Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Magn Reson Imaging. Author manuscript; available in PMC 2025 December 01.

Published in final edited form as:

J Magn Reson Imaging. 2024 December ; 60(6): 2290–2308. doi:10.1002/jmri.29358.

AI Applications to Breast MRI: Today and Tomorrow

Roberto Lo Gullo, MD[1,*], **Joren Brunekreef, PHD**[2,*], **Eric Marcus, PHD**[2], **Lynn K Han, BS**[3], **Sarah Eskreis-Winkler, MD PhD**[1], **Sunitha B Thakur, PhD**[1,4], **Ritse Mann, MD**[2,5], **Kevin Groot Lipman, PhD**[2,5], **Jonas Teuwen, PHD**[2,5,**], **Katja Pinker, MD PhD**[1,**]

[1]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[2]AI for Oncology, Netherlands Cancer Institute, Amsterdam, the Netherlands

[3]Weill Cornell Medical College, New York-Presbyterian Hospital, New York, NY, USA

[4]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[5]Department of Medical Imaging, Radboud University Medical Center, Nijmegen, the Netherlands

Abstract

In breast imaging, there is an unrelenting increase in the demand for breast imaging services, partly explained by continuous expanding imaging indications in breast diagnosis and treatment. Since the human workforce providing these services is not growing at the same rate, the implementation of artificial intelligence (AI) in breast imaging has gained significant momentum to maximize workflow efficiency and increase productivity while concurrently improving diagnostic accuracy and patient outcomes. Thus far, the implementation of AI in breast imaging is at the most advanced stage with mammography and digital breast tomosynthesis techniques, followed by ultrasound, whereas the implementation of AI in breast magnetic resonance imaging (MRI) is not moving along as rapidly due to the complexity of MRI examinations and fewer available dataset. Nevertheless, there is persisting interest in AI-enhanced breast MRI applications, even as the use of and indications of breast MRI continue to expand. This review presents an overview of the basic concepts of AI imaging analysis and subsequently reviews the use cases for AI-enhanced MRI interpretation, i.e., breast MRI triaging and lesion detection, lesion classification, prediction of treatment response, risk assessment, and image quality. Finally, it provides an outlook on the barriers and facilitators for the adoption of AI in breast MRI.

Keywords

breast; magnetic resonance imaging; artificial intelligence; machine learning; deep learning

Corresponding Author: Katja Pinker, MD, PhD. Department of Radiology, Breast Imaging Service, Memorial Sloan Kettering Cancer Center, 300 E 66th Street, New York, NY 10065, USA. pinkerdk@mskcc.org.

*co-first authors

**co-senior authors

Conflicts of Interest: The other authors of this manuscript declare no conflicts of interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Introduction

In the field of breast imaging, there is an unrelenting increase in the demand for breast imaging services, partly explained by continuous expanding imaging indications in breast diagnosis and treatment. Unfortunately, the human workforce providing these services is not growing at the same rate. Therefore, artificial intelligence (AI) applications in breast imaging has gained significant momentum to maximize workflow efficiency and increase productivity while concurrently improving diagnostic accuracy and patient outcomes.

Thus far, breast imaging AI is at the most advanced stage with mammography and digital breast tomosynthesis, followed by ultrasound (1). While the implementation of AI in breast magnetic resonance imaging (MRI) is not moving along as rapidly, due to the complexity of MRI examinations and fewer available datasets (given that MRI is not as frequently performed as mammography or ultrasound) (2,3), there is persisting interest in breast MRI AI. Breast MRI is the most effective imaging modality for detecting breast cancer (4), playing an essential role in screening women at high risk of developing breast cancer (i.e., ≥ 20% lifetime risk). Breast MRI can identify more breast cancers compared to mammography, detect cancers at an earlier stage (5), and reduce the likelihood of interval cancers. It is also highly effective as a screening tool for women with a personal history of breast cancer and women with extremely dense breast tissue, albeit these women may not be classified as being at high risk (6,7).

This review presents an overview of the basic concepts of AI and the application of AI to breast MRI, focusing on narrow AI applications. Broad AI applications in breast imaging (e.g., image reconstruction, positioning, and quality assurance) have been reviewed elsewhere (8).

Concepts of AI Image Analysis

Generally, AI applications in imaging have come to rely increasingly on deep learning methods over classical machine learning methods. In deep learning, automated feature selection is integrated within the machine learning framework, allowing the model to self-learn the most representative image features for solving a particular problem (Figure 1). Although deep learning models can outperform classical machine learning models, one drawback is their lack of explainability, where it can be challenging to understand which image features the model has learned to use when solving the specified problem. Deep learning models are often referred to as “black boxes” that do not provide the same level of intuition as, for example, a radiomics model based on handcrafted features. However, there are increasing efforts to provide explainability (9). Another drawback is that the performance of deep learning models greatly depends upon the quantity and quality of the available data used to train them; compared to transfer learning, supervised learning, and semi-supervised learning, deep learning models need larger amounts of training data. Other drawbacks include overfitting and catastrophic forgetting. Overfitting occurs when the model gives accurate predictions for training data but not for new data. To address overfitting, weight regularization can be applied which results in a simpler model that is forced to learn only the relevant patterns in the training data. Catastrophic forgetting occurs as, during continual learning, deep networks tend to swiftly replace previously acquired

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

knowledge from earlier tasks. A neural variable risk minimization framework with neural variable optimizers has been shown to relieve overfitting and catastrophic forgetting (10).

An approach to improve the performance of deep learning models in low-data scenarios is *transfer learning,* whereby knowledge learned while training for a related task (where more data might be available) can be leveraged for the task at hand. For example, imaging features that predict breast cancer subtype are somewhat related to imaging features that predict breast cancer treatment response. If there are abundant MRI data to train a prediction model for breast cancer subtype (e.g., 5,000 MRI scans) but less data to train a model for treatment response (e.g., 200 MRI scans), transfer learning can help (11). A subtype prediction model can serve as a "pre-trained" model, used to initialize the treatment response model. Fine-tuning, using the 200 MRI scans, can be employed to convert the model from a subtype prediction model to a treatment response model. Notably, a recent development in this field is *self-supervised learning*, where pre-training occurs on unlabeled data. Removing the need for experts to label input data vastly increases the amount of data for pre-training. Self-supervised methods have proven to be highly effective at boosting deep learning model performance (12,13).

One of the most used deep learning architectures in imaging is the so-called convolutional neural network (CNN). By relying on a set of subsequent convolution operations, CNNs can extract both small and larger-scale features from the input image. Regarding breast MRI, which is most naturally a 3D imaging modality, two main types of CNNs can be constructed: full 3D CNNs that utilize the volumetric nature of MRI data, and 2D CNNs that take individual slices of MRI scans as input. The choice between these is typically made based on the available resources. Full 3D CNNs can integrate the most information simultaneously but are computationally intensive. 2D CNNs are less resource-demanding but may not capture useful spatial correlations in the data. On the other hand, the 2D approach may occasionally be more effective in a low-data scenario. An example study employing a 3D CNN is Zhou et al. (14), where the network was trained to diagnose and localize breast cancer in breast MRI volumes. Elsewhere, Portnoi et al. (15) compressed 3D volumes to 2D maximum intensity projections (MIP) to accommodate for limited computer memory.

A more recent development concerns so-called *deep generative models* that learn to represent the distribution of training data and thereby generate *unseen* samples. An example application of generative models in the context of breast MRI is *image synthesis*, where missing sequences from an MRI protocol can be (partially) reconstructed based on data from available sequences. This is typically performed using a *generative adversarial network (GAN)* (16), where two deep learning models are trained in hybrid: one is the *generator* which learns to generate images from a given distribution; the other is the *discriminator* which learns to distinguish real from fake images. By training these models simultaneously, the generator is forced to construct increasingly realistic images to "fool" the discriminator. We list several studies utilizing GANs for breast MRI image synthesis under the "image quality" use case.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Use Case: Breast MRI Triaging and Lesion Detection

The widespread use of MRI for breast cancer screening is hindered by substantial costs as well as lengthy scan and interpretation times. Breast MRI interpretation requires a careful review of the current imaging examination but also a comparison with multiple previous imaging examinations. While lengthy scan and interpretation times can be mitigated using abbreviated protocols typically comprising one pre- and one post-contrast T1-weighted sequence and sometimes an additional T2-weighted sequence, breast MRI screening is still performed as a full protocol (17,18). In this context, most breast MRI screening examinations yield entirely negative results, requiring no further investigation; thus, radiologists would invest a significant portion of their time reviewing scans that in fact do not show any suspicious lesions (6,19,20). Another issue related to interpretation is that the quality of interpretation can be influenced by the overall number of examinations being interpreted and the placement of a specific examination in the queue (21). Thus, AI has been investigated as an automated method to enhance the clinical workflow by triaging low suspicion examinations from high suspicion examinations. AI is a promising for handling high-volume repetitive tasks like examination triaging as part of breast cancer screening as it does not experience the fatigue that can cause errors among radiologists in similar circumstances. AI-based triaging tools can be implemented into the daily workflow so that low suspicion examinations can be assigned to specific radiologists or scheduled to be interpreted at the end of the workday. This way, more time, attention, and expertise can be channeled toward the more complex cases. Below, we elaborate on studies that have applied AI to breast MRI for the purpose of triaging and lesion detection. Table 1 summarizes these studies.

In 2022, Jing et al. (22) reported results from their retrospective study involving 837 breast MRIs performed in 438 women whereby the ultrafast section of the acquired full MRI protocol was used as the basis for AI triage. They investigated a deep learning model for automatically detecting negative scans to improve workflow efficiency. The training and validation sets comprised 659 exams in total (or 1,318 single breasts, of which 118 were abnormal and 1,200 normal), whereas the independent test set comprised 178 exams (or 356 single breasts, of which 55 were abnormal and 301 normal). The scenario was simulated whereby negative results from breast MRI time-resolved angiography with stochastic trajectories (TWIST) sequences would not necessitate radiologist interpretation or any additional patient assessment. The deep learning model demonstrated an area under the curve (AUC) of 0.81 on the independent test set. Using two thresholds (0.37 and 0.25) corresponding to a sensitivity of 100% and 95% on the validation set, the model achieved high sensitivity rates of 95% and 98% as well as high negative predictive values of 97% and 98% on the independent test set. The threshold of 0.25 resulted in a workload reduction of 6.2% (11 of 178) and no cancers being overlooked. The threshold of 0.37 resulted in a workload reduction of 15.7% (28 of 178), but 3 breast lesions (one 8-mm mucinous carcinoma and two benign lesions) were misclassified. The authors noted that one possible reason for false negatives is small lesion size, as all missed lesions were < 10 mm. Regarding false positives, the authors recommended proper handling of dense and Breast Imaging-Reporting and Data System (BI-RADS) 2 breasts in the future.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In the same year, Verburg et al. (23) reported results from their retrospective study involving 4,581 breast MRIs from the Dense Tissue and Early Breast Neoplasm Screening (DENSE) trial. Like Jing et al., they also utilized a deep learning model to triage negative examinations for the purpose of reducing radiologist workload. Patients who participated in the DENSE trial previously underwent a full multiparametric MRI protocol, but Verburg et al. used only the pre-contrast and first post-contrast images of the dynamic contrast-enhanced (DCE) series to develop the deep learning model. The model was trained with left and right breasts separate from each other, with data from seven hospitals split (80%:20%) into a training set comprising 9,386 breasts (9,386 with lesions and 9,386 without lesions) and a validation set comprising 2,346 breasts (1,173 with lesions and 1,173 without lesions); unseen data from one independent hospital was used as the test set comprising 9,164 breasts (838 with lesions and 8,326 without lesions). With a threshold corresponding to 100% sensitivity for malignant lesions, the model correctly flagged 90.7% of breast MRIs with lesions as abnormal, recommending them for radiologist interpretation. At the same time, the model ruled out 39.7% of breast MRIs without lesions. The model achieved an AUC of 0.83 for distinguishing between breast MRIs with and without lesions and did not miss any lesion that was in fact malignant. Notably, accurate triaging of negative breast MRIs without lesions was influenced somewhat by the presence of background parenchymal enhancement (BPE), whereby breast MRIs with minimal BPE were more frequently excluded compared to breast MRIs with higher levels of BPE.

A more recent study by Bhowmik et al. was published in 2023 (24). This study involved 16,535 consecutive contrast-enhanced breast MRIs in 8,354 women using a full breast MRI protocol. Training and validation of the deep learning model were based on a dataset of 14,768 breast MRIs (13,918 screening examinations and 850 examinations for the purpose of determining the extent of disease in patients with a recent cancer diagnosis). Examinations with BI-RADS scores of 1 or 2 were assigned an “extremely low suspicion” label whereas examinations with BI-RADS scores of 3–6 were assigned a “possibly suspicious” label. The external validation dataset comprised 1,687 examinations (1,441 screening examinations and 246 examinations for determining the extent of disease in patients with a recent cancer diagnosis). Of the 1,441 screening examinations, 59 contained cancer (100% were detected by the model) and 1,382 did not contain cancer (11.5% were triaged as “extremely low suspicion” by the model). Of the 246 non-screening examinations, the model accurately categorized all 246 (100%) malignant examinations as “possibly suspicious.” Correspondingly, the model achieved a specificity of 19.15% and a sensitivity of 100% on the external validation dataset, providing a potential workload reduction of 11%. Additionally, while the previous studies by Jing et al. and Verburg et al. did not compare model performance against radiologist performance, the study by Bhowmik et al. compared model performance against 2 fellowship-trained breast imaging radiologists for 80 breast MRI examinations. Of the 79 screening examinations within the reader study, 33 contained cancer, all (100%) of which were categorized as “possibly suspicious” by the model, whereas 1 reader flagged all (100%) as “possibly suspicious” and the remaining reader missed 1 cancer in an examination complicated by marked BPE (Figure 2). Of the 47 negative screening examinations within the reader study, the model classified 9 (19.15%) as “extremely low suspicion” and 38 (80.85%) as “possibly suspicious”. Thus, the reader study

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

suggests that the model is promising as a triage tool to improve the clinical workflow rather than serving as an independent reader. The model used pre- and post-contrast T1-weighted-based 2D MIP images and MIP slabs, but as the authors noted, future work could extend the model's architecture to also incorporate 3D images to improve the model's sensitivity, particularly regarding small or indistinct cancers.

Notably, in 2017, the United States Food and Drug Administration approved an AI-based software named Quant X to aid radiologists in the assessment of cancerous and non-cancerous breast lesions (25).

Use Case: Lesion Classification

One of the primary applications of AI in breast imaging is to distinguish between benign and malignant lesions (26). Several studies have reported on the use of CNN models to autonomously detect and diagnose breast lesions (14,27–31), with most showing encouraging results. Studies reporting better results, as indicated by higher AUC values, incorporated more comprehensive data inputs (multiparametric MRI inputs as well as clinical data) for training the CNN model. When AI performance was compared to that of human readers, AI has demonstrated comparable or slightly better performance, particularly when the radiologists are junior radiologists. Below, we elaborate on several studies that have applied AI for breast lesion classification. Table 2 summarizes these studies.

In 2018, Antropova at al. (29) reported results for three support vector machine (SVM) classifiers from DCE-MRIs to differentiate benign from malignant lesions. A pre-trained VGGNet with CNN architecture was used to extract features from DCE-MRIs of 690 cases (212 benign and 478 malignant). The classifier based on extracted features from the MIP image from the second post-contrast subtraction produced the best AUC of 0.88, compared to the classifier based on extracted features from the central slice of the second post-contrast image (AUC of 0.80) and the classifier based on extracted features from the central slice from the second post-contrast subtraction image (AUC of 0.84). Notably, Antropova et al. showed that while most pre-trained CNNs can only take 2D images as input, MIP images can be used to produce 2D input images from 4D DCE-MRIs.

In 2019, Truhn et al. (30) reported results comparing a CNN model based on pretrained ResNet18 architecture against two radiomic models (one using L1 regularization and one using principal component analysis) and three breast radiologists for the classification of breast lesions as benign or malignant. Based on features extracted from multiparametric MRI, the CNN model (AUC of 0.88) outperformed the two radiomic models (AUC of 0.81 for the radiomic model using L1 regularization and AUC of 0.78 for the radiomic model using principal component analysis) ($p < 0.001$); however, the CNN model had subpar performance compared to the breast radiologists (AUC of 0.98) ($p < 0.001$). The authors noted that the pre-tained architecture only allowed up to 3 input channels and thereby only 3 of all the available multiparametric MRI sequences were used; they theorized that more input (i.e., using GAN to augment the input data) would yield improved performance.

Also in 2019, Dalmiş et al. (32) reported results from their study in which they analyzed 368 malignant and 208 benign lesions using deep learning. Lesions were annotated manually

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

by placing a region of interest (ROI) at the lesion's center on subtraction volumes generated from ultrafast TWIST volumes, T2-weighted volumes, and apparent diffusion coefficient (ADC) volumes. The "DenseNet" architecture was employed in training a CNN based on TWIST imaging patches and another CNN based on T2-weighted imaging patches. Likelihood values representing the likelihood of malignancy from CNNs, ADC values, and patient information (e.g., age, *BRCA* status) were combined using random forest classifiers to generate final malignancy scores at the lesion level. The best performance was achieved by incorporating all imaging data and patient information (AUC of 0.852). When the level of sensitivity achieved by the radiologists in the original dataset was applied, the best-performing CNN yielded 130 false positives, i.e., 19 fewer than the number of benign lesions biopsied. Notably, the best-performing CNN and radiologists differed substantially in their misclassified lesions, suggesting that deep learning may be a useful adjunct to radiologists.

Herent et al. (28) reported in 2019 results from a CNN model that they built based on ResNet-50 architecture to both detect and characterize lesions according to histological subtype, in response to a data challenge organized during the *Journées Francophones de Radiologie* in Paris in October 2018. The model, trained on 335 single 2D contrast-enhanced T1-weighted fat-suppressed MRIs from 335 patients, achieved a weighted AUC of 0.816 on an independent test set of 168 MRIs, ranking first in the challenge. They urged the extension of their model to incorporate 3D data to increase its generalizability.

In 2020, Adachi et al. (31) reported results for an AI system to detect and diagnose breast lesions as benign on MIP images from DCE-MRI. The AI system, trained on DCE-MRI MIP images (using a training set of 214 cases and validation set of 72 cases) using RetinaNet architecture, showed better performance compared to 4 human readers (varying experience from 1–20 years) without AI or with AI (AUC of 0.925, 0.884, and 0.899, respectively) (using a test set of 85 cases), suggesting that AI may be useful as an adjunct for both junior and senior radiologists.

In 2021, Fujioka et al. (33) reported on the effectiveness of various CNN architectures in predict the likelihood of malignancy. Training and validation sets comprised 286 DCE-MRI MIP images (31 normal, 75 benign, and 180 malignant). The various CNN architectures included DenseNet121, DenseNet169, InceptionResNetV2, InceptionV3, NasNetMobile, and Xception, each trained over 500 epochs. The resulting models were tested on a dataset comprising 72 images (12 normal, 13 benign, and 47 malignant). Additionally, 2 human readers (a breast surgeon and a breast radiologist) independently interpreted single craniocaudal MIP images from this test dataset, scoring the likelihood of malignancy per patient by using BI-RADS categories. The CNN models exhibited a mean AUC of 0.830 (range, 0.750–0.895). The top-performing CNN model, InceptionResNetV2, achieved a sensitivity of 74.5%, a specificity of 96.0%, and an AUC of 0.895. By contrast, Reader 1 achieved a sensitivity of 72.3%, a specificity of 88.0%, and an AUC of 0.823, while Reader 2 achieved a sensitivity of 78.7%, a specificity of 80.0%, and an AUC of 0.849. It is worth noting that while some of the CNN models displayed higher AUC values compared to the human readers, these differences were insignificant ($p > 0.125$), and both the top-performing CNN model and readers did not identify over half of false-negative and false-positive cases.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To reduce misclassifications, the authors suggested that training data should include more small lesions as well as non-mass lesions.

Seeking to improve on Antropova et al.'s 2018 results whereby a VGGNet CNN classifier incorporating features from the MIP image from the second post-contrast subtraction in particular produced the best AUC of 0.88 (29), Hu et al. reported in 2021 (34) their development of a VGGNet CNN classifier incorporating 4D information (volumetric and temporal) from DCE-MRI to distinguish between benign and malignant breast lesions, hypothesizing that this classifier would outperform the one incorporating traditional MIP information. Thus, instead of using an "image MIP" approach whereby volumetric data are collapsed into MIP images at the image level, Hu et al. used a "feature MIP" approach: the maximum of the features of the CNN were assembled in the two dimensions directly within the deep neural network architecture. Additionally, in contrast to Antropova et al.'s study, Hu et al. used four dynamic timepoints from the DCE sequence in the CNN input instead of relying on a single post-contrast subtraction image. Both the image MIP and feature MIP classifiers were developed using a training and validation set of 1,455 lesions (716 malignant masses, 357 malignant non-mass enhancements, 230 benign masses, and 152 benign non-mass enhancements) and subsequently tested on an independent test set of 535 lesions (293 malignant masses, 128 malignant non-mass enhancements, 70 benign masses, and 44 benign non-mass enhancements). As hypothesized, the feature MIP classifier outperformed the image MIP classifier (AUCs of 0.93 (95% CI: 0.91–0.96) and 0.91 (95% CI: 0.87–0.94), respectively) ($P = 0.03$). Correspondingly, the feature MIP classifier allocated higher probabilities of malignancy to malignant cases and lower probabilities of malignancy to benign cases compared to the image MIP classifier.

In 2022, Zhu et al. (35) reported on the impact of various multiparametric MRI sequences (including DCE-MRI alone, diffusion-weighted imaging (DWI) alone, and combined DCE and DWI sequences) on ResNet-based models to distinguish between benign and malignant breast lesions. Notably, they also combined the ResNet-based models with a VNet-based segmentation model. This resulted in two single-input ResNet models and one multi-input ResNet model. The dataset for lesion segmentation comprised 2,823 patients while the dataset for lesion characterization comprised 3,303 patients with 3,607 lesions (2,213 malignant lesions and 1,394 benign lesions). The multi-input DCE plus DWI model attained higher sensitivity, specificity, and accuracy compared to the single-input DWI and DCE-MRI models (sensitivity: 83.1%, 77.2%, and 77.6%, respectively; specificity: 87.4%, 63.0%, and 70.9%, respectively; and accuracy: 84.6%, 72.1%, and 75.2%, respectively). Moreover, the specificity of the multi-input DCE plus DWI model was higher than that of radiologists who achieved a specificity of 40.4% when BI-RADS categories 4 and 5 were both considered malignant and 84.1% when only BI-RADS category 5 was considered malignant. On an external test dataset, the performance of all three deep learning models decreased only marginally, with AUCs of 0.81, 0.83, and 0.89, respectively.

Also in 2022, Wang et al. (36) reported on the performance of a deep learning model to classify breast non-mass enhancement (NME) lesions in particular. This study included 965 NME lesions (539 benign and 426 malignant); 754 NME lesions from one MRI scanner were randomly divided into a training set (482 lesions), validation set (121 lesions),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and “test set A” (151 lesions). An additional 211 NME lesions from a different MRI scanner composed “test set B”. The model was developed using ResNet-50 architecture and leveraged axial and sagittal MIP images derived from early post-contrast subtracted breast MR images. For each case, two radiologists, one senior and one junior, also independently reviewed the MIP images and assigned a BI-RADS score. The model achieved AUCs of 0.859 and 0.816 in test sets A and B, respectively, demonstrating comparable performance to the senior radiologist in test set A ($p = 0.558$), significantly outperforming the senior radiologist in test set B ($p = 0.041$) and the junior radiologist in both test sets ($p < 0.001$ and $p = 0.009$, respectively). With AI assistance, the junior radiologist’s AUC improved from 0.740 to 0.862 on test set A ($p < 0.001$) and from 0.732 to 0.843 on test set B ($p < 0.001$), confirming the hypothesis that AI can help classify NME lesions on breast MRI and contribute to enhanced performance for junior radiologists.

Witowski et al. (37) in 2022 described a deep learning system to predict the probability of breast cancer in 21,537 bilateral DCE-MRIs from 13,463 patients and possibly reduce false-positives from DCE-MRI clinical interpretation. The deep learning system was built using 3D pre- and post-contrast DCE-MRI volumes to mimic clinical practice, achieving an AUC of 0.92 on the internal test set and AUCs of of 0.80–0.97 on external international datasets. Moreover, the deep learning system also showed comparable diagnostic accuracy to five breast radiologists (2–12 years of experience). A simulated hybrid system (deep learning system + radiologist predictions) resulted in improved breast cancer predictions. In subgroup analysis, the deep learning system showed no differences across various histological subtypes or patient demographics, but there was slight performance deterioration with more BPE. For patients with BI-RADS 3 or 4 lesions, the deep learning system was able to reduce the number of unnecessary biopsies. Notably, the deep learning system had much better performance in datasets comprising only invasive cancers (vs. datasets comprising multiple malignant subtypes).

Cong et al. (38) in 2023 described a deep learning method for detecting and classifying breast lesions as malignant or benign, incorporating CNNs and a long short-term memory cascaded network, with input from multiparametric MRI. The study used 569 local cases as an internal cohort and 125 cases from a public dataset as an external cohort. Excluding DCE-MRI, the best performance was achieved by a 5-Seq system combining input from T2-weighted imaging, T1-weighted imaging, DWI with a b-value of 50, DWI with a b-value of 850, and ADC (AUC of 0.964), with improvements in AUC values of +0.065 for reader 1 (12 years of experience), +0.082 for reader 2 (10 years of experience), and +0.142 for reader 3 (3 years of experience). Including DCE-MRI, the best performance was achieved by combining input from DCE-MRI, DWI with a b-value of 850, and T2-weighted imaging (AUC of 0.983). The results suggest that using input without DCE-MRI may still achieve comparable performance to when DCE-MRI input is used. On the external dataset, the 5-Seq system achieved an AUC of 0.892.

Use Case: Breast Cancer Characterization

While radiomics-based classical machine learning approaches have been investigated for breast cancer characterization, more recent studies have focused on using deep learning

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(39–41). Common limitations observed across these studies include their retrospective nature, limited sample size, especially for specific cancer subtypes, often preventing the division of cases into a training dataset and a validation dataset, which can lead to the risk of overfitting. Additionally, testing was frequently limited as molecular subtypes were not confirmed through formal genetic testing but rather through immunohistochemistry surrogates. However, in clinical practice, molecular subtypes are often determined using immunohistochemistry surrogates. Table 3 summarizes these studies.

Leithner et al. (40) reported in 2019 initial results using radiomic signatures extracted from the first post-contrast T1-weighted sequence, showing that the extracted radiomic signatures were promising to distinguish between breast cancer molecular subtypes (with 79.4% accuracy for distinguishing between luminal A and luminal B breast cancer and 77.1% accuracy for distinguishing between luminal B and triple negative breast cancer on the validation set) (Figure 3). The same authors (42) used information from both DCE-MRI and ADC maps, yielding an AUC of 0.86 for the separation of triple negative \from other cancers subtypes.

A year later, in 2020, Bitencourt et al. (39) reported that several radiomic-based MRI features were significantly associated with HER2 expression in the breast tumor, and when one of the radiomic-based MRI features was combined with two other conventional MRI parameters in a machine learning model, the model was able to predict HER2 expression with an accuracy of 97.4%, a sensitivity of 99.3%, and a specificity of 81.3% (28).

Meanwhile, in 2021, Lo Gullo et al. (41) reported results showing that radiomic-based MRI features showed value in predicting PD-L1 expression in the breast tumor; indeed, the final machine learning model incorporating three radiomic-based MRI features achieved an accuracy of 88.2%, a sensitivity of 90.7%, and a specificity of 85.1% (Figure 4).

Of the studies that have used a deep learning approach, several have focused particularly on detecting invasive breast cancer. For example, Zhu et al. (43) in 2019 reported using deep learning to predict occult invasive disease on MRI in 131 patients with a recent biopsy-confirmed diagnosis of ductal carcinoma in situ. The best AUC of 0.70 was achieved using a “deep features” approach whereby a modified pre-trained GoogleNet architecture was used together with an SVM model (given that SVMs are less prone to overfitting compared to deep neural networks) to predict occult invasive disease, achieving a sensitivity of 69% and a specificity of 8%. As opposed to the “deep features” approach, the “transfer learning” approach was hampered by overfitting. Meanwhile, in 2021, Moroianu and Rusu (44) reported using a pre-trained VGG16 architecture to build a model to predict invasive breast cancer on pre-treatment MRI in 81 patients who had undergone surgery, achieving an AUC of 0.83 ± 0.05, a sensitivity of 83% ± 16%, and a specificity of 71% ± 11%.

Two studies have used deep learning to distinguish between breast molecular subtypes. In 2021, Zhang et al. (45) reported the comparison of a conventional CNN with a convolutional long short-term memory (CLSTM) network to distinguish between HR+/HER2−, HER2+, and triple negative subtypes. 244 patients were divided into a training set (n = 99 patients who underwent 1.5T MRI) and two testing sets (n = 83 and 62 patients who underwent

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3T MRI). DCE images were used for both networks. After training the networks, the models were fine-tuned on the first test set and evaluated on the second, and vice versa. Using transfer learning, the CNN model achieved accuracies of 91% and 78% on the first and second test sets, respectively. The CLSTM model performed slightly worse, with accuracies of 83% and 74% on the same test sets. Meanwhile, Yin et al. (46) in 2022 reported the use of CNNs based on preoperative multiparametric MRI to predict breast cancer molecular subtypes (luminal A, luminal B, HER2-enriched, and triple negative). A modified pre-trained ResNet18 architecture was used to develop CNN models based on post-contrast T1-weighted images, T2-weighted images, and ADC maps, respectively. 136 patients were divided into training, validation, and testing sets (6:2:2 ratio). The models based on post-contrast T1-weighted images achieved the highest AUCs (AUCs of 0.76–0.92), outperforming both the ADC-based models (AUCs of 0.69–0.85) and the T2-weighted-based models (AUCs 0.64–0.70). T1-weighted imaging-based models were especially good for separating HER2-enriched and triple negative cancers, probably as these cancers are highly vascularized and show high contrast enhancement early in the post-contrast phase.

Use Case: Prediction of Treatment Response

There is a wide range of treatment strategies for breast cancer, each having its own unique combination of benefits, side effects, and costs. Moreover, the efficacy of each treatment strategy can vary significantly between individuals. In this context, it is highly advantageous to predict a patient's response to a specific treatment strategy before treatment is even initiated. Several studies have shown promising results for using breast MRI to predict response to neoadjuvant chemotherapy (NAC) in particular. NAC is a widely used treatment strategy that is followed by surgery to remove remaining tumor tissue. Regarding NAC, it is clinically relevant to predict whether a patient will reach pathologic complete response (pCR) prior to (or in an early stage of) treatment. In recent years, there has been major interest in investigating whether AI can be employed to predict treatment response, with promising results. Common limitations across these studies include the limited size of the study populations and the lack of data specific to different breast cancer subtypes. Additionally, many studies lack specific details regarding the NAC regimens administered to patients. Further, the majority of studies incorporate only limited sequences from MRI studies, suggesting that the inclusion of additional sequences as well as clinical data may enhance prediction. Table 4 summarizes these studies.

Several studies have shown that radiomic features from MRI are promising to predict the likelihood of a patient's response to NAC, and we have included a selection of those studies here (39,47–53). In 2019, Liu et al. (47) reported using radiomic features from pre-treatment multiparametric MRI data to train SVM models to predict pCR. The model based on pre-treatment multiparametric MRI (combining T2-weighted imaging, ADC maps, and post-contrast T1-weighted imaging) achieved an AUC of 0.79 for predicting pCR on the primary cohort (n = 128 patients) of a multicenter study, outperforming the models based on T2-weighted imaging alone, ADC maps alone, and post-contrast T1-weighted imaging alone (AUCs of 0.69, 0.69, and 0.64), respectively. This model exhibited slightly degraded but still reasonable performance on the three validation cohorts (n = 99, 107, and 80 patients),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

attaining AUCs of 0.70, 0.68, and 0.79, respectively. When the best radiomic signature was combined with clinicopathologic risk factors, the AUC improved to 0.86 on the primary cohort and 0.79, 0.71, and 0.80 on the external validation cohorts.

In 2022, Umutlu et al. (55) reported results from their study involving 73 patients with breast cancer, using ^{18}F-FDG PET/MRI radiomics analysis of the breast. The SVM model combining pre-treatment data from all MRI sequences and PET data achieved the best AUC of 0.8 with a negative predictive value of 79.5%. In subgroup analysis (HR+/HER2− patients and TN/HER2+ patients analyzed separately), the highest AUC of 0.94 was observed when combining data from all MRI sequences and PET data in the HR+/HER2− group.

Also in 2022, Gilad and Freiman (48) introduced a novel physiologically decomposed DWI machine learning model incorporating both radiomic features extracted from pre-treatment 3D DWI scans as well as clinical data, achieving an AUC of 0.79 for predicting pCR to NAC, beating the previous best AUC of 0.78 on the leaderboard in response to the Breast Multi-parametric MRI for prediction of NAC Response (BMMR2) challenge.

In terms of studies using deep learning approaches, both 2D and 3D CNN architectures have proven effective for predicting pCR to NAC (56–62).

In 2018, Ha et al. (63) reported building a 3D CNN based on the first post-contrast T1-weighted sequence of pre-treatment MRI to predict axillary response to NAC (axillary pCR vs axillary non-pCR). Patients with breast cancer who were included in the study (n = 127) were randomly divided into a training set and a test set (80:20 ratio). The training dataset was further subdivided into five folds, ensuring class balance, for cross-validation during algorithm training. The CNN consisted of ten convolutional layers, four max-pooling layers, and dropout of 50% after a fully connected layer to avoid overfitting. The CNN achieved an overall accuracy of 83%, a sensitivity of 93%, a specificity of 77%, and an AUC of 0.93. In 2019, the same group of researchers (64) reported a similar 3D CNN to perform three-class prediction of pCR (pCR, partial response, and no response/progression). Data from 141 patients with locally advanced breast cancer were divided into a training and test set (80:20 ratio). The 3D CNN achieved an overall accuracy of 88% and an AUC of 0.98 in three-class prediction of neoadjuvant treatment axillary response on the test set. The two studies show great performance increases not found in works with similar approaches and similarly sized datasets.

In 2020, Liu et al. (57) showed that their 2D CNN based on first post-contrast images predicted pCR based on pre-treatment MRI with an AUC of 0.72 in 131 patients from the public I-SPY TRIAL breast MRI dataset. The authors mentioned the need to investigate incorporating clinical information into the algorithm to improve performance.

In 2021, Joo et al. (56) reported results from their study that included 536 patients with invasive breast cancer. They were interested to see if combining clinical information with pre-treatment MRI data would improve performance. Their model, combining clinical information and imaging-derived 3D information from pre-treatment T1-weighted subtraction images and T2-weighted images achieved an AUC of 0.89 to predict pCR on the validation set, compared to an AUC of 0.83 when only the clinical information was used.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In 2021, Comes et al. (58) constructed a hybrid model, where a classical machine learning algorithm was trained on clinical data (e.g., ER, PR, HER2, and molecular subtype) and a combination of pre- and early-treatment 2D image features extracted by a pre-trained CNN, achieving AUC values of 0.93 and 0.90 and accuracy values of 91.4% and 92.3% on the fine-tuning dataset and the independent test, respectively.

Of note, a direct comparison between deep learning and classical machine learning methods was made by Peng et al. (62), whereby a 2D CNN (AUC of 0.83) was found to slightly outperform several classical machine learning models based on radiomic features (best AUC of 0.78).

Use Case: Risk Assessment

Currently, breast MRI plays an essential role in screening women at high risk of developing breast cancer. Additionally, apart from knowing the risk of developing breast cancer, knowing the risk of breast cancer recurrence after treatment is also valuable for choosing the optimal treatment strategy and for determining the post-treatment monitoring schedule. Below, we elaborate on studies that have applied AI to breast MRI for risk assessment. Table 5 summarizes these studies.

Predicting the risk of developing breast cancer—In 2019, Saha et al. (65) reported results from their retrospective study involving 133 high-risk women, where they assessed if machine learning models could predict breast cancer in the subsequent two years following screening breast MRI. The machine learning model based on BPE features extracted from the non-fat-saturated sequence of screening breast MRI achieved an AUC of 0.70, outperforming the machine learning model based on BPE features extracted from the fat-saturated MRI sequence (AUC of 0.63) as well as radiologist subjective evaluation of BPE (0.59–0.60).

Also in 2021, Portnoi et al. (15) reported developing a deep learning model to predict 5-year breast cancer risk based on a 2D projection of a single 3D volumetric image (axial view of the MIP) from screening breast MRI. In their study, which involved 1,183 high-risk women, the deep learning model achieved a mean AUC of 0.64 ± 0.09, outperforming the traditional Tyrer–Cuzick risk model which achieved an AUC of 0.49 ± 0.09.

Meanwhile, in 2022, Eskreis-Winkler et al. (66) reported automated deep learning models based on MRI slabs and MRI MIPs (Figure 5) separately for BPE classification using standard-of-care qualitative BPE assessment from subjective radiologist review in high-risk screening patients, with the slab model outperforming the MIP model (AUCs of 0.84 and 0.79, respectively). In the reader study, the slab model also significantly outperformed subjective radiologist review for BPE assessment.

Predicting the risk of breast cancer recurrence—Several multigene assays (e.g., MammaPrint (67) and Oncotype DX® (68)) are available for assessing breast cancer recurrence risk, but they are typically costly and are not available for every patient. Breast MRI has been shown to be helpful for predicting Oncotype Dx recurrence scores (RS),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

thereby having the potential to either identify patients that should undergo the assay or even serving as a proxy to estimate breast cancer recurrence risk (69–71).

In 2016, Li et al. (72) reported results from their study involving 84 women that MRI-based radiomic signatures exhibited significant associations with the recurrence scores of four types of multigene assays: MammaPrint (R^2 = 0.30, r = 0.55), Oncotype DX (R^2 = 0.25, r = 0.5), PAM50 ROR-S (R^2 = 0.32, r = 0.56), and PAM50 ROR-P (R^2 = 0.28, r = 0.53) (p < 0.0001 for all). Additionally, when radiomic signatures were applied to logistic regression classifiers as the decision variable in a binary classification task to distinguish between good and bad prognoses (defined by thresholding the recurrence scores for each multigene assay), the classifiers achieved AUC values of 0.88 ± 0.05, 0.76 ± 0.06, and 0.68 ± 0.08 based on thresholding MammaPrint, Oncotype DX, PAM50 ROR-S RS, respectively.

Similarly, in 2022, Fan et al. (73) reported results from their study involving 130 women that an elastic net regression model based on eleven MRI-based radiogenomic features was significantly associated with the recurrence scores of Oncotype DX (R^2 = 0.33) (p < 0.001). When the predicted recurrence scores by the elastic net regression model were assessed for associations with survival (in a separate prognostic assessment dataset comprising 116 women), predicted recurrence scores > 29.9 were linked to both adverse overall and recurrence-free survival outcomes (p = 0.001 and p = 0.007, respectively). When the predicted recurrence scores by the elastic net regression model were assessed for associations with treatment response to NAC (in a separate treatment assessment dataset comprising 135 women), women who exhibited a favorable response to NAC had higher predicted recurrence scores (30.51 ± 6.92) compared to non-responders (27.35 ± 4.04) (p = 0.001); notably, to predict response to NAC, a model combining the predicted recurrence scores with clinical features and other imaging features had the best performance (AUC = 0.85) (p < 0.001).

While Li et al. and Fan et al.'s studies relied on a classical machine learning approach to predict breast cancer recurrence scores, Ha et al. (74) employed a 2D CNN on tumor-containing slices of 3D MRI scans to classify patients (n = 134) into three Oncotype DX risk groups: low risk (recurrence scores < 18), intermediate risk (recurrence scores = 18–30), and high risk (recurrence scores > 30). The 2D CNN model trained for three-class prediction achieved a mean accuracy of 81% (95% CI ± 4%), with an AUC of 0.92 (SD ± 0.01), whereas the 2D CNN model trained for two-class prediction (low risk vs. moderate and high risk) achieved a mean accuracy of 87% (95% CI ± 5%) with an AUC of 0.92 (SD ± 0.01).

Use Case: Image Quality

Clinicians need high-quality breast MRI examinations to make reliable and effective diagnoses. Machine learning can be used to assess or improve the quality of MRI scans, in some cases allowing for reduced acquisition times or higher diagnostic accuracies. We discuss a selection of recent work in this domain. We group these works into three categories, namely 1) image quality assessment, 2) accelerated image reconstruction, and 3) generating contrast-enhanced images from non-contrast MRI data. The full list of relevant literature is summarized in Table 6 and selected studies are discussed below.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Image quality assessment—As DWI sequences can be prone to image artifacts, Kapsner et al. (75) trained a CNN to detect artifacts on high b-value MIPs derived from DWI scans, attaining an AUC of 0.91 on the test set. Meanwhile, Allen et al. (76) applied deep learning reconstruction techniques to improve the quality of T2-weighted images (assessed via signal-to-noise ratio, contrast-to-noise ratio, and edge sharpness), at the expense of a 20% increase in acquisition time.

Accelerated image reconstruction—MRI scanners acquire data in the *frequency domain* (also referred to as *k-space*) and images are obtained from these data through a process called *reconstruction*. Acquisition time can be reduced by undersampling in the frequency domain, although this leads to reduced image quality. Both traditional "compressed sensing" and deep learning have proven effective at reconstructing high-quality images from strongly undersampled k-space data. The following deep learning applications all make use of the variational network (VN) architecture proposed in (77). In a restrospective study by Sauer et al. (78), the VN approach enabled a simulated 39% reduction in DWI acquisition time. Images reconsctructed using the VN approach had excellent structural similarity to standard DWI images. Additionally, VN and an additional super-resolution algorithm had the highest subjective image quality scores. Similar results were found in a prospective study by Lee et al. (79), who showed that the VN approach enabled a 47.2% reduction in acquisition time. Images reconstructed with the VN approach outperformed standard images from single-short echo planar imaging (ss-EPI) for DWI on both quantitative analysis and a subjective evaluation by two radiologists. In a prospective study by Wilpert et al. (80), 65 patients underwent ss-EPI for DWI (average 5:02 minutes) followed by deep learning accelerated DWI (average 2:44 minutes), witth deep learning accelerated DWI improving lesion conspicuity and maintaining overall image quality.

Contrast-enhanced image generation—Whereas administration of a gadolinium-based contrast agent can improve theperformance of breast MRI, this is not without drawbacks (81). Tsui et al. (82) in their meta-analysis described multiple machine learning approaches to generate synthetic post-contrast images. Recently, GANs have also been employed to generate contrast-enhanced images from non-contrast MRI data. To this end, Chung et al. (83) trained a 3D CNN to simulate contrast-enhanced T1-weighted imaging with five pre-contrast sequences as input. Per four breast radiologists, almost all simulated images were of diagnostic quality, and quantitative analysis showed strong similarity between simulated and real images. Elsewhere, Kim et al. (84) used a GAN to generate T1-weighted contrast-enhanced images from pre-contrast T1-weighted images. The resulting images exhibited a structural similarity index map of 0.86 when evaluated on the tumor region of interest, outperforming earlier approaches by a small amount.

Barriers to the Adoption of AI in Breast MRI

A recent scoping review analyzed the barriers to the conception and implementation of AI in breast imaging (85). An obvious barrier is the lack of well annotated publicly available datasets; this pertains especially to supervised learning which requires a lot of data with annotations and labels provided by experts in the field. Another barrier related to datasets is that many are based on data from a single institution, hampering generalizability.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Some potential solutions to these challenges are to use diverse data sources from multiple institutions at the outset or perform external validation after initial results. Recently, there has been a growing number of publicly available datasets (2) to facilitate data diversity and allow external validation. Validation with clinical trials, ideally randomized clinical trials, is also needed, whereby initial accurate predictions can be confirmed in large and heterogeneous populations to validate and verify the clinical effectiveness of the AI model. This is important to identify unsuspected biases in the AI system and confirm that the AI system does in fact improve clinical outcomes.

Notably, for breast MRI, standardization is particularly challenging and encompasses many aspects including hardware, imaging protocols, imaging parameters, post-processing, and image analysis methods. However, solutions to overcome standardization issues have been proposed (71) and there is a move towards deep learning models that have shown greater robustness across different MRI scanners.

Workflow integrations also presents a challenge (86). While several guidelines have described good practices for conducting and reporting AI-based research, there are few guidelines on workflow integration, e.g., “how well does the AI product integrate into the the existing IT system?” or “how does it impact the clinical workflow?” Integrating an AI product into the institutional picture and archiving communications system may need a significant redesign to support the AI product and it is important to know ahead of time if the information technology team and the AI vendor can support a redesign.

Additional issues (87) relate to continuous learning, AI systems can be brittle and fail to operate effectively with newly acquired or updated devices, different patient demographics, or with a different disease.

Ultimately, the integration of AI into clinical practice must address barriers related to ethical and legal concerns, as well as human factors. This includes understanding how humans contribute input to train the algorithm and how they interpret the resulting output. A study by Dratsh et al. (88) defined background automation bias as “the propensity for humans to favor suggestions from automated decision-making systems.” In this prospective study, 27 radiologists read 50 mammograms with AI assistance. The percentage of correctly rated mammograms by inexperienced, moderately experienced and very experienced radiologists was significantly impacted by the correctness of the AI prediction. Inexperienced radiologists were more likely to follow the incorrect suggestions provided by the AI system compared with moderately and very experienced readers, although all readers were prone to automation bias. This effect of human and machine interaction must be considered to ensure safe deployment and accurate diagnostic performance when combining human readers and AI. A significant question arises concerning the accountability of the AI developer for these decisions, and who bears liability if an AI-based device errs is still unknown. Other ethical-legal considerations are related to privacy and and confidentiality asa large amount of sensitive data might either be harvested illicitly or collected from unknown sources because of the lack of unique and clear regulations (89).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Future Directions in the Field of AI

While AI has made huge progress in several areas such as language translation, speech recognition, and natural image recognition, there is still a lot of progress that needs to be made in the medical field.

Most of the current applications of AI in medicine have addressed narrowly defined tasks using data from a single modality, for example CT or MRI. This does not reflect the real world where clinicians analyze data from multiple sources and modalities (diagnostic scans, lab tests, clinical history, genetic testing) when making a diagnosis, making a prognostic evaluation, or when deciding the best treatment plan (93). The AI model, in order to function optimally, should also be able to use and analyze all data sources that are available to the clinician. Although the integration of such diverse types of data remains challenging, the advancement of multimodal AI models, which integrate data from various modalities, holds the potential to partially address this challenge. The integration of multimodal data has demonstrated better diagnostic performance across a range of different tasks, e.g., to identify pulmonary embolism (94) and to differentiate between common causes of acute respiratory failure (95).

In the last few years, there has been a transition from architectures with strong modality-specific biases—e.g., CNNs for images, or recurrent neural networks for text and physiological signals—to a relatively novel architecture called the Transformer, which has demonstrated good performance across a wide variety of input and output modalities and tasks (96).

Ultimately, including other data such as polygenic risk scores, cancer predisposition gene variants, circulating cell-free tumor DNA results, AI results of prior diagnostic studies, and other data within the electronic health record may increase the performance of AI in the detection of breast cancer and allow precision medicine.

Conclusion

This review provides a summary of several clinically relevant use cases for AI in breast MRI, presenting the available results from the literature to date for each use case. While barriers to the adoption of AI in healthcare exist, several solutions have been identified to overcome these barriers. Furthermore, with a strong push not only from the imaging and AI community but also other stakeholders such as referring physicians, patients, referrers, and payers, it can be expected that we will soon be able to derive value from AI-enhanced solutions in multiple healthcare scenarios, with breast MRI being no exception.

Acknowledgments:

The authors thank Joanne Chin, MFA, ELS, for editing this manuscript.

Grant Support:

This work was partially supported by the NIH/NCI Cancer Center Support Grant P30 CA008748.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Source of Funding:

All authors report no industry support of the project. Katja Pinker received payment for activities not related to the present article including lectures including service on speakers bureaus and for travel/ accommodations/meeting expenses unrelated to activities listed from the European Society of Breast Imaging (MRI educational course, annual scientific meeting).

References

1. Bitencourt A, Daimiel Naranjo I, Lo Gullo R, Rossi Saccarelli C, Pinker K. AI-enhanced breast imaging: Where are we and where are we heading? Eur J Radiol 2021;142:109882. [PubMed: 34392105]
2. Dishner KA, McRae-Posani B, Bhowmik A, et al. A Survey of Publicly Available MRI Datasets for Potential Use in Artificial Intelligence Research. J Magn Reson Imaging 2023.
3. Balkenende L, Teuwen J, Mann RM. Application of Deep Learning in Breast Cancer Imaging. Semin Nucl Med 2022;52(5):584–596. [PubMed: 35339259]
4. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. Radiology 2019;292(3):520–536. [PubMed: 31361209]
5. Saadatmand S, Geuzinge HA, Rutgers EJT, et al. MRI versus mammography for breast cancer screening in women with familial risk (FaMRIsc): a multicentre, randomised, controlled trial. Lancet Oncol 2019;20(8):1136–1147. [PubMed: 31221620]
6. Bakker MF, de Lange SV, Pijnappel RM, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. N Engl J Med 2019;381(22):2091–2102. [PubMed: 31774954]
7. Tilanus-Linthorst MM, Obdeijn IM, Bartels KC, de Koning HJ, Oudkerk M. First experiences in screening women at high risk for breast cancer with MR imaging. Breast Cancer Res Treat 2000;63(1):53–60. [PubMed: 11079159]
8. Hickman SE, Baxter GC, Gilbert FJ. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. Br J Cancer 2021;125(1):15–22. [PubMed: 33772149]
9. 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. [PubMed: 35576821]
10. Xie Z, He F, Fu S, Sato I, Tao D, Sugiyama M. Artificial Neural Variability for Deep Learning: On Overfitting, Noise Memorization, and Catastrophic Forgetting. Neural Comput 2021;33(8):2163–2192. [PubMed: 34310675]
11. Romero M, Interian Y, Solberg T, Valdes G. Targeted transfer learning to improve performance in small medical physics datasets. Med Phys 2020;47(12):6246–6256. [PubMed: 33007112]
12. Grill J-B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning. Advances in neural information processing systems 2020;33:21271–21284.
13. Chen T, Kornblith S, Norouzi M, Hinton G. A Simple Framework for Contrastive Learning of Visual Representations. In: Hal D III, Aarti S, editors. Proceedings of the 37th International Conference on Machine Learning. Volume 119. Proceedings of Machine Learning Research: PMLR; 2020. p. 1597--1607.
14. Zhou J, Luo LY, Dou Q, et al. Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. J Magn Reson Imaging 2019;50(4):1144–1151. [PubMed: 30924997]
15. Portnoi T, Yala A, Schuster T, et al. Deep Learning Model to Assess Cancer Risk on the Basis of a Breast MR Image Alone. AJR Am J Roentgenol 2019;213(1):227–233. [PubMed: 30933651]
16. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. Part of Advances in Neural Information Processing Systems 27. In: Ghahramani Z, Weilling M, Cortes C, Lawrence N, Weinberger KQ, editors. 28th Annual Conference on Neural Information Processing Systems 2014. Montreal, Canada; 2014.
17. Kuhl CK, Schrading S, Strobel K, Schild HH, Hilgers RD, Bieling HB. Abbreviated breast magnetic resonance imaging (MRI): first postcontrast subtracted images and maximum-intensity

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

projection-a novel approach to breast cancer screening with MRI. J Clin Oncol 2014;32(22):2304–2310. [PubMed: 24958821]

18. Comstock CE, Gatsonis C, Newstead GM, et al. Comparison of Abbreviated Breast MRI vs Digital Breast Tomosynthesis for Breast Cancer Detection Among Women With Dense Breasts Undergoing Screening. Jama 2020;323(8):746–756. [PubMed: 32096852]
19. Kuhl CK, Strobel K, Bieling H, Leutner C, Schild HH, Schrading S. Supplemental Breast MR Imaging Screening of Women with Average Risk of Breast Cancer. Radiology 2017;283(2):361–370. [PubMed: 28221097]
20. Lee JM, Ichikawa L, Valencia E, et al. Performance Benchmarks for Screening Breast MR Imaging in Community Practice. Radiology 2017;285(1):44–52. [PubMed: 28582633]
21. Backmann HA, Larsen M, Danielsen AS, Hofvind S. Does it matter for the radiologists' performance whether they read short or long batches in organized mammographic screening? Eur Radiol 2021;31(12):9548–9555. [PubMed: 34110427]
22. Jing X, Wielema M, Cornelissen LJ, et al. Using deep learning to safely exclude lesions with only ultrafast breast MRI to shorten acquisition and reading time. Eur Radiol 2022;32(12):8706–8715. [PubMed: 35614363]
23. Verburg E, van Gils CH, van der Velden BHM, et al. Deep Learning for Automated Triaging of 4581 Breast MRI Examinations from the DENSE Trial. Radiology 2022;302(1):29–36. [PubMed: 34609196]
24. Bhowmik A, Monga N, Belen K, et al. Automated Triage of Screening Breast MRI Examinations in High-Risk Women Using an Ensemble Deep Learning Model. Invest Radiol 2023;58(10):710–719. [PubMed: 37058323]
25. Madrzyk A Artificial intelligence software for breast cancer diagnosis makes TIME's list of Best Inventions for 2019. The Forefront, University of Chicago Medicine. 2019.
26. Zhao X, Bai JW, Guo Q, Ren K, Zhang GJ. Clinical applications of deep learning in breast MRI. Biochim Biophys Acta Rev Cancer 2023;1878(2):188864. [PubMed: 36822377]
27. Sheth D, Giger ML. Artificial intelligence in the interpretation of breast cancer on MRI. J Magn Reson Imaging 2020;51(5):1310–1324. [PubMed: 31343790]
28. Herent P, Schmauch B, Jehanno P, et al. Detection and characterization of MRI breast lesions using deep learning. Diagn Interv Imaging 2019;100(4):219–225. [PubMed: 30926444]
29. Antropova N, Abe H, Giger ML. Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. J Med Imaging (Bellingham) 2018;5(1):014503. [PubMed: 29430478]
30. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-enhancing Lesions at Multiparametric Breast MRI. Radiology 2019;290(2):290–297. [PubMed: 30422086]
31. Adachi M, Fujioka T, Mori M, et al. Detection and Diagnosis of Breast Cancer Using Artificial Intelligence Based assessment of Maximum Intensity Projection Dynamic Contrast-Enhanced Magnetic Resonance Images. Diagnostics (Basel) 2020;10(5).
32. Dalmiş MU, Gubern-Mérida A, Vreemann S, et al. Artificial Intelligence-Based Classification of Breast Lesions Imaged With a Multiparametric Breast MRI Protocol With Ultrafast DCE-MRI, T2, and DWI. Invest Radiol 2019;54(6):325–332. [PubMed: 30652985]
33. Fujioka T, Yashima Y, Oyama J, et al. Deep-learning approach with convolutional neural network for classification of maximum intensity projections of dynamic contrast-enhanced breast magnetic resonance imaging. Magn Reson Imaging 2021;75:1–8. [PubMed: 33045323]
34. Hu Q, Whitney HM, Li H, Ji Y, Liu P, Giger ML. Improved Classification of Benign and Malignant Breast Lesions Using Deep Feature Maximum Intensity Projection MRI in Breast Cancer Diagnosis Using Dynamic Contrast-enhanced MRI. Radiol Artif Intell 2021;3(3):e200159. [PubMed: 34235439]
35. Zhu J, Geng J, Shan W, et al. Development and validation of a deep learning model for breast lesion segmentation and characterization in multiparametric MRI. Front Oncol 2022;12:946580. [PubMed: 36033449]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

36. Wang L, Chang L, Luo R, et al. An artificial intelligence system using maximum intensity projection MR images facilitates classification of non-mass enhancement breast lesions. Eur Radiol 2022;32(7):4857–4867. [PubMed: 35258676]
37. Witowski J, Heacock L, Reig B, et al. Improving breast cancer diagnostics with deep learning for MRI. Sci Transl Med 2022;14(664):eabo4802. [PubMed: 36170446]
38. Cong C, Li X, Zhang C, et al. MRI-Based Breast Cancer Classification and Localization by Multiparametric Feature Extraction and Combination Using Deep Learning. J Magn Reson Imaging 2024;59(1):148–161. [PubMed: 37013422]
39. Bitencourt AGV, Gibbs P, Rossi Saccarelli C, et al. MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer. EBioMedicine 2020;61:103042. [PubMed: 33039708]
40. Leithner D, Horvat JV, Marino MA, et al. Radiomic signatures with contrast-enhanced magnetic resonance imaging for the assessment of breast cancer receptor status and molecular subtypes: initial results. Breast Cancer Res 2019;21(1):106. [PubMed: 31514736]
41. Lo Gullo R, Wen H, Reiner JS, et al. Assessing PD-L1 Expression Status Using Radiomic Features from Contrast-Enhanced Breast MRI in Breast Cancer Patients: Initial Results. Cancers (Basel) 2021;13(24).
42. Leithner D, Mayerhoefer ME, Martinez DF, et al. Non-Invasive Assessment of Breast Cancer Molecular Subtypes with Multiparametric Magnetic Resonance Imaging Radiomics. J Clin Med 2020;9(6).
43. Zhu Z, Harowicz M, Zhang J, et al. Deep learning analysis of breast MRIs for prediction of occult invasive disease in ductal carcinoma in situ. Comput Biol Med 2019;115:103498. [PubMed: 31698241]
44. Moroianu S, Rusu M. Detecting invasive breast carcinoma on dynamic contrast-enhanced MRI: SPIE: 2021.
45. Zhang Y, Chen JH, Lin Y, et al. Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. Eur Radiol 2021;31(4):2559–2567. [PubMed: 33001309]
46. Yin H, Bai L, Jia H, Lin G. Noninvasive assessment of breast cancer molecular subtypes on multiparametric MRI using convolutional neural network with transfer learning. Thorac Cancer 2022;13(22):3183–3191. [PubMed: 36203226]
47. Liu Z, Li Z, Qu J, et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. Clin Cancer Res 2019;25(12):3538–3547. [PubMed: 30842125]
48. Gilad M, Freiman M. PD-DWI: Predicting Response to Neoadjuvant Chemotherapy in Invasive Breast Cancer with Physiologically-Decomposed Diffusion-Weighted MRI Machine-Learning Model. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Cham: Springer Nature Switzerland; 2022. p. 36–45.
49. Drukker K, Edwards A, Doyle C, Papaioannou J, Kulkarni K, Giger ML. Breast MRI radiomics for the pretreatment prediction of response to neoadjuvant chemotherapy in node-positive breast cancer patients. J Med Imaging (Bellingham) 2019;6(3):034502. [PubMed: 31592438]
50. Jahani N, Cohen E, Hsieh MK, et al. Prediction of Treatment Response to Neoadjuvant Chemotherapy for Breast Cancer via Early Changes in Tumor Heterogeneity Captured by DCE-MRI Registration. Sci Rep 2019;9(1):12114. [PubMed: 31431633]
51. Hwang KP, Elshafeey NA, Kotrotsou A, et al. A Radiomics Model Based on Synthetic MRI Acquisition for Predicting Neoadjuvant Systemic Treatment Response in Triple-Negative Breast Cancer. Radiol Imaging Cancer 2023;5(4):e230009. [PubMed: 37505106]
52. Gan L, Ma M, Liu Y, et al. A Clinical-Radiomics Model for Predicting Axillary Pathologic Complete Response in Breast Cancer With Axillary Lymph Node Metastases. Front Oncol 2021;11:786346. [PubMed: 34993145]
53. Tahmassebi A, Wengert GJ, Helbich TH, et al. Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Chemotherapy and Survival Outcomes in Breast Cancer Patients. Invest Radiol 2019;54(2):110–117. [PubMed: 30358693]

54. Cardoso MR, Silva AAR, Talarico MCR, et al. Metabolomics by NMR Combined with Machine Learning to Predict Neoadjuvant Chemotherapy Response for Breast Cancer. Cancers (Basel) 2022;14(20).
55. Umutlu L, Kirchner J, Bruckmann NM, et al. Multiparametric (18)F-FDG PET/MRI-Based Radiomics for Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer. Cancers (Basel) 2022;14(7).
56. Joo S, Ko ES, Kwon S, et al. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Sci Rep 2021;11(1):18800. [PubMed: 34552163]
57. Liu MZ, Mutasa S, Chang P, Siddique M, Jambawalikar S, Ha R. A novel CNN algorithm for pathological complete response prediction using an I-SPY TRIAL breast MRI database. Magn Reson Imaging 2020;73:148–151. [PubMed: 32889091]
58. Comes MC, Fanizzi A, Bove S, et al. Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast DCE-MRIs. Sci Rep 2021;11(1):14123. [PubMed: 34238968]
59. Duanmu H, Huang PB, Brahmavar S, et al. Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Deep Learning with Integrative Imaging, Molecular and Demographic Data. In: Martel AL, Abolmaesumi P, Stoyanov D, et al., editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Cham: Springer International Publishing; 2020. p. 242–252.
60. Kavya R, Nathaniel B, Andrew J, Anant M. A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI. ProcSPIE. Volume 10575; 2018. p. 105750C.
61. Qu YH, Zhu HT, Cao K, Li XT, Ye M, Sun YS. Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning (DL) method. Thorac Cancer 2020;11(3):651–658. [PubMed: 31944571]
62. Peng Y, Cheng Z, Gong C, et al. Pretreatment DCE-MRI-Based Deep Learning Outperforms Radiomics Analysis in Predicting Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer. Front Oncol 2022;12:846775. [PubMed: 35359387]
63. Ha R, Chang P, Karcich J, et al. Predicting Post Neoadjuvant Axillary Response Using a Novel Convolutional Neural Network Algorithm. Ann Surg Oncol 2018;25(10):3037–3043. [PubMed: 29978368]
64. Ha R, Chin C, Karcich J, et al. Prior to Initiation of Chemotherapy, Can We Predict Breast Tumor Response? Deep Learning Convolutional Neural Networks Approach Using a Breast MRI Tumor Dataset. J Digit Imaging 2019;32(5):693–701. [PubMed: 30361936]
65. Saha A, Grimm LJ, Ghate SV, 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):456–464. [PubMed: 30648316]
66. Eskreis-Winkler S, Sutton EJ, D'Alessio D, et al. Breast MRI Background Parenchymal Enhancement Categorization Using Deep Learning: Outperforming the Radiologist. J Magn Reson Imaging 2022;56(4):1068–1076. [PubMed: 35167152]
67. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415(6871):530–536. [PubMed: 11823860]
68. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351(27):2817–2826. [PubMed: 15591335]
69. Zhang M, Sadinski M, Haddad D, et al. Background Parenchymal Enhancement on Breast MRI as a Prognostic Surrogate: Correlation With Breast Cancer Oncotype Dx Score. Front Oncol 2020;10:595820. [PubMed: 33614481]
70. Kim HJ, Choi WJ, Kim HH, Cha JH, Shin HJ, Chae EY. Association between Oncotype DX recurrence score and dynamic contrast-enhanced MRI features in patients with estrogen receptor-positive HER2-negative invasive breast cancer. Clin Imaging 2021;75:131–137. [PubMed: 33548871]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

71. Woodard GA, Ray KM, Joe BN, Price ER. Qualitative Radiogenomics: Association between Oncotype DX Test Recurrence Score and BI-RADS Mammographic and Breast MR Imaging Features. Radiology 2018;286(1):60–70. [PubMed: 28885890]
72. Li H, Zhu Y, Burnside ES, et al. MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays. Radiology 2016;281(2):382–391. [PubMed: 27144536]
73. Fan M, Cui Y, You C, et al. Radiogenomic Signatures of Oncotype DX Recurrence Score Enable Prediction of Survival in Estrogen Receptor-Positive Breast Cancer: A Multicohort Study. Radiology 2022;302(3):516–524. [PubMed: 34846204]
74. Ha R, Chang P, Mutasa S, et al. Convolutional Neural Network Using a Breast MRI Tumor Dataset Can Predict Oncotype Dx Recurrence Score. J Magn Reson Imaging 2019;49(2):518–524. [PubMed: 30129697]
75. Kapsner LA, Balbach EL, Folle L, et al. Image quality assessment using deep learning in high b-value diffusion-weighted breast MRI. Sci Rep 2023;13(1):10549. [PubMed: 37386021]
76. Allen TJ, Henze Bancroft LC, Unal O, et al. Evaluation of a Deep Learning Reconstruction for High-Quality T2-Weighted Breast Magnetic Resonance Imaging. Tomography 2023;9(5):1949–1964. [PubMed: 37888744]
77. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. Magn Reson Med 2018;79(6):3055–3071. [PubMed: 29115689]
78. Sauer ST, Christner SA, Lois AM, et al. Deep Learning k-Space-to-Image Reconstruction Facilitates High Spatial Resolution and Scan Time Reduction in Diffusion-Weighted Imaging Breast MRI. J Magn Reson Imaging 2023.
79. Lee EJ, Chang YW, Sung JK, Thomas B. Feasibility of deep learning k-space-to-image reconstruction for diffusion weighted imaging in patients with breast cancers: Focus on image quality and reduced scan time. Eur J Radiol 2022;157:110608. [PubMed: 36403564]
80. Wilpert C, Neubauer C, Rau A, et al. Accelerated Diffusion-Weighted Imaging in 3 T Breast MRI Using a Deep Learning Reconstruction Algorithm With Superresolution Processing: A Prospective Comparative Study. Invest Radiol 2023;58(12):842–852. [PubMed: 37428618]
81. Sardanelli F, Cozzi A, Trimboli RM, Schiaffino S. Gadolinium Retention and Breast MRI Screening: More Harm Than Good? AJR Am J Roentgenol 2020;214(2):324–327. [PubMed: 31799869]
82. Tsui B, Calabrese E, Zaharchuk G, Rauschecker AM. Reducing Gadolinium Contrast With Artificial Intelligence. J Magn Reson Imaging 2023.
83. Chung M, Calabrese E, Mongan J, et al. Deep Learning to Simulate Contrast-enhanced Breast MRI of Invasive Breast Cancer. Radiology 2023;306(3):e213199. [PubMed: 36378030]
84. Kim E, Cho HH, Kwon J, Oh YT, Ko ES, Park H. Tumor-Attentive Segmentation-Guided GAN for Synthesizing Breast Contrast-Enhanced MRI Without Contrast Agents. IEEE J Transl Eng Health Med 2023;11:32–43. [PubMed: 36478773]
85. Lokaj B, Pugliese MT, Kinkel K, Lovis C, Schmid J. Barriers and facilitators of artificial intelligence conception and implementation for breast imaging diagnosis in clinical practice: a scoping review. Eur Radiol 2023.
86. Omoumi P, Ducarouge A, Tournier A, et al. To buy or not to buy-evaluating commercial AI solutions in radiology (the ECLAIR guidelines). Eur Radiol 2021;31(6):3786–3796. [PubMed: 33666696]
87. Food and Drug Administration. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions: Draft Guidance for Industry and Food and Drug Administration Staff. 2023.
88. Dratsch T, Chen X, Rezazade Mehrizi M, et al. Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. Radiology 2023;307(4):e222176. [PubMed: 37129490]
89. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. Insights Imaging 2018;9(5):745–753. [PubMed: 30112675]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

90. Mazurowski MA. Artificial Intelligence in Radiology: Some Ethical Considerations for Radiologists and Algorithm Developers. Acad Radiol 2020;27(1):127–129. [PubMed: 31818378]
91. Gerke S, Minssen T, Cohen G. Chapter 12 - Ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in Healthcare: Academic Press; 2020. p. 295–336.
92. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14(12):749–762. [PubMed: 28975929]
93. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med 2022;28(9):1773–1784. [PubMed: 36109635]
94. Huang SC, Pareek A, Zamanian R, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. Sci Rep 2020;10(1):22147. [PubMed: 33335111]
95. Jabbour S, Fouhey D, Kazerooni E, Wiens J, Sjoding MW. Combining chest X-rays and electronic health record (EHR) data using machine learning to diagnose acute respiratory failure. J Am Med Inform Assoc 2022;29(6):1060–1068. [PubMed: 35271711]
96. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc.; 2017. p. 6000–6010.
97. Wessling D, Gassenmaier S, Olthof SC, et al. Novel deep-learning-based diffusion weighted imaging sequence in 1.5 T breast MRI. Eur J Radiol 2023;166:110948. [PubMed: 37481831]
98. Müller-Franzes G, Huck L, Tayebi Arasteh S, et al. Using Machine Learning to Reduce the Need for Contrast Agents in Breast MRI through Synthetic Images. Radiology 2023;307(3):e222211. [PubMed: 36943080]
99. Wang P, Nie P, Dang Y, et al. Synthesizing the First Phase of Dynamic Sequences of Breast MRI for Enhanced Lesion Identification. Front Oncol 2021;11:792516. [PubMed: 34950593]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

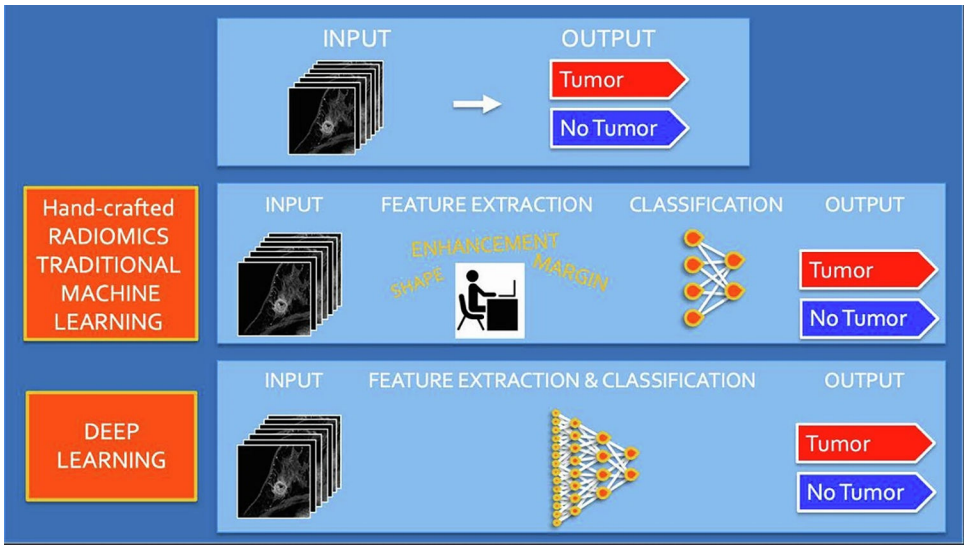

Figure 1.

Radiomics analysis workflow using hand-crafted feature extraction along with traditional machine learning techniques and deep learning for classification and modelling. Reprinted with permission from: Bitencourt A, Daimiel Naranjo I, Lo Gullo R, Rossi Saccarelli C, Pinker K. AI-enhanced breast imaging: Where are we and where are we heading? Eur J Radiol 2021;142:109882.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

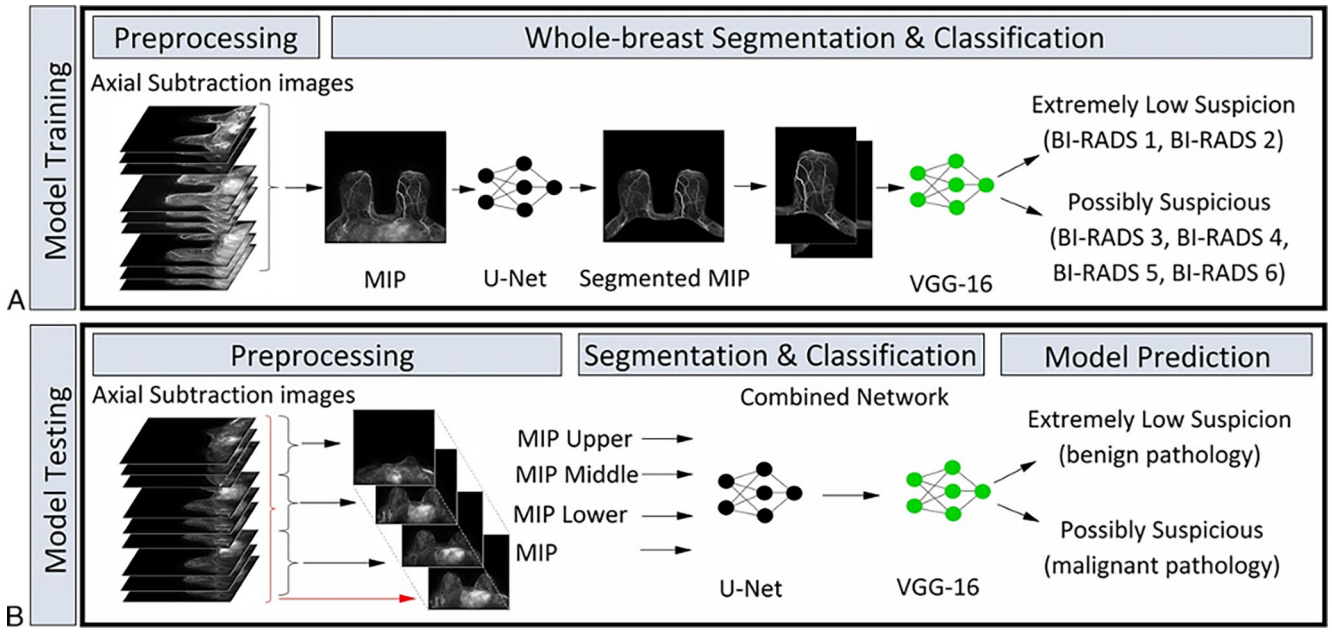

Figure 2.

Flow diagram illustrating deep learning model triaging of breast magnetic resonance imaging examinations into “extremely low suspicion” and “possibly suspicious” examinations. A, Model training. B, Model testing. MIP, maximum intensity projection; BI-RADS, Breast Imaging Reporting and Data System. Reprinted with permission from: Bhowmik A, Monga N, Belen K, et al. Automated Triage of Screening Breast MRI Examinations in High-Risk Women Using an Ensemble Deep Learning Model. Invest Radiol 2023;58(10):710–719.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

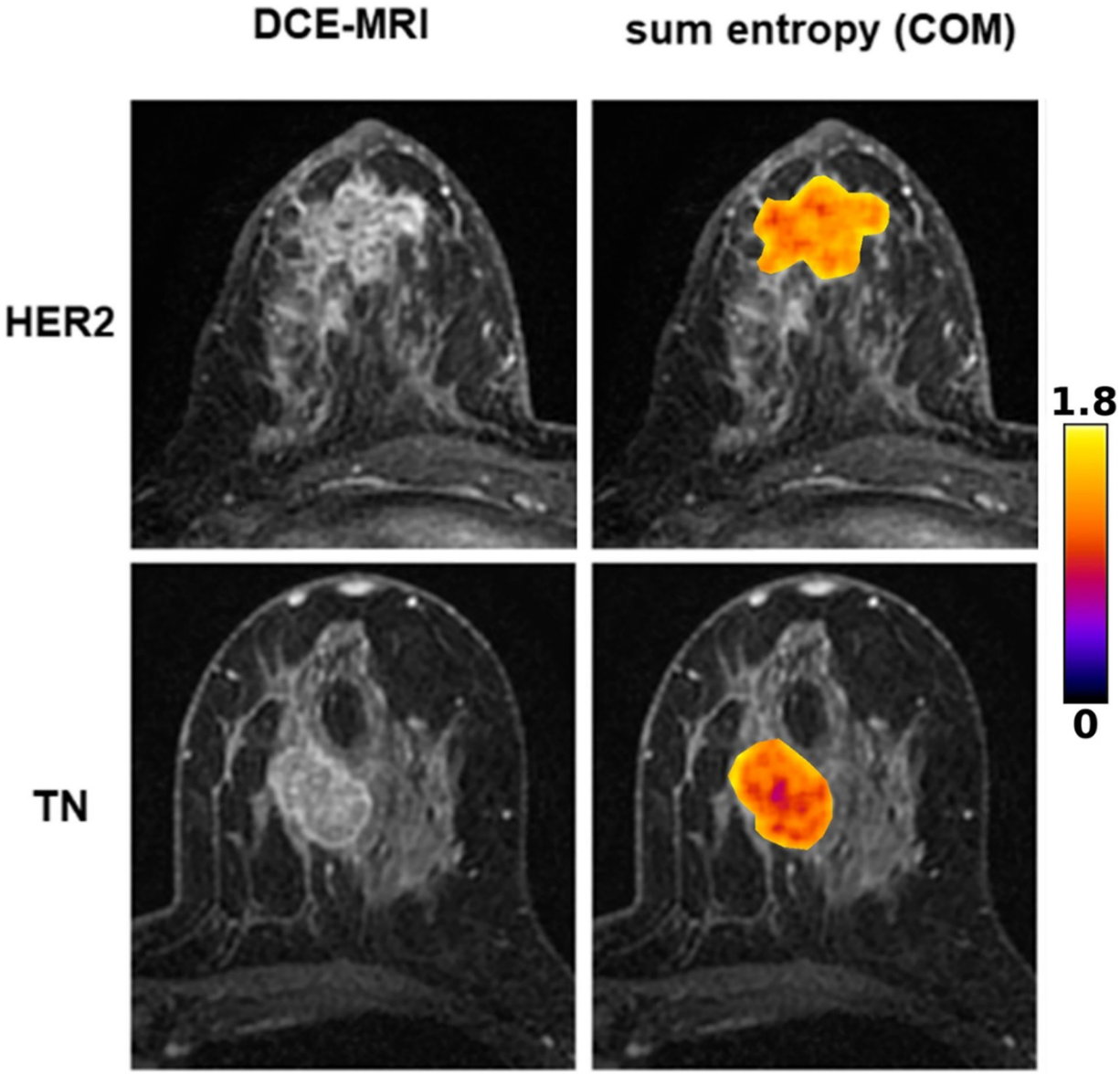

Figure 3.

Original CE-MRI images and corresponding color-coded sum entropy feature map as overlay of the tumor area of triple-negative (TN) and HER2-enriched (HER2) breast cancer. TN shows a clearly lower sum entropy than HER2. Reprinted under the terms of the Creative Commons license (http://creativecommons.org/licenses/by/4.0/) from: Leithner D, Horvat JV, Marino MA, et al. Radiomic signatures with contrast-enhanced magnetic resonance imaging for the assessment of breast cancer receptor status and molecular subtypes: initial results. Breast Cancer Res 2019;21(1):106.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

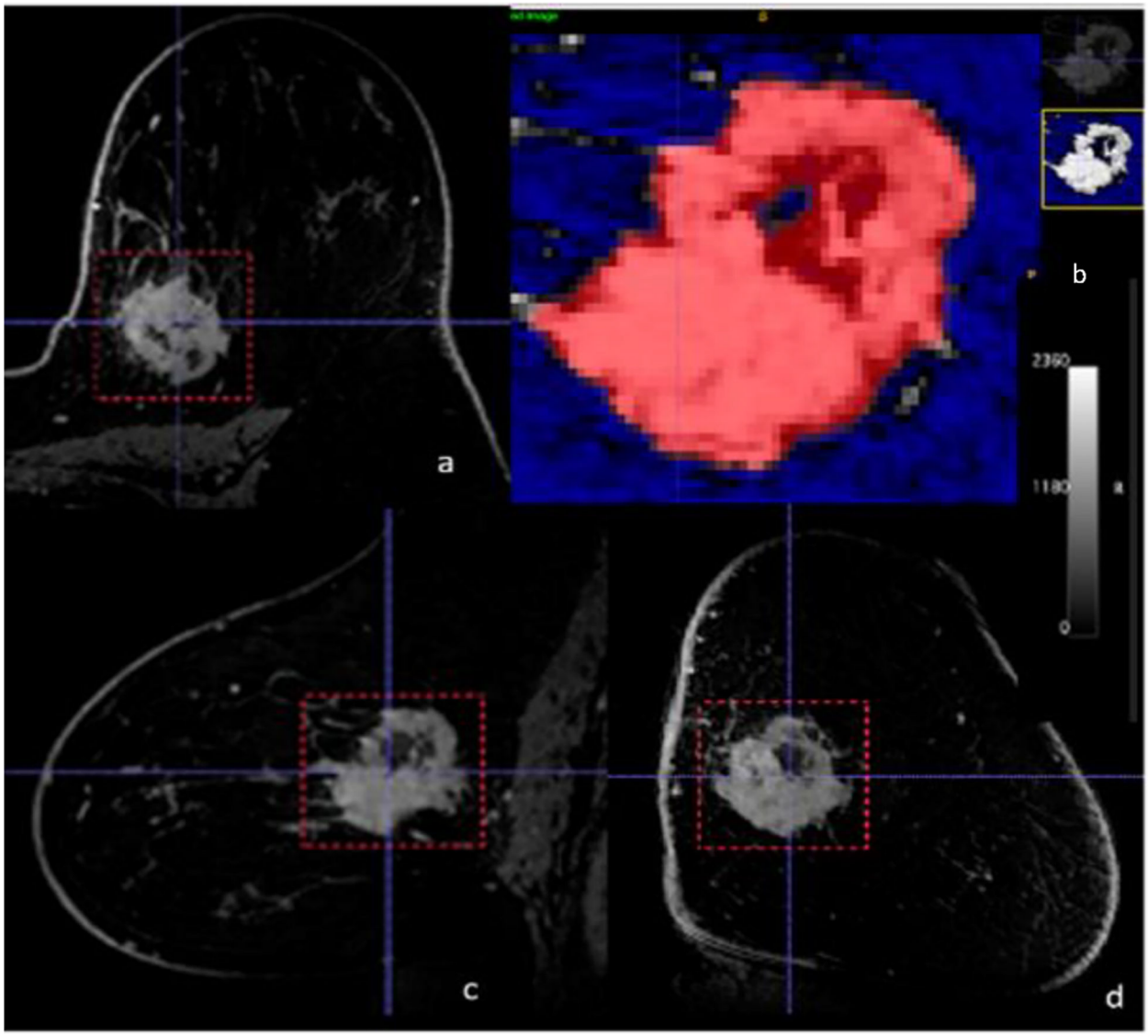

Figure 4.

Example of semi-automatic tumor segmentation with dynamic contrast-enhanced MRI—(a) axial, (b) axial detail, (c) sagittal, (d) coronal—for radiomics analysis in a 43-year-old female with a biopsy-proven PD-L1-positive poorly differentiated triple negative breast cancer in the 10:00 axis of the left breast. Reprinted under the terms of the Creative Commons license (http://creativecommons.org/licenses/by/4.0/) from: Lo Gullo R, Wen H, Reiner JS, et al. Assessing PD-L1 Expression Status Using Radiomic Features from Contrast-Enhanced Breast MRI in Breast Cancer Patients: Initial Results. Cancers (Basel) 2021;13(24).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

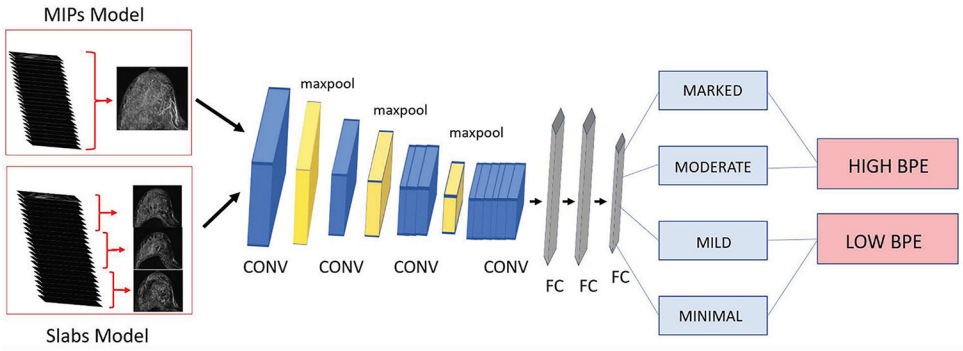

Figure 5.

Artificial intelligence (AI) model architecture schematic. A VGG-19 architecture was trained to classify images into four background parenchymal enhancement (BPE) categories, which were then pooled into “high BPE” and “low BPE” categories. In the maximum intensity projection (MIP) AI model, axial MIPs generated from the first subtraction phase were used as the model input. In the Slab AI model, axial slices from the first subtraction phase were pooled into three maximum intensity slabs, which were each used as an independent model input. CONV = convolutional layer; FC = fully connected layer. Reprinted with permission from: Eskreis-Winkler S, Sutton EJ, D’Alessio D, et al. Breast MRI Background Parenchymal Enhancement Categorization Using Deep Learning: Outperforming the Radiologist. J Magn Reson Imaging 2022;56(4):1068–1076.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Use Case of AI-Enhanced Breast MRI: Breast MRI Triaging and Lesion Detection.

Author(s)	Year	Method	Patient Population Size	MRI Sequences Used	Results
Verburg et al. (23)	2021	CNN	4,581	Pre- and post-contrast T1W DCE	AUC 0.83 (95% CI: 0.80–0.85)
Jing et al. (22)	2022	CNN	438	TWIST	AUC 0.81 (95% CI: 0.75–0.88)
Bhowmik et al. (24)	2023	CNN	8,354	Pre- and post-contrast T1W DCE	Sensitivity 100%, specificity 11.5%

Abbreviations: AUC, area under the curve; CI, confidence interval; CNN, convolutional neural network; DCE, dynamic contrast-enhanced; T1W, T1-weighted; TWIST, time-resolved angiography with stochastic trajectories

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Use Case of AI-Enhanced Breast MRI: Lesion Classification.

Author(s)	Year	Method	Patient Population Size	Results
Antropova et al. (29)	2018	CNN	690	AUC 0.88
Truhn et al. (30)	2019	CNN/radiomics	447	AUC 0.88/AUC 0.81
Dalmiş et al. (32)	2019	CNN	465	AUC 0.852
Herent et al. (28)	2019	CNN	335	AUC 0.893
Zhou et al. (14)	2019	CNN	1537	AUC 0.859
Adachi et al. (31)	2020	CNN	321	AUC 0.925
Fujioka et al. (33)	2021	CNN	286	Best AUC 0.895
Hu et al. (34)	2021	CNN	1979	Best AUC 0.93
Zhu et al. (35)	2022	CNN	2.823	Best AUC 0.885
Wang et al. (36)	2022	CNN	903	AUCs 0.859 & 0.816 (test sets A & B)
Witowski et al. (37)	2022	CNN	21.537	AUCs 0.797–0.969
Cong et al. (38)	2023	CNN	569	AUC 0.91

Abbreviations: AUC, area under the curve; CNN, convolutional neural network

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Use Case of AI-Enhanced Breast MRI: Breast Cancer Characterization.

Author(s)	Year	Method	Patient Population Size	Lesion Types	Results
Leithner et al. (40)	2019	Radiomics + ML	91	Mass 70 NME 21	Best accuracy 79.4%
Zhu et al. (43)	2019	CNN (transfer learning) + SVM	131	N/A	AUC 0.70
Bitencourt et al. (39)	2020	Radiomics + ML	311	Mass 147 NME 34 Mass + NME 130	Accuracy 97.4%
Lo Gullo et al. (41)	2021	Radiomics + ML	62	N/A	Accuracy 88.2%
Moroianu & Rusu (44)	2021	CNN (transfer learning)	81	N/A	AUC 0.83
Zhang et al. (45)	2021	(Recurrent) CNN (transfer learning)	244	Mass 244	Best AUC 0.91
Yin et al. (46)	2022	CNN (transfer learning)	136	N/A	Best AUC 0.92

Abbreviations: AUC, area under the curve; CNN, convolutional neural network; ML, machine learning; SVM support vector machine; NME, non-mass enhancement

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4.

Use Case of AI-Enhanced Breast MRI: Prediction of Treatment Response.

Author(s)	Year	Method	Patient Population Size	Results
Ha et al. (63)	2018	CNN	127	AUC 0.93
Kavya et al. (60)	2018	CNN	166	AUC 0.85
Tahmassebi et al. (53)	2019	Radiomics + XGBoost	38	Best AUC 0.86
Liu et al. (47)	2019	Radiomics + SVM	414	Radiomic signature AUC 0.79, RMM AUC 0.86
Drukker et al. (49)	2019	Radiomics + Linear Discriminant Classifier	158	Best AUC 0.82
Jahani et al. (50)	2019	Radiomics + Logistic Regression	127	AUC 0.78
Bitencourt et al. (39)	2020	Radiomics + Coarse Decision Trees	311	Accuracy 0.839
Ha et al. (64)	2020	CNN	141	AUC 0.98
Liu et al. (57)	2020	CNN	131	AUC 0.72
Duanmu et al. (59)	2020	CNN	112	AUC 0.80
Qu et al. (61)	2020	CNN	302	AUC 0.553 (pre-NAC)
Gan et al. (52)	2021	Radiomics + Logistic Regression	248	Best AUC 0.878
Joo et al. (56)	2021	CNN	536	AUC 0.89
Comes et al. (58)	2021	CNN with SVM (transfer learning)	134	AUC 0.93
Umutlu et al. (55)	2022	Radiomics + SVM	73	AUC 0.8
Gilad & Freiman (48)	2022	Radiomics + XGBoost	191	AUC 0.79
Peng et al. (62)	2022	CNN + ML	356	Best AUC 0.83
Hwang et al. (51)	2023	Radiomics + Logistic Regression	181	AUC 0.72

Abbreviations: CNN, convolutional neural network; NAC, neoadjuvant chemotherapy; RMM, radiomics of multiparametric MRI; SVM, support vector machine

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 5.

Use Case of AI-Enhanced Breast MRI: Risk Assessment.

Author(s)	Year	Method	Patient Population Size	Results
Li et al. (72)	2016	Radiomics + Logistic Regression	84	AUC 0.88 (MammaPrint), 0.76 (Oncotype DX), 0.68 (PAM50)
Woodard et al. (71)	2018	Radiomics + Linear Regression	408	Significant correlations
Ha et al. (74)	2018	CNN	134	AUC 0.92
Saha et al. (65)	2019	Radiomics + logistic regression	133	AUC 0.70
Portnoi et al. (15)	2019	Radiomics + logistic regression & CNN	1183	Best AUC 0.64
Zhang et al. (69)	2020	K-Means Clustering	80	Significant correlations
Kim et al. (70)	2021	Radiomics + Logistic Regression	473	Significant correlations
Fan et al. (73)	2022	Radiomics + Elastic Net Regression	130	R^2 0.33
Eskreis-Winkler et al. (66)	2022	CNN	3,705	Best AUC 0.96

Abbreviations: AUC, area under the curve; CNN, convolutional neural network

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6.

Use Case of AI-Enhanced Breast MRI: Image Quality.

Author(s)	Year	Method	Patient Population Size	Results
Kapsner et al. (75)	2023	CNN	1158	AUC 0.91 (presence of artifacts)
Lee et al. (79)	2022	VN	87	47% STR
Wessling et al. (97)	2023	VN	55	40% STR
Sauer et al. (78)	2023	VN	133	39% STR
Wilpert et al. (76,80)	2023	VN	65	46% STR
Allen et al.	2023	CNN	76	Improved quality metrics
Chung et al. (83)	2022	CNN	96	Diagnostic-quality images
Müller-Franzes et al. (98)	2023	GAN	5086	Mixed
Wang et al. (99)	2021	GAN	97	Realistic “first phases of dynamic” images
Kim et al. (84)	2021	GAN	446	Realistic contrast-enhanced T1-weighted images

Abbreviations: AUC, area under the curve; CNN, convolutional neural network; VN, variational network; GAN, generative adversarial network; STR, scan time reduction