

RESEARCH ARTICLE

Machine learning with textural analysis of longitudinal multiparametric MRI and molecular subtypes accurately predicts pathologic complete response in patients with invasive breast cancer

Aaquib Syed¹, Richard Adam¹, Thomas Ren¹, Jinyu Lu², Takouhie Maldjian¹, Tim Q. Duong^{1*}

1 Department of Radiology, Montefiore Health System and Albert Einstein College of Medicine, Bronx, New York, United States of America, **2** Oncology Division, Department of Medicine, Montefiore Health System and Albert Einstein College of Medicine, Bronx, New York, United States of America

* Tim.duong@einsteinmed.org



OPEN ACCESS

Citation: Syed A, Adam R, Ren T, Lu J, Maldjian T, Duong TQ (2023) Machine learning with textural analysis of longitudinal multiparametric MRI and molecular subtypes accurately predicts pathologic complete response in patients with invasive breast cancer. PLoS ONE 18(1): e0280320. <https://doi.org/10.1371/journal.pone.0280320>

Editor: Marco Giannelli, Pisa University Hospital, ITALY

Received: April 12, 2022

Accepted: December 27, 2022

Published: January 17, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0280320>

Copyright: © 2023 Syed et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data was obtained as a public dataset in the BMMR2 Challenge. The full

Abstract

Purpose

To predict pathological complete response (pCR) after neoadjuvant chemotherapy using extreme gradient boosting (XGBoost) with MRI and non-imaging data at multiple treatment timepoints.

Material and methods

This retrospective study included breast cancer patients (n = 117) who underwent neoadjuvant chemotherapy. Data types used included tumor ADC values, diffusion-weighted and dynamic-contrast-enhanced MRI at three treatment timepoints, and patient demographics and tumor data. GLCM textural analysis was performed on MRI data. An extreme gradient boosting machine learning algorithm was used to predict pCR. Prediction performance was evaluated using the area under the curve (AUC) of the receiver operating curve along with precision and recall.

Results

Prediction using texture features of DWI and DCE images at multiple treatment time points (AUC = 0.871; 95% CI: (0.768, 0.974; p<0.001) and (AUC = 0.903 95% CI: 0.854, 0.952; p<0.001) respectively), outperformed that using mean tumor ADC (AUC = 0.850 (95% CI: 0.764, 0.936; p<0.001)). The AUC using all MRI data was 0.933 (95% CI: 0.836, 1.03; p<0.001). The AUC using non-MRI data was 0.919 (95% CI: 0.848, 0.99; p<0.001). The highest AUC of 0.951 (95% CI: 0.909, 0.993; p<0.001) was achieved with all MRI and all non-MRI data at all time points as inputs.

dataset is available at (<https://doi.org/10.7937/TCIA.KK02-6D95>).

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Conclusion

Using XGBoost on extracted GLCM features and non-imaging data accurately predicts pCR. This early prediction of response can minimize exposure to toxic chemotherapy, allowing regimen modification mid-treatment and ultimately achieving better outcomes.

Introduction

Neoadjuvant chemotherapy (NAC) [1] in the setting of locally advanced breast cancer can reduce tumor size, making breast conservation surgery feasible and obviating the need for mastectomy. Pathological complete response (pCR) is a desirable endpoint of NAC entailing no residual invasive tumor is present at surgery post NAC [2, 3]. Patients with pCR are more likely to be candidates for breast-conserving surgery and to have longer progression-free and overall survival [2, 3]. Therefore, pCR can be used as a surrogate for favorable outcome. The ability to predict pCR prior to treatment would help in determining which patients will benefit from NAC and which will not. Furthermore, predicting pCR can allow for changes in treatment regimens, maximizing the chances of pCR. Thus, the accurate prediction of pCR could help in identifying patients who are likely to respond to specific NAC drugs while enabling oncologists to alter treatments mid-course if needed in order to maximize successful outcomes while minimizing the adverse effects of unnecessary chemotherapy.

Currently, the efficacy of a chemotherapy regimen is tested invasively through core needle biopsy. Clinical biomarkers, such as Ki67, give a limited assessment of the entire tumor as they are obtained by core needle biopsy and therefore may not be representative of the entire tumor. Noninvasive imaging can overcome this problem of tumor heterogeneity, as the entire tumor is depicted on images [4]. In fact, the capacity of pre and post NAC MRI to depict response to treatment and predict pCR has already been demonstrated [5–7].

Machine learning (ML) has been used to predict eventual pCR. Radiomic features (such as volume, sphericity, DCE MRI signal of wash in and wash out) [8–12] and deep learning analysis of whole MR images [13], DCE dynamics [14] that include demographics and molecular receptor subtypes [15, 16] have been used to predict pCR. ML analysis has also been applied to include diffusion MRI [17–20]. However, the role of diffusion MRI to predict pCR is understudied and the results remain controversial [21].

The goal of this study was to apply the extreme gradient boosting (XGBoost) algorithm to predict pCR using multiparametric MRI data along with non-imaging data at multiple treatment timepoints as inputs. Extreme gradient boosting (XGBoost) was chosen due to the relatively small sample size of the dataset, as XGBoost has been shown to be effective even with limited data. Furthermore, XGBoost has been proven to be effective at classifying with tabular data. This approach has the potential to non-invasively identify patients who are likely to respond to neoadjuvant chemotherapy at diagnosis or early treatment. This approach may prove useful for treatment planning, treatment execution, and mid-treatment adjustment to achieve better outcomes.

Materials and methods

Data sources

In this retrospective study, data from the Breast Multiparametric MRI for prediction of NAC Response-2 (BMMR2) competition training dataset, which was curated from the ACRIN-6698

sub trial of the I-SPY 2 TRIAL (NCT Number: NCT01564368), was used to create machine learning models to predict pCR. Patients from the ACRIN-6698 multicenter trial were previously reported in a paper by Partridge et. Al titled “Diffusion-weighted MRI Findings Predict Pathologic Response in Neoadjuvant Treatment of Breast Cancer: The ACRIN 6698 Multicenter Trial”, published in *Radiology*. Data collected in that trial was published as an open dataset. Our paper utilizes that data (n = 117 patients, of which 36 had pCR). While the previous paper attempted to predict pCR solely using DWI MRI with logistic regression as the model technique, we use extreme gradient boosting with multiple types of MRI data as well as patient demographic data to predict pCR. As the data containing the training and testing sets came from a public dataset, no IRB was required.

All 117 female patients in the dataset were diagnosed with invasive breast cancer and underwent 12 weeks of paclitaxel followed by 4 weeks of anthracycline treatment. Each patient sample contained collections of MRI images taken at 3 distinct timepoints, namely, tp0: pre-treatment, tp1: 3 weeks post paclitaxel (\pm experimental agent), and tp2: 12 weeks post paclitaxel (\pm experimental agent).

Non-imaging data included patient age, race, lesion type (one of multiple masses, single mass or non-mass enhancement), hormone receptor status (hormone receptor (HR) positive/negative, human epidermal growth factor receptor 2 (HER2) positive/negative), and Scarff-Bloom-Richardson (SBR) grade. Of the 117 patients, 12 patients had missing data (11 had an unknown race, and 1 had an unknown SBR grade). These values were filled using backward fill, which populates the missing data with the data from the next patient. There were no other missing data (Fig 1).

MRI data, including DWI and DCE MRI, were performed bilaterally in the axial orientation using a 1.5- or 3.0-T field strength magnet and a dedicated breast radiofrequency coil as described in [22]. Acquisition parameters are provided in the data source in our data availability statement. DCE images were aligned with their corresponding rectangular tumor masks with position metadata, which highlighted the tumor and the surrounding area. Textural features were extracted using the 3rd dynamic DCE data for all 3 time points in the rectangular region of interest mask. DWI images were aligned with manually defined binary segmentation data provided in the dataset. Features were calculated in the smallest rectangle that encompassed the segmented tumor using $b = 800 \text{ s/mm}^2$ DWI imaging data. ADC tumor segmentations were carefully aligned with their corresponding ADC map with position metadata. Tumor ADC values, and changes in these values, were extracted for ML analysis. All images were interpolated to 0.7825 mm x 0.7825 mm voxel spacing with a 2.5 mm slice thickness, which was the median spacing and thickness for all patients. When applying machine learning techniques on multicenter data, data harmonization [23] may be necessary. However, we did not perform additional data harmonization as all MRI acquisitions were taken with ISPY2 acquisition requirements as described in [22]. Furthermore, we ensured that data from different field strengths and different molecular subtypes, etc., were not over-represented in either training and testing data sets by using 5-fold cross validation so that all data has the chance to be in both the test and training sets.

Ground truth

pCR determination was done using histopathologic analysis at study sites by institutional pathologists (blinded to functional tumor volume (FTV) and ADC MRI measures) according to the I-SPY 2 trial protocol using the residual cancer burden system. Following U.S. FDA rationale and guidelines, pCR was the reference standard for determining response to neoadjuvant chemotherapy in our study, defined and reported as no residual invasive disease in either

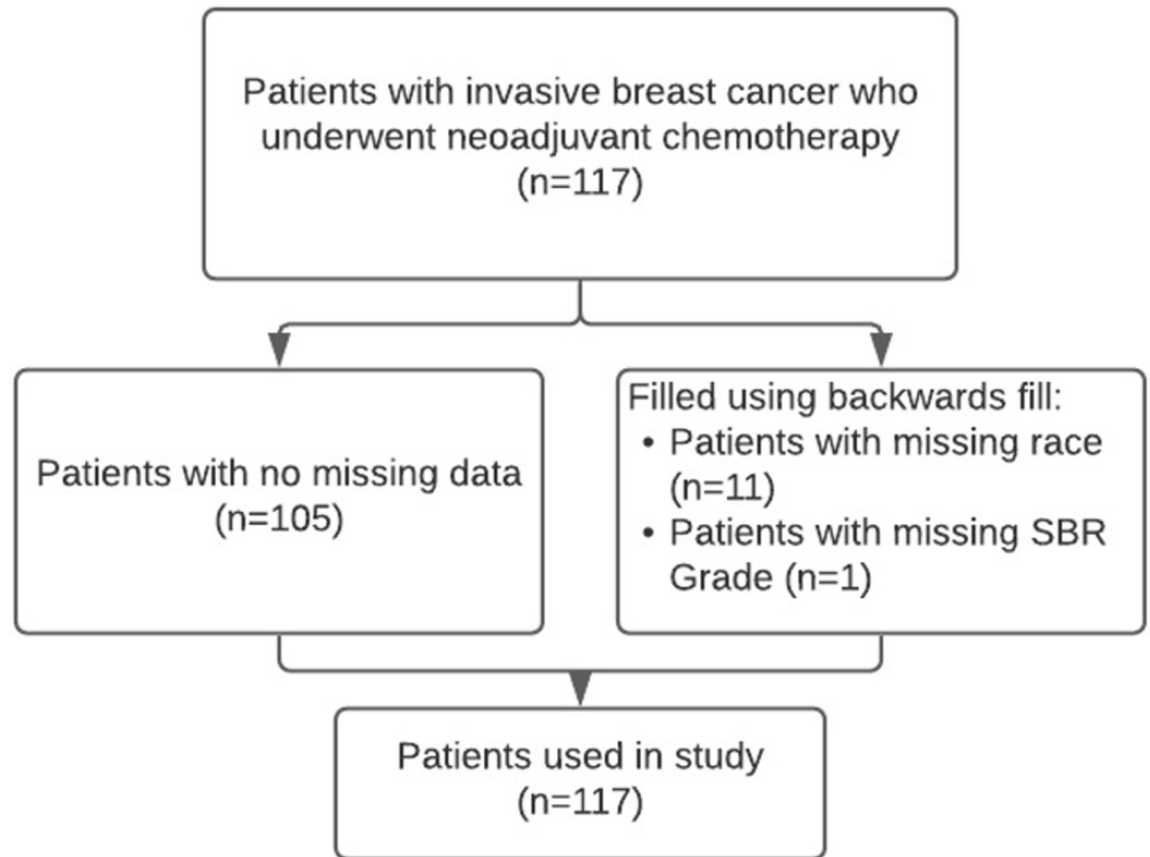


Fig 1. Study population characteristics.

<https://doi.org/10.1371/journal.pone.0280320.g001>

breast or axillary lymph nodes after neoadjuvant therapy. Patients were categorized as having pCR or non-pCR based on postsurgical histopathologic examination findings.

Features used

Texture analysis was performed using a small bounding box enclosing the tumor as determined by functional tumor volume provided by ISPY-2 data. Calculations were performed using the Scikit-Image library in Python. Images were cast to 8-bits, and the number of bins was set to the maximum value of 256 to maximize the number of grey levels counted. GLCM features (Energy, Homogeneity, Contrast, Dissimilarity, ASM, and Correlation) were calculated at distances of 1, 3, and 5 pixels at angles of 0 , $\frac{\pi}{2}$, $\frac{\pi}{4}$, and $\frac{3\pi}{4}$ radians, representing all cardinal and ordinal directions. These features were calculated for all 3 treatment time points for both DCE and DWI MRI images, except for the 5-pixel direction for DWI images, as sometimes the segmentation was too small. [Table 1](#) summarizes the significance of each feature and the formula used to calculate it. This resulted in a total of 360 GLCM features. Combined with 6 non-imaging patient data features, and 6 features extracted from ADC parametric map, the total number of features used was 372. Feature selection to reduce the number of features was unnecessary, as XGBoost inherently selects for the most important features when splitting leaves of decision trees, automatically ignoring irrelevant features.

Table 1. GLCM textural features used.

GLCM Feature Name	Formula	Purpose
Contrast	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - j)^2 P(i, j)$	A measure of the intensity difference between a pixel and its neighbor, 0 for a constant image.
Energy	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j)^2$	A measure of the sum of squared elements in the GLCM, 1 for a constant image.
Correlation	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{(i - \mu_i)(j - \mu_j)P(i, j)}{\sigma_i \sigma_j}$	A measure of how correlated a pixel is to its neighbors over the whole image.
Homogeneity	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i, j)}{1 + i - j }$	A measure of the closeness of the distribution of elements in the GLCM compared to the diagonal.
Dissimilarity	$\sum_{x=1}^K \sum_{y=1}^K x - y P_{xy}$	A measure of distance between pairs of objects (pixels) in the region of interest.
Angular Second Moment (ASM)	$\sum_{i=0}^N \sum_{j=0}^N P(i, j)^2$	A measure of the uniformity of distribution of grey level in the image

<https://doi.org/10.1371/journal.pone.0280320.t001>

Extreme gradient boosting models

A total of 13 XGBoost models were created using different combinations of data and time points. These XGBoost models were created in Python using the Scikit-Learn API. Minority oversampling was used to balance the frequency of each class in the data set by randomly oversampling the minority (pCR) class. The oversampled balanced dataset consisted of 162 patients with 50% pCR and 50% non-PCR outcomes.

Bayesian optimization along with a sequential domain reduction transformer was used to find optimal values of these hyperparameters. The mean AUC after 5-fold cross-verification was selected as the variable to maximize. 15 rounds of random exploration and 80 rounds of optimization were used. If the optimal value was an extreme of the bound, the bounds were adjusted, and optimization was run again. Fig 2 shows this process for a sample hyperparameter.

Statistical analysis

Distributions of patient characteristics, such as the distribution of lesion types and race, were compared using the χ^2 test for homogeneity. Patient age and maximum tumor diameter

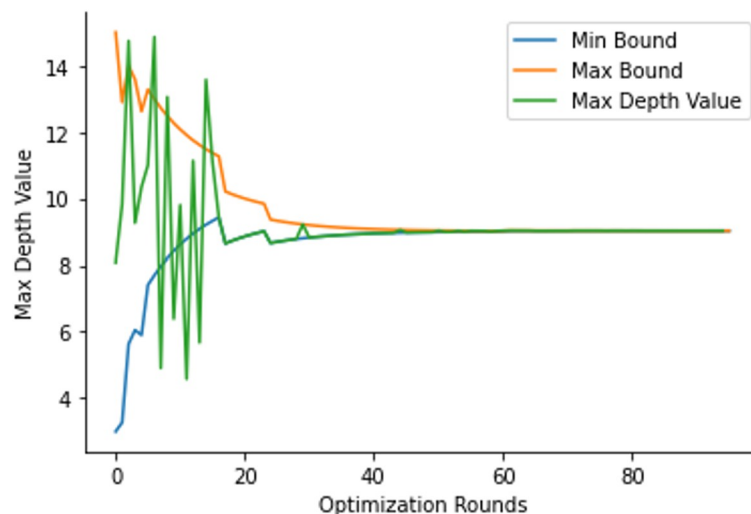


Fig 2. Sequential domain reduction transformer on the “max depth” hyperparameter.

<https://doi.org/10.1371/journal.pone.0280320.g002>

distributions between classes were compared using a t-test. Investigation of the receptor status characteristics were done using 2 sample z-test of proportions. F-scores were used to calculate the importance of each individual feature in predicting pCR.

K-fold cross-validation is considered the gold standard for determining model performance after training. 5 folds of verification were used to designate 20% of the dataset as testing and 80% as training for each fold. All performance metrics were calculated as the mean value of each testing metric of the 5 folds after 1000 rounds of training. Standard error of the mean was calculated as the variability of this mean AUC.

Models were analyzed primarily through the area under the receiver operator curve (AUC). AUC has been shown to be the optimal method for comparing AI models. Precision and recall were secondary metrics for model performance. P-values < 0.05 were considered significant and were calculated as the probability that the null hypothesis is true.

Results

Patient characteristics and the sample sizes (n = 117) are described in [Table 2](#). There were 36 patients with pCR and 81 without pCR. There was no significant difference in the mean age for patients with pCR (49.08 +/- 10.31 years) and patients without pCR (49.0 +/- 11.73 years) (p = 0.971). There was also no significant difference in the longest diameter of the tumor between classes (p = 0.252). The distributions of patient races (p = 0.205), lesion types (p = 0.409), and SBR grades (p = 0.488) showed no significant difference. There was a significant difference in distributions of receptor statuses (p<0.001), with the patient without pCR

Table 2. Patient demographics and sample sizes.

	Patients with pCR (n = 36)	Patients without pCR (n = 81)	P-Value
Age	49.08 ± 10.31 years	49.0 ± 11.73 years	0.971
Race	White (n = 26)	White (n = 61)	0.205
	Asian (n = 3)	Asian (n = 7)	
	Black (n = 1)	Black (n = 8)	
	Unknown (n = 6)	Unknown (n = 5)	
Lesion Type	Multiple masses (n = 16)	Multiple masses (n = 49)	0.409
	Multiple NME (n = 2)	Multiple NME (n = 3)	
	Single mass (n = 16)	Single mass (n = 27)	
	Single NME (n = 2)	Single NME (n = 2)	
Receptor Status	HR + / HER2 + (n = 8)	HR + / HER2 + (n = 12)	<0.001
	HR + / HER2 - (n = 8)	HR + / HER2 - (n = 43)	
	HR- / HER2 - (n = 12)	HR- / HER2 - (n = 24)	
	HR- / HER2 + (n = 8)	HR- / HER2 + (n = 2)	
SBR Grade	III [High] (n = 22)	III [High] (n = 55)	0.488
	II [Intermediate] (n = 13)	II [Intermediate] (n = 23)	
	I [Low] (n = 0)	I [Low] (n = 3)	
	Unknown (n = 0)	Unknown (n = 1)	
Longest Diameter	3.67 +/- 2.21 cm	4.18 +/- 2.19 cm	0.252

Distributions of patient characteristics, such as the distribution of lesion types and race, were compared using the χ^2 test for homogeneity. Patient age and maximum tumor diameter distributions between classes were compared using a t-test. Investigation of the receptor status characteristics were done using 2 sample z-test of proportions. NME: non-mass enhancing; HR: hormone receptor; HER2: human epidermal growth factor receptor 2; pCR: complete pathological response; SBR: Scarff-Bloom-Richardson.

<https://doi.org/10.1371/journal.pone.0280320.t002>

class having a greater proportion of patients that have HR+/HER2- status type ($p < 0.001$) and a lesser proportion of patients with the HR-/HER2+ status type ($p < 0.001$).

Fig 3A shows post-contrast DCE MRIs for a pCR patient and **Fig 3B** shows these MRIs for a non pCR patient at the pre-treatment, early treatment, and mid-treatment time points. The tumors were hyperintense relative to background tissue. Tumor of the non-pCR patient shrunk moderately whereas the tumor of the pCR patient shrunk markedly with time.

Fig 4A shows post-contrast DWI MRIs for a pCR patient and **Fig 4B** shows these MRIs for a non pCR patient. DWI signals of the tumors were hyperintense relative to background tissue. Tumor of the non-pCR patient shrunk moderately whereas the tumor of the pCR patient shrunk markedly with time.

Fig 5 describes the 10 most important texture features out of 372 features. All top 10 features were GLCM textural features derived from MRI data. In comparison, patient data scored far lower in terms of F-score. Patient age, race, and lesion type had F-scores of 1, 0, and 3 respectively. Receptor status, SBR grade, and longest diameter of the tumor had F-scores of 6, 3, and 3 respectively.

For pCR patients, regional tumor ADCs were 0.59 ± 0.05 , 0.81 ± 0.04 , and $1.11 \pm 0.06 \times 10^{-3} \text{ mm}^2/\text{s}$ at pre-treatment, early treatment, and mid-treatment, respectively (**Fig 6**). For non-pCR patients, the corresponding tumor ADC values were 0.59 ± 0.03 , 0.74 ± 0.03 , and $0.91 \pm 0.05 \times 10^{-3} \text{ mm}^2/\text{s}$. ADC values for both PCR and non-PCR patients increased with treatment, with pCR patients showing a larger increase progressively.

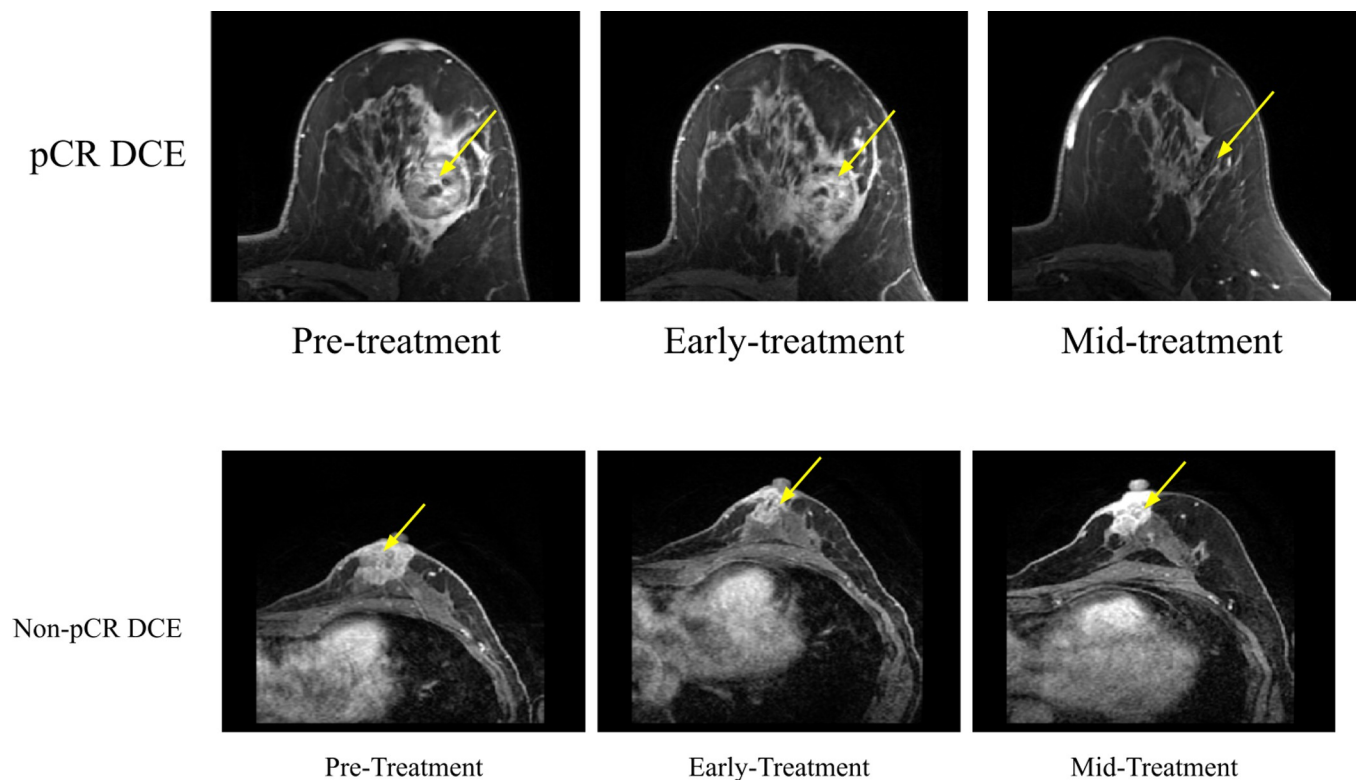


Fig 3. Post-contrast MRIs from DCE for (A) a pCR patient and (B) a non-PCR patient at the pre-treatment, early treatment, and mid-treatment time points. For (A), the 54 year old Caucasian patient had invasive ductal carcinoma with HR- and HER2-. For (B), the 51 year old Caucasian patient had invasive ductal carcinoma with HR+ and HER2-. The yellow arrow indicates the approximate location of the tumor. MRI images were taken approx. 2.5 minutes after contrast injection with a phase duration of $80\text{sec} < \text{phase duration} < 100\text{sec}$. pCR: complete pathological response, DCE: dynamic contrast enhanced.

<https://doi.org/10.1371/journal.pone.0280320.g003>

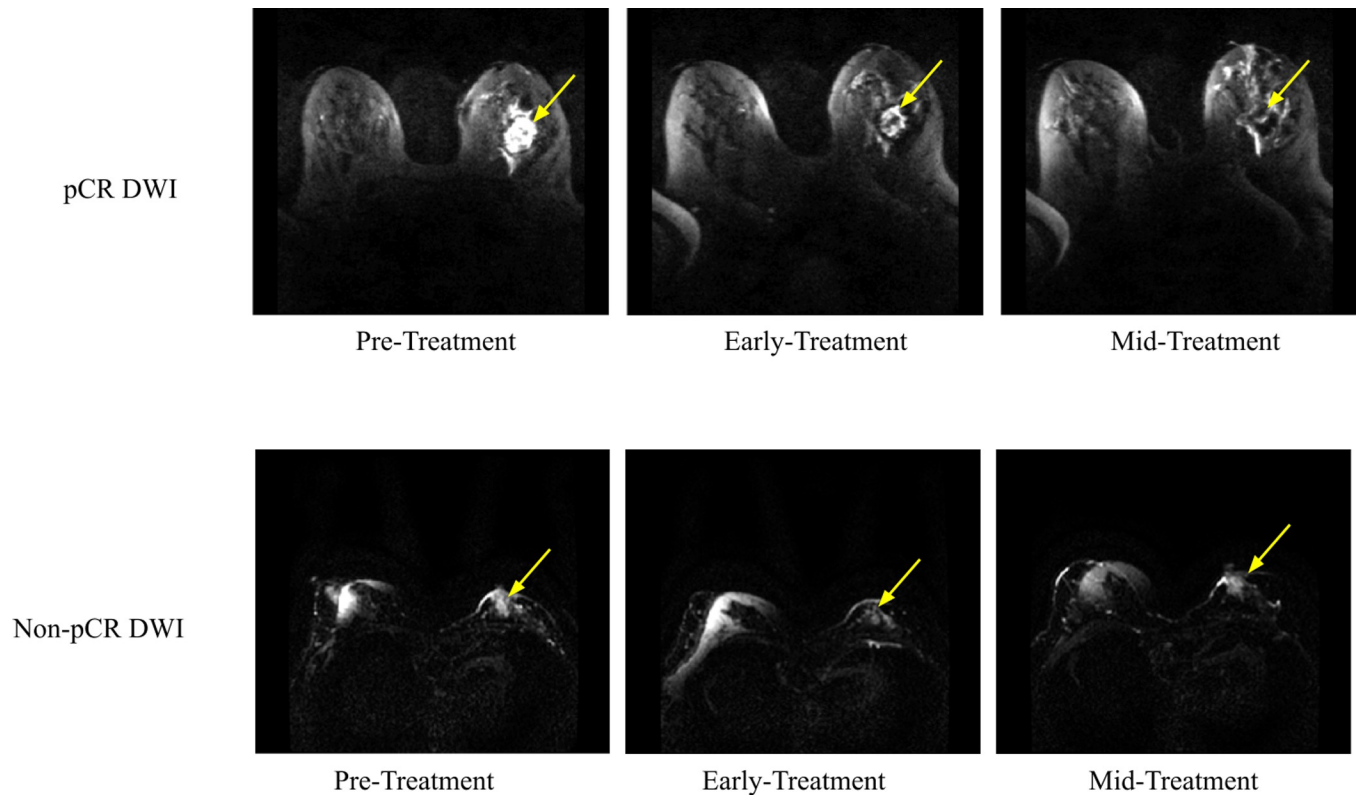


Fig 4. DWIs for (A) a pCR patient and (B) a non-PCR patient at the pre-treatment, early treatment, and mid-treatment time points. For (A), the 54 year old Caucasian patient had invasive ductal carcinoma with HR- and HER2-. For (B), the 51 year old Caucasian patient had invasive ductal carcinoma with HR+ and negative HER2-. The yellow arrow indicates the approximate location of the tumor. DWI acquisitions were done using a b-value of 800 s/mm². DWI: diffusion-weighted imaging).

<https://doi.org/10.1371/journal.pone.0280320.g004>

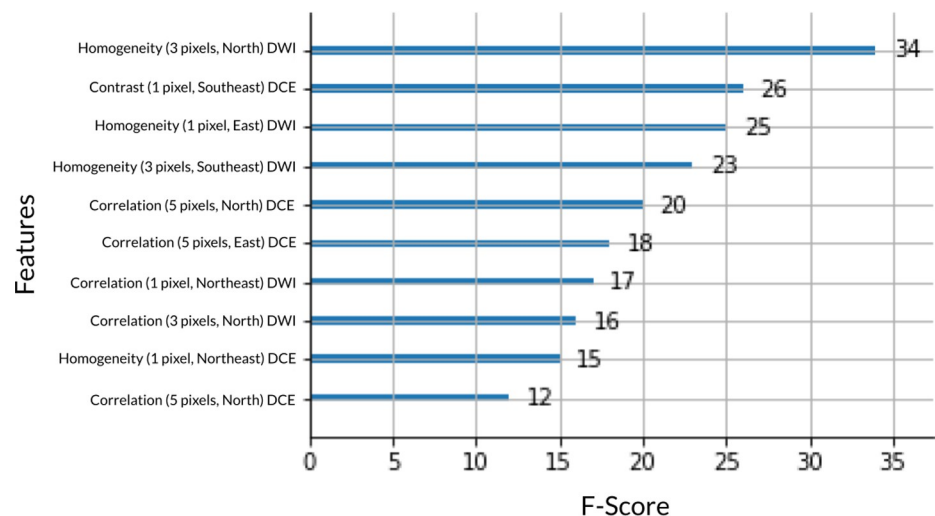


Fig 5. Top 10 most important texture features out of 372 features by F-score. GLCM features are shown with pixel distance value (1, 3 or 5 pixels), direction (North/South, East/West, North East/South West, or North West/South East), and MRI imaging type (DCE: dynamic contrast enhanced or DWI: diffusion-weighted imaging). GLCM: gray-level co-occurrence matrix.

<https://doi.org/10.1371/journal.pone.0280320.g005>

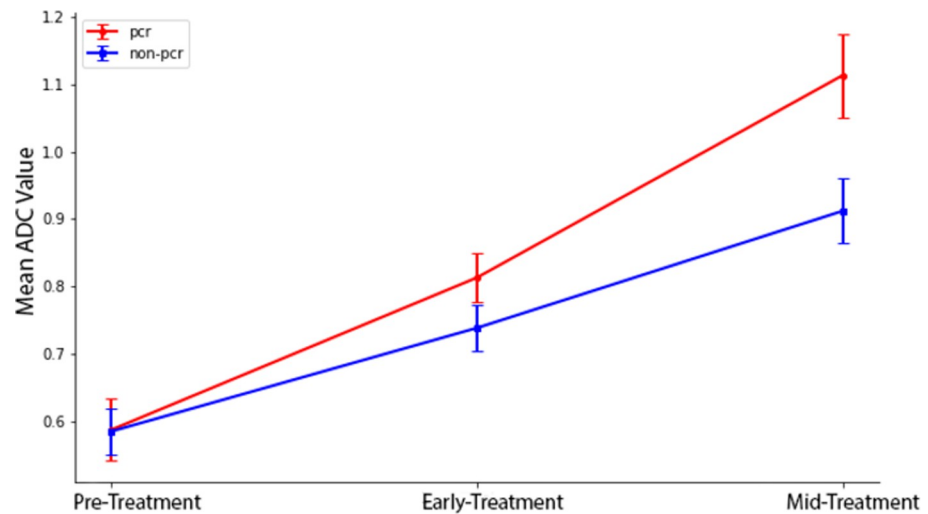


Fig 6. Mean ADC values for all non-pCR and pCR patients at the pre-treatment, early treatment, and mid-treatment time points. Error bars are standard deviations. pCR: complete pathological response.

<https://doi.org/10.1371/journal.pone.0280320.g006>

Prediction of pCR

Table 3 shows the AUC, precision, and recall performance of XGBoost for different data inputs after five-fold cross-validation. These AUC values were obtained using all treatment time points as input. With the mean tumor ADC, the AUC was 0.850 (95% CI: 0.764, 0.936; $p < 0.001$). With texture features of DWI, the AUC was 0.871 (95% CI: 0.768, 0.974; $p < 0.001$). With texture features of DCE images, the AUC was 0.903 (95% CI: 0.854, 0.952; $p < 0.001$). Combining texture features of both DCE and DWI, the AUC was 0.916 (95% CI: 0.851, 0.981; $p < 0.001$). All MRI data yielded an AUC of 0.933 (95% CI: 0.836, 1.03; $p < 0.001$). By comparison, using non-imaging data yielded an AUC of 0.910 (95% CI: 0.848, 0.99; $p < 0.001$). Prediction using all available imaging and non-imaging data yielded an AUC of 0.951 (95% CI: 0.909, 0.993; $p < 0.001$).

We also evaluated the effects of using different combinations of treatment time points on prediction performance. For prediction models using all available MRI and non-MRI data, AUC using tp0, tp1, tp2, and tp0+tp1 were respectively, 0.918 (95% CI: 0.856, 0.98; $p < 0.001$),

Table 3. Model performances for models trained on ADC, DCE(GLCM), DWI(GLCM), their combinations, all non-imaging data, and the combination of all MRI and all non-MRI data with 5-fold cross-validation.

Features	AUC	Precision	Recall
ADC	0.850 (0.764, 0.936; $p < 0.001$)	0.752 (0.666, 0.838; $p < 0.002$)	0.827 (0.753, 0.901; $p < 0.001$)
DWI GLCM	0.871 (0.768, 0.974; $p < 0.001$)	0.779 (0.713, 0.845; $p < 0.001$)	0.926 (0.861, 0.991; $p < 0.001$)
DCE GLCM	0.903 (0.854, 0.952; $p < 0.001$)	0.856 (0.808, 0.904; $p < 0.001$)	0.939 (0.891, 0.987; $p < 0.001$)
DCE+DWI GLCM	0.916 (0.851, 0.981; $p < 0.001$)	0.779 (0.678, 0.880; $p < 0.003$)	0.915 (0.861, 0.969; $p < 0.001$)
All MRI Data	0.933 (0.836, 1.030; $p < 0.001$)	0.824 (0.780, 0.868; $p < 0.001$)	0.889 (0.848, 0.930; $p < 0.001$)
All Non-MRI Data	0.919 (0.848, 0.990; $p < 0.001$)	0.762 (0.665, 0.859; $p < 0.003$)	0.914 (0.888, 0.940; $p < 0.001$)
All MRI + Non-MRI Data	0.951 (0.909, 0.993; $p < 0.001$)	0.815 (0.690, 0.940; $p < 0.004$)	0.926 (0.844, 1.008; $p < 0.001$)

Values in parentheses are 95% confidence intervals. P-values are probability of AUC being 0.5. DWI: diffusion-weighted imaging; ADC: apparent diffusion coefficient; AUC: area under the curve; GLCM: gray-level co-occurrence matrix.

<https://doi.org/10.1371/journal.pone.0280320.t003>

0.844 (95% CI: 0.705, 0.983; $p = 0.004$), 0.920 (95% CI: 0.858, 0.982; $p < 0.001$), and 0.938 (95% CI: 0.89, 0.986; $p < 0.001$). The AUC of using all time points were 0.951 (95% CI: 0.91, 0.992; $p < 0.001$) (as shown in [Table 4](#)).

Discussion

This study aimed to predict pCR non-invasively using XGBoost models with multiparametric MRI data at multiple treatment timepoints along with non-imaging data. This paper is the first to apply an extreme gradient boosting algorithm to the ISPY-2 data set to predict pathological complete response, and the second ever to use such an algorithm for predicting NAC response. This paper is also the first ever to use GLCM features to do so.

In the original I-SPY2 study, Partridge et al. [22] used logistic regression to evaluate if change in tumor ADC values (not texture analysis) is predictive of pCR in breast cancer patients. They found that changes in tumor ADC values was moderately predictive of pCR at mid-treatment (AUC = 0.60; 95% CI: 0.52, 0.68). In a test subset, a model combining tumor subtype and mid-treatment changes in ADC significantly improved predictive performance (AUC = 0.72; 95% CI: 0.61, 0.83).

There have only been a few studies that have applied deep learning on MR images themselves as inputs to predict pCR. Braman et al. [7] applied deep learning to predict NAC in HER2 patients from pre-treatment 2D DCE MRI in a retrospective study. They explored both pre-contrast and late post-contrast phases of DCE MRI and found an AUC of 0.74 (± 0.03). Qu et al. [24] built a CNN model using pre and post-NAC. Tumor regions were manually segmented by two expert radiologists on enhanced T1-weighted images. They found an AUC of 0.553 (95% CI: 0.416, 0.683) for pre-NAC data.

There have also only been a few studies that have applied non-deep learning on MR images themselves as inputs to predict pCR. Suo et al. [17] evaluated mono-exponential (ADC), bi-exponential and stretched-exponential from diffusion MRI data to predict pCR. They also included tumor size and relative enhancement ratio from DCE MRI data at 3 time points: before treatment, at mid-treatment, and after treatment with NAC. They found that flow-insensitive ADC change at mid-treatment was the most predictive feature, with an AUC of 0.831 (95% CI: 0.747, 0.915; $P < 0.001$). Combining this with receptor statuses, the AUC increased to 0.905 (95% CI: 0.843, 0.966; $P < 0.001$). Bian et al. [25] analyzed radiomic signatures based on T2W imaging, diffusion-weighted imaging, dynamic contrast-enhanced imaging and their combination to predict pCR. Logistic regression was then used to assess the association between features and clinical risk factors. The combined radiomic signature and nomogram model achieved an AUC of 0.91 (95% CI: 0.86, 1.00). Chen et al. [26] similarly analyzed the ability of radiomic signatures extracted from MRI data to predict pCR. They achieved an AUC of 0.848 by combining DCE-MRI and ADC maps. Eun et al. [19] performed non-GLCM texture analysis of pre and mid-treatment T2-weighted MRI, DCE, DWI, and ADC

Table 4. Model performances for models trained on all features restricted to tp0, tp1, tp2, tp0 + tp1 and all timepoints using 5-fold cross-validation.

Timepoints	AUC	Precision	Recall
Tp0	0.918 (0.856, 0.980; $p < 0.001$)	0.778 (0.716, 0.840; $p < 0.001$)	0.913 (0.885, 0.941; $p < 0.001$)
Tp1	0.844 (0.705, 0.983; $p = 0.004$)	0.793 (0.653, 0.933; $p = 0.007$)	0.841 (0.709, 0.973; $p = 0.003$)
Tp2	0.920 (0.858, 0.982; $p < 0.001$)	0.832 (0.77, 0.894; $p < 0.001$)	0.951 (0.929, 0.973; $p < 0.001$)
Tp0 + Tp1	0.938 (0.89, 0.986; $p < 0.001$)	0.805 (0.757, 0.853; $p < 0.001$)	0.963 (0.918, 1.008; $p < 0.001$)
Tp0 + Tp1 + Tp2	0.951 (0.910, 0.992; $p < 0.001$)	0.815 (0.690, 0.940; $p = 0.004$)	0.926 (0.844, 1.008; $p < 0.001$)

Values in parentheses are 95% confidence intervals. P-values are probability of AUC being 0.5. AUC: area under the curve; Tp: time point.

<https://doi.org/10.1371/journal.pone.0280320.t004>

mapping. The random forest classifier to predict pCR showed the highest diagnostic performance with mid-treatment DCE MRI (AUC = 0.82; 95% CI: 0.74, 0.88). Huang et al. [27] used a multilayer perceptron trained on radiomics features generated from MRI images (ADC maps, DCE, and fat-suppressed T2-weighted imaging), and reported an AUC of 0.900 AUC (95% CI: 0.849, 0.935) on the validation dataset. Tahmassebi et al. [20] was the only other paper to employ XGBoost to predict pCR. Using quantitative pharmacokinetic DCE features and ADC values in different classifier algorithms (support vector machine, linear discriminant analysis, logistic regression, random forests, stochastic gradient descent, decision tree, adaptive boosting and extreme gradient boosting), they achieved a final AUC of 0.86 with an extreme gradient boosting.

Limitations

This study was performed on a relatively small multi-center dataset. As more data releases, our findings need to be replicated on a larger dataset to improve the generalizability of the model as well as account for the racial bias towards white women in this dataset. We have only evaluated XGBoost, other machine learning methods such as LightGBM should also be explored. Finally, we have only extracted GLCM features from MRI imaging data. Using additional features such as radiomic textural features may improve the model and give it more perspectives on the images. We employed images as provided and performed only visual image quality checks. We did not perform distortion correction, eddy current correction, check for multisite scanner consistency, among others, as raw data were not available. It has been shown that accurate estimates and reproducibility of diffusion indices in clinical studies require careful data quality assurance [28]. Moreover, large multicenter studies have shown a non-negligible variability in quantitative diffusion indices measured using different scanner systems [29, 30]. Therefore, MRI scanner system should be adequately characterized in diffusion-MRI of the breast [31]. Deep learning methods could also be applied to predict pCR [24, 32–37] (see review [38]) but were not analyzed.

Conclusions

XGBoost models using multiparametric MRI data along with demographic and molecular subtype data, accurately predict pCR to NAC. With further development and testing on larger multi-institutional sample sizes, this approach has the potential to non-invasively identify patients who are likely to respond to neoadjuvant chemotherapy at diagnosis or early treatment. This approach may prove useful for treatment planning, treatment execution, and mid-treatment adjustment to achieve better outcomes.

Author Contributions

Conceptualization: Aaquib Syed, Takouhie Maldjian, Tim Q. Duong.

Data curation: Aaquib Syed.

Formal analysis: Aaquib Syed, Tim Q. Duong.

Funding acquisition: Aaquib Syed.

Investigation: Aaquib Syed, Thomas Ren, Tim Q. Duong.

Methodology: Aaquib Syed, Takouhie Maldjian, Tim Q. Duong.

Project administration: Tim Q. Duong.

Resources: Tim Q. Duong.

Software: Aaquib Syed, Thomas Ren.

Supervision: Aaquib Syed, Richard Adam, Thomas Ren, Jinyu Lu, Takouhie Maldjian, Tim Q. Duong.

Validation: Aaquib Syed, Thomas Ren, Takouhie Maldjian, Tim Q. Duong.

Visualization: Aaquib Syed.

Writing – original draft: Aaquib Syed, Richard Adam, Jinyu Lu, Takouhie Maldjian, Tim Q. Duong.

Writing – review & editing: Aaquib Syed, Richard Adam, Jinyu Lu, Takouhie Maldjian, Tim Q. Duong.

References

1. Curigliano G, Burstein HJ, Winer EP, Gnant M, Dubsy P, Loibl S, et al. Correction to: De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Annals of Oncology*. 2018; 29: 2153. <https://doi.org/10.1093/annonc/mdx806> PMID: 29733336
2. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *The Lancet*. 2014; 384: 164–172. [https://doi.org/10.1016/S0140-6736\(13\)62422-8](https://doi.org/10.1016/S0140-6736(13)62422-8) PMID: 24529560
3. Cortazar P, Geyer CE. Pathological Complete Response in Neoadjuvant Treatment of Breast Cancer. *Ann Surg Oncol*. 2015; 22: 1441–1446. <https://doi.org/10.1245/s10434-015-4404-8> PMID: 25727556
4. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*. 2012; 48: 441–446. <https://doi.org/10.1016/j.ejca.2011.11.036> PMID: 22257792
5. Li X, Arlinghaus LR, Ayers GD, Chakravarthy AB, Abramson RG, Abramson VG, et al. DCE-MRI analysis methods for predicting the response of breast cancer to neoadjuvant chemotherapy: Pilot study findings: DCE-MRI to Predict Breast Cancer Treatment Response. *Magn Reson Med*. 2014; 71: 1592–1602. <https://doi.org/10.1002/mrm.24782> PMID: 23661583
6. Li X, Abramson RG, Arlinghaus LR, Kang H, Chakravarthy AB, Abramson VG, et al. Multiparametric Magnetic Resonance Imaging for Predicting Pathological Response After the First Cycle of Neoadjuvant Chemotherapy in Breast Cancer: Investigative Radiology. 2015; 50: 195–204. <https://doi.org/10.1097/RLI.000000000000100> PMID: 25360603
7. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res*. 2017; 19: 57. <https://doi.org/10.1186/s13058-017-0846-1> PMID: 28521821
8. Mani S, Chen Y, Li X, Arlinghaus L, Chakravarthy AB, Abramson V, et al. Machine learning for predicting the response of breast cancer to neoadjuvant chemotherapy. *J Am Med Inform Assoc*. 2013; 20: 688–695. <https://doi.org/10.1136/amiajnl-2012-001332> PMID: 23616206
9. Tahmassebi A, Gandomi AH, Fong S, Meyer-Baese A, Foo SY. Multi-stage optimization of a deep model: A case study on ground motion modeling. Olier I, editor. *PLoS ONE*. 2018; 13: e0203829. <https://doi.org/10.1371/journal.pone.0203829> PMID: 30231077
10. Cain EH, Saha A, Harowicz MR, Marks JR, Marcom PK, Mazurowski MA. Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set. *Breast Cancer Res Treat*. 2019; 173: 455–463. <https://doi.org/10.1007/s10549-018-4990-9> PMID: 30328048
11. Lo Gullo R, Eskreis-Winkler S, Morris EA, Pinker K. Machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy. *The Breast*. 2020; 49: 115–122. <https://doi.org/10.1016/j.breast.2019.11.009> PMID: 31786416
12. Houssein EH, Emam MM, Ali AA, Suganthan PN. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Systems with Applications*. 2021; 167: 114161. <https://doi.org/10.1016/j.eswa.2020.114161>
13. El Adoui M, Drisis S, Benjelloun M. A PRM approach for early prediction of breast cancer response to chemotherapy based on registered MR images. *Int J CARS*. 2018; 13: 1233–1243. <https://doi.org/10.1007/s11548-018-1790-y> PMID: 29790078

14. Ravichandran K, Braman N, Janowczyk A, Madabhushi A. A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI. In: Mori K, Petrick N, editors. *Medical Imaging 2018: Computer-Aided Diagnosis*. Houston, United States: SPIE; 2018. p. 11. <https://doi.org/10.1117/12.2294056>
15. Schettini F, Pascual T, Conte B, Chic N, Brasó-Maristany F, Galván P, et al. HER2-enriched subtype and pathological complete response in HER2-positive breast cancer: A systematic review and meta-analysis. *Cancer Treatment Reviews*. 2020; 84: 101965. <https://doi.org/10.1016/j.ctrv.2020.101965> PMID: 32000054
16. Kalinowski L, Saunus JM, McCart Reed AE, Lakhani SR. Breast Cancer Heterogeneity in Primary and Metastatic Disease. In: Ahmad A, editor. *Breast Cancer Metastasis and Drug Resistance*. Cham: Springer International Publishing; 2019. pp. 75–104. https://doi.org/10.1007/978-3-030-20301-6_6
17. Suo S, Yin Y, Geng X, Zhang D, Hua J, Cheng F, et al. Diffusion-weighted MRI for predicting pathologic response to neoadjuvant chemotherapy in breast cancer: evaluation with mono-, bi-, and stretched-exponential models. *J Transl Med*. 2021; 19: 236. <https://doi.org/10.1186/s12967-021-02886-3> PMID: 34078388
18. Zhao R, Lu H, Li Y-B, Shao Z-Z, Ma W-J, Liu P-F. Nomogram for Early Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Dynamic Contrast-enhanced and Diffusion-weighted MRI. *Academic Radiology*. 2022; 29: S155–S163. <https://doi.org/10.1016/j.acra.2021.01.023> PMID: 33593702
19. Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim J-A, et al. Texture Analysis with 3.0-T MRI for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer. *Radiology*. 2020; 294: 31–41. <https://doi.org/10.1148/radiol.2019182718> PMID: 31769740
20. Tahmassebi A, Wengert GJ, Helbich TH, Bago-Horvath Z, Alaei S, Bartsch R, et al. Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant Chemotherapy and Survival Outcomes in Breast Cancer Patients: Investigative Radiology. 2019; 54: 110–117. <https://doi.org/10.1097/RLI.0000000000000518> PMID: 30358693
21. van der Hoogt KJJ, Schipper RJ, Winter-Warnars GA, Ter Beek LC, Loo CE, Mann RM, et al. Factors affecting the value of diffusion-weighted imaging for identifying breast cancer patients with pathological complete response on neoadjuvant systemic therapy: a systematic review. *Insights Imaging*. 2021; 12: 187. <https://doi.org/10.1186/s13244-021-01123-1> PMID: 34921645
22. Partridge SC, Zhang Z, Newitt DC, Gibbs JE, Chenevert TL, Rosen MA, et al. Diffusion-weighted MRI Findings Predict Pathologic Response in Neoadjuvant Treatment of Breast Cancer: The ACRIN 6698 Multicenter Trial. *Radiology*. 2018; 289: 618–627. <https://doi.org/10.1148/radiol.2018180273> PMID: 30179110
23. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8: 118–127. <https://doi.org/10.1093/biostatistics/kxj037> PMID: 16632515
24. 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):651–8. Epub 20200116. <https://doi.org/10.1111/1759-7714.13309> PMID: 31944571; PubMed Central PMCID: PMC7049483.
25. Bian T, Wu Z, Lin Q, Wang H, Ge Y, Duan S, et al. Radiomic signatures derived from multiparametric MRI for the pretreatment prediction of response to neoadjuvant chemotherapy in breast cancer. *BJR*. 2020; 93: 20200287. <https://doi.org/10.1259/bjr.20200287> PMID: 32822542
26. Chen X, Chen X, Yang J, Li Y, Fan W, Yang Z. Combining Dynamic Contrast-Enhanced Magnetic Resonance Imaging and Apparent Diffusion Coefficient Maps for a Radiomics Nomogram to Predict Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Patients: Journal of Computer Assisted Tomography. 2020; 44: 275–283. <https://doi.org/10.1097/RCT.0000000000000978> PMID: 32004189
27. Huang Y, Chen W, Zhang X, He S, Shao N, Shi H, et al. Prediction of Tumor Shrinkage Pattern to Neoadjuvant Chemotherapy Using a Multiparametric MRI-Based Machine Learning Model in Patients With Breast Cancer. *Front Bioeng Biotechnol*. 2021; 9: 662749. <https://doi.org/10.3389/fbioe.2021.662749> PMID: 34295877
28. Jones DK. Precision and accuracy in diffusion tensor magnetic resonance imaging. *Top Magn Reson Imaging*. 2010; 21: 87–99. <https://doi.org/10.1097/RMR.0b013e31821e56ac> PMID: 21613874
29. Fedeli L, Belli G, Ciccarone A, Coniglio A, Esposito M, Giannelli M, et al. Dependence of apparent diffusion coefficient measurement on diffusion gradient direction and spatial position—A quality assurance intercomparison study of forty-four scanners for quantitative diffusion-weighted imaging. *Phys Med*. 2018; 55: 135–141. <https://doi.org/10.1016/j.ejmp.2018.09.007> PMID: 30342982

30. Fedeli L, Benelli M, Busoni S, Belli G, Ciccarone A, Coniglio A, et al. On the dependence of quantitative diffusion-weighted imaging on scanner system characteristics and acquisition parameters: A large multicenter and multiparametric phantom study with unsupervised clustering analysis. *Phys Med*. 2021; 85: 98–106. <https://doi.org/10.1016/j.ejmp.2021.04.020> PMID: 33991807
31. Giannelli M, Sghedoni R, Iaconi C, Iori M, Traino AC, Guerrisi M, et al. MR Scanner Systems Should Be Adequately Characterized in Diffusion-MRI of the Breast. Woloschak GE, editor. *PLoS ONE*. 2014; 9: e86280. <https://doi.org/10.1371/journal.pone.0086280> PMID: 24489711
32. Duanmu H, Huang PB, Brahmava S, Lin S, Ren T, Kong RJ, et al. Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Deep Learning with Integrative Imaging, Molecular and Demographic Data. *MICCAI 2020: Medical Image Computing and Computer Assisted Intervention 2020*. p. 242–52.
33. 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; 73:148–51. Epub 20200902. <https://doi.org/10.1016/j.mri.2020.08.021> PMID: 32889091; PubMed Central PMCID: PMC8111786.
34. Huynh BQ, Antropova N, Giger ML. Comparison of breast DCE-MRI contrast time points for predicting response to neoadjuvant chemotherapy using deep convolutional neural network features with transfer learning. *SPIE Medical Imaging*. 2017;10134. <https://doi.org/10.1117/12.2255316>
35. Duanmu H, Ren T, Duong TQ. Deep learning prediction of pathological complete response, residual cancer burden, and progression-free survival in breast cancer patients. *PlosOne*. 2022, in press.
36. Ren T, Lin S, Huang P, Duong TQ. Convolutional Neural Network of Multiparametric MRI Accurately Detects Axillary Lymph Node Metastasis in Breast Cancer Patients With Pre Neoadjuvant Chemotherapy. *Clin Breast Cancer*. 2022; 22(2):170–7. Epub 20210713. <https://doi.org/10.1016/j.clbc.2021.07.002> PMID: 34384696.
37. Hussain L, Huang P, Nguyen T, Lone KJ, Ali A, Khan MS, et al. Machine learning classification of texture features of MRI breast tumor and peri-tumor of combined pre- and early treatment predicts pathologic complete response. *Biomed Eng Online*. 2021; 20(1):63. Epub 20210628. <https://doi.org/10.1186/s12938-021-00899-z> PMID: 34183038.
38. Khan N, Adam R, Huang P, Maldjian T, Duong TQ. Deep Learning Prediction of Pathologic Complete Response in Breast Cancer Using MRI and Other Clinical Data: A Systematic Review. *Tomography*. 2022; 8(6):2784–95. Epub 20221121. <https://doi.org/10.3390/tomography8060232> PMID: 36412691.