, Mus R, den Heeten GJ, et al. Computer-aided detection of masses at mammography: interactive decision support versus prompts. *Radiology* (2013) 266(1):123–9. doi: 10.1148/radiol.12120218
155. Elmore JG, Lee CI. Artificial intelligence in medical imaging—learning from past mistakes in mammography. *JAMA Health Forum* (2022) 3(2):e215207–e. doi: 10.1001/jamahealthforum.2021.5207