Breast Cancer Research and Treatment (2019) 173:455–463
https://doi.org/10.1007/s10549-018-4990-9

EPIDEMIOLOGY

# 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

**Elizabeth Hope Cain[1] · Ashirbani Saha[1] · Michael R. Harowicz[1,2] · Jeffrey R. Marks[3] · P. Kelly Marcom[4] · Maciej A. Mazurowski[1,5]**

Received: 29 March 2018 / Accepted: 1 October 2018 / Published online: 16 October 2018


## Abstract

**Purpose** To determine whether a multivariate machine learning-based model using computer-extracted features of pre-treatment dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) can predict pathologic complete response (pCR) to neoadjuvant therapy (NAT) in breast cancer patients.

**Methods** Institutional review board approval was obtained for this retrospective study of 288 breast cancer patients at our institution who received NAT and had a pre-treatment breast MRI. A comprehensive set of 529 radiomic features was extracted from each patient's pre-treatment MRI. The patients were divided into equal groups to form a training set and an independent test set. Two multivariate machine learning models (logistic regression and a support vector machine) based on imaging features were trained to predict pCR in (a) all patients with NAT, (b) patients with neoadjuvant chemotherapy (NACT), and (c) triple-negative or human epidermal growth factor receptor 2-positive (TN/HER2+) patients who had NAT. The multivariate models were tested using the independent test set, and the area under the receiver operating characteristics (ROC) curve (AUC) was calculated.

**Results** Out of the 288 patients, 64 achieved pCR. The AUC values for predicting pCR in TN/HER+ patients who received NAT were significant (0.707, 95% CI 0.582–0.833, $p < 0.002$).

**Conclusions** The multivariate models based on pre-treatment MRI features were able to predict pCR in TN/HER2+ patients.

**Keywords** Pathologic complete response · Neoadjuvant therapy · Breast cancer · Breast cancer MRI · MRI radiomics · Machine learning · Logistic regression · Support vector machines

✉ Elizabeth Hope Cain
Elizabeth.cain@duke.edu

Ashirbani Saha
as698@duke.edu

Michael R. Harowicz
michael.harowicz@gmail.com

Jeffrey R. Marks
jeffrey.marks@duke.edu

P. Kelly Marcom
kelly.marcom@duke.edu

Maciej A. Mazurowski
maciej.mazurowski@duke.edu

[1] Department of Radiology, Duke University School of Medicine, 2301 Erwin Road, Durham, NC 27705, USA

[2] Department of Radiology, Johns Hopkins University, Baltimore, MD, USA

[3] Department of Surgery, Duke University School of Medicine, 2301 Erwin Road, Durham, NC 27705, USA

[4] Department of Medicine, Duke University School of Medicine, 2301 Erwin Road, Durham, NC 27705, USA

[5] Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA

## Introduction

Neoadjuvant therapy (NAT) allows for early treatment of breast cancer, aiming to reduce tumor burden and allow patients to undergo breast-conserving surgery rather than mastectomy [1–6]. One trial reported up to 23% of patients with primary operable breast cancer can have breast-conserving surgery instead of mastectomy if treated with neoadjuvant therapy [7]. However, response rates to NAT vary among patients and have been shown to depend on subtype [8–16]. In fact, up to 30% of patients may not respond to NAT [17]. These patients incur toxicities and costs of treatment but gain proportionally less benefit [7, 8]. Identifying a method to predict which breast cancer patients will achieve pCR and which will have residual invasive disease (RD) following NAT would allow for improved stratification of patients to more appropriate treatment regimens [18].

Among other means of predicting response to NAT, such as using tumor stage and receptor status [19], features extracted from pre-treatment MRI have shown promise for this purpose [8, 18, 20–28]. Using MRI could provide a fast and non-invasive way to further stratify patients at no additional cost to patients who already require MRI as part of a standard pre-operative workup for breast cancer. However, all prior studies evaluating this possibility analyzed relatively small cohorts (ranging from 11 patients [18] to 151 patients [25]) and most did not use an independent dataset for validation. Furthermore, the methods used to determine response to NAT varied. While several studies [18, 21, 24, 26] used the response evaluation criteria in solid tumors (RECIST) [29], others [20, 22] used pathologic assessment. Additionally, the definition of complete response among these papers differed considerably.

Our study builds on past research by showing a promising association between pre-treatment MRI characteristics and pCR in breast cancer, particularly in patients with TN/HER2+ tumors. Due to our larger dataset, we were able to create an independent validation cohort for each subpopulation. With this independent validation set, we were able to validate the work of previous studies [20]. Of particular interest, no prior studies have reported the prediction of pCR response in TN/HER2+ patients using an independent test set. The extensive set of 529 features allowed for comprehensive analysis of different aspects of MRIs including size, shape, enhancement, and texture of tumors and the surrounding tissue.

## Materials and methods

### Patient population

In this retrospective institutional review board-approved study, we analyzed data for 288 patients treated with NAT. To arrive at this cohort, we reviewed 1150 consecutive female invasive breast cancer patients without prior history of breast cancer, breast surgery (definitive or non-definitive), or breast cancer therapy, with available pre-treatment, pre-operative bilateral MRI performed at Duke University Hospital between January 1, 2000 and March 23, 2014. Out of these patients, 374 were treated with NAT (including chemotherapy, endocrine, and anti-HER2neu therapies) according to medical histories and had sufficient data to establish pathologic response to NAT. To clearly distinguish between patients with pCR and those with residual invasive disease, patients with in situ carcinoma only following NAT were excluded ($n = 15$). Of the resulting 359 patients, 288 had all necessary MR sequences available and suitable for our image pre-processing steps. Details regarding patient exclusions can be found in Fig. 1.

### Pathology data collection

The pCR or pathological non-complete response (pNR) status of patients in our cohort was established using the following methods. First, pathology reports from the initial surgical intervention (mastectomy or lumpectomy) were reviewed and the tumor re-staging (ypTNM) information was collected. When a ypTNM stage was not reported, the complete pathology reports were reviewed. Then, the pCR/pNR status was identified according to the criteria in Table 1. Ultimately, pCR was defined as having no remaining invasive or in situ disease in the breast or lymph nodes, and pNR was defined as any residual invasive disease in either breast or lymph nodes.

Additionally, patients whose treatment regimens included neoadjuvant chemotherapy (NACT), either alone or in addition to other neoadjuvant therapies (including endocrine or anti-HER2neu therapies), were identified from our cohort of 288 patients. The receptor status of these patients was determined from the clinical and immunohistochemical analysis of the tissue samples obtained from these patients.

### Imaging data & annotation

For all patients, pre-treatment dynamic contrast-enhanced MR (DCE-MR) images were available with the following sequences: (a) T1-weighted fat-saturated pre-contrast sequence, (b) multiple (2–6) T1-weighted fat-saturated post-contrast sequences that were acquired following intravenous administration of a contrast agent, and (c) T1-weighted non-fat-saturated sequence. One of eight post-fellowship trained breast radiologists annotated each patient's images by placing a three-dimensional box around tumors in the subtracted sequence (obtained by subtracting the pre-contrast from the first post-contrast sequence).

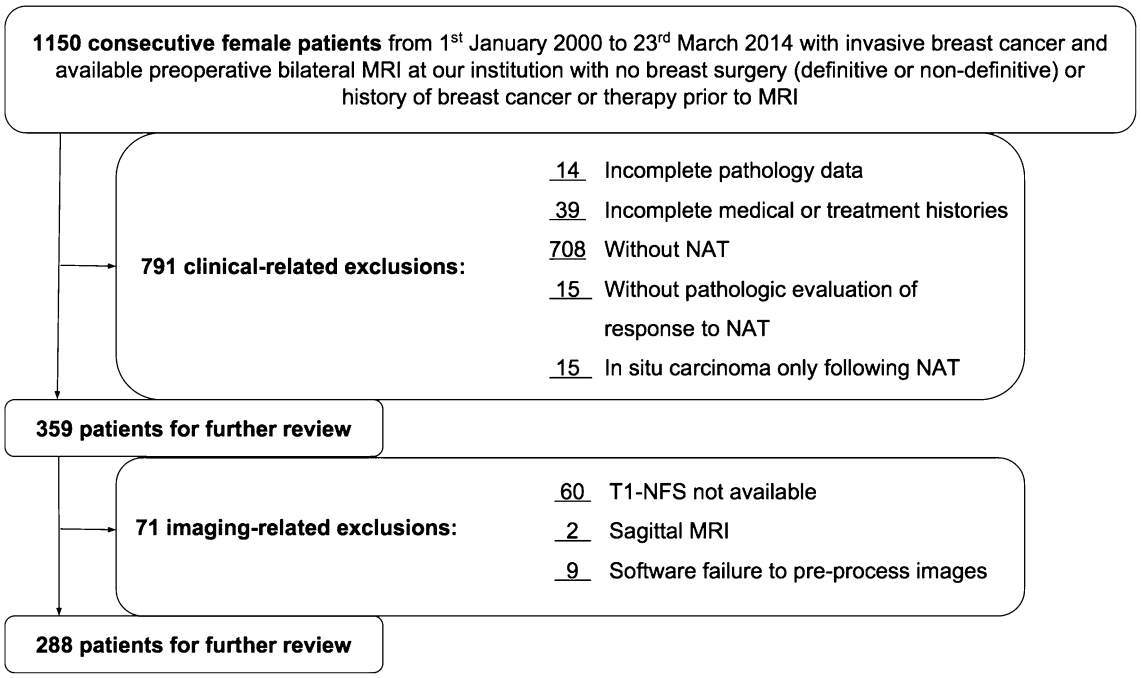


**Fig. 1** Patient population and exclusions

**Table 1** Definitions used for pCR and pNR in this study

| | Pathologic complete response (pCR) | Pathologic non-complete response (pNR) |
|---|---|---|
| ypTNM stage | ypT0, N0, M0/X (in cases where the ypTNM stage could not be confirmed as indicated by TX or NX, pathology reports were reviewed to determine response status) | ypT1 or higher or ypN1 or higher |
| Pathologic evaluation (ypTNM stage unavailable) | No remaining invasive or in situ disease in either breast or lymph nodes | Any remaining invasive disease in breast or lymph nodes noted |

## Image Processing and feature extraction

The tumor mask was obtained by applying the fuzzy C-means algorithm [30] inside the annotation box created by the radiologists for each patient. Based on the masks extracted and the available sequences, we extracted 529 radiomic features [31]. These features include breast and fibroglandular tissue (FGT) volumetric features (5), tumor size and morphology (10), tumor enhancement (30), FGT enhancement (82), tumor enhancement texture (135), FGT enhancement texture (135), tumor enhancement variation (35), FGT enhancement variation (34), spatial variation of tumor heterogeneity (4), and combination of tumor and FGT enhancement (18). Any FGT-related feature was extracted twice to include only one type of FGT mask at a time.

## Model development and statistical analysis

The entire cohort of patients was divided to form training and test sets of equal size. All machine learning models were generated using the training set only. This model generation included feature selection and training two different classifiers using the selected features as follows. Using the entire training set, a stepwise multilinear regression-based feature selection procedure was used to select features (*stepwisefit* function in Matlab 2016b, The Mathworks, Natick, MA) for predicting pCR. The selected features were used to train a multivariate logistic regression classifier (*fitglm* function in Matlab 2016b) and a support vector machine classifier (*fitcsvm* and *fitSVMposterior* functions in Matlab 2016b). The trained models were used to predict pCR in the test set. The

performance of each of the trained models was evaluated through analysis of the area under the receiver operating characteristics (ROC) curve (AUC). The AUC and its confidence intervals were calculated using the *proc* package in R [32] using the Delong method [33]. The output of the multivariate model was used as the covariate in a logistic regression model to determine the significance of its association with NAT. In addition to performing the feature selection and training classifiers for all patients in the training set, we repeated it for two additional subpopulations based on therapies received and receptor status in the training set: (a) patients who received NACT and (b) Triple-negative or HER2+ (TN/HER2+) patients who received NAT. We have not analyzed TN/HER2+ patients with NACT as a separate cohort because this group was almost identical to the second group analyzed (only one TN/HER2+ patient received NAT that did not include chemotherapy). The classifiers trained on these two subpopulations in the training set were evaluated in the corresponding subpopulations in the test set. Because we had two models per subpopulation, the significance level considered was $(0.05/(3\times2)=0.0083)$ by Bonferroni correction. Additionally, the individual AUCs of the all the selected features (from the training set) were calculated in the test set to investigate specific prognostic features as an exploratory analysis.

## Results

### Characteristics of the study population

Characteristics of all patients are shown in Table 2. Of the 288 patients reviewed, 64 patients achieved pCR and 224 had a pNR. Among all three subgroups, there was very little difference in median age, age range, racial makeup, and menopausal status. Furthermore, there was little difference in receptor status and Nottingham grade when comparing patients who received NACT to all patients. The first two subgroups differed from our third subgroup (TN/HER2+ patients who received NAT) in size (269–288 patients in the first two subpopulations and 151 patients in the third), receptor status, and tumor grade (38% of tumors in the TN/HER2+ subpopulation with grade 2 and 58% with grade 3 compared to 45–47% grade 2 and 46–48% grade 3 in the first two subpopulations), which is expected in these populations. Further details regarding the characteristics of the study population relevant to each task are shown in Table 2. The distribution of pCR and pNR in different subsets of the study population and across the training and test sets is shown in Table 3.

### Performance of prognostic models

The accuracies of the trained multivariate models in terms of AUC are shown in Table 4. The multivariate models (both SVM and logistic regression) were prognostic of pCR in TN/HER2+ patients that received neoadjuvant therapy ($p<0.002$). The prognostic value of the model predicting pCR in the entire cohort was marginally significant ($p=0.01$). Using the training set, a variable number of features were selected for the entire cohort and each of the subpopulations. In total, 12 unique features were selected for the three cohorts using the training set.

The results of our exploratory analysis using the 12 selected features are shown in Fig. 2. Two features were selected for TN/HER2+ patients who received NAT. Among these features, a tumor-based feature (change_in_variance_of_uptake) was selected for the subpopulation of all patients who received NAT as well as the TN/HER2+ subpopulation. This feature quantifies the change in variance of tumor uptake by finding the minimum ratio of the variances of tumor voxels in two consecutive time points. For this feature, lower values predict a higher chance of achieving pCR. This tumor-based feature had the highest AUC among the 12 features selected in all subpopulations evaluated. An additional feature, extracted using FGT, was also selected and found to be significant in TN/HER2+ patients who received NAT. This feature, 'SER_Partial_tissue_vol_cu_mm_T1,' is computed as a volumetric measure of FGT enhancement (extracted from T1 non-fat-saturated sequences) using the signal enhancement ratio of FGT voxels. For this feature, higher values predicted a lower chance of achieving pCR. Of all the features evaluated, 6 were extracted from tumor alone, 5 were extracted from FGT alone, and one was extracted from both tumor and FGT.

## Discussion

In this study, we analyzed 288 breast cancer patients and demonstrated that multivariate machine learning models based on MRI features are able to predict pathologic complete response to neoadjuvant therapy in breast cancer. We selected important features from an initial set of 529 features using training data and subsequently developed multivariate machine learning models. We evaluated the performance of these models in our independent test set, validating their prognostic value in TN/HER+ patients.

Our study used a set of 529 MR features extracted from the pre-treatment breast MRIs of patients in our cohort. We used feature selection [34] to identify a smaller number of features (<8) for further analysis. Models using less features ($n=2$) performed better compared to models using a higher number of features ($n=7$). The exploratory analysis

**Table 2** Clinical and pathological characteristics of study population grouped taskwise

| Details | Patients having NAT | | | Patients having NACT | | | TN/HER2+ patients having NAT | | |
|---|---|---|---|---|---|---|---|---|---|
| | Overall | Train | Test | Overall | Train | Test | Overall | Train | Test |
| N | 288 | 144 | 144 | 269 | 135 | 134 | 151 | 76 | 75 |
| Median age (years) | | 49.8 | 47.5 | | 48.6 | 46.5 | | 49.3 | 45.1 |
| Age range (years) | 24–76 | 25–76 | 24–73 | 24–76 | 25–76 | 24–73 | 24–73 | 25–72 | 24–73 |
| Race | | | | | | | | | |
| White | 184 (64%) | 89 (62%) | 95 (66%) | 169 (63%) | 81 (60%) | 88 (66%) | 98 (65%) | 49 (64%) | 49 (65%) |
| Black | 85 (30%) | 49 (34%) | 36 (25%) | 83 (31%) | 48 (36%) | 35 (26%) | 46 (30%) | 25 (33%) | 21 (28%) |
| Other | 16 (6%) | 5 (3%) | 11 (8%) | 15 (6%) | 5 (4%) | 10 (7%) | 7 (5%) | 2 (3%) | 5 (7%) |
| NA | 3 (1%) | 1 (1%) | 2 (1%) | 2 (1%) | 1 (1%) | 1 (1%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Menopausal status | | | | | | | | | |
| Pre | 158 (55%) | 77 (53%) | 81 (56%) | 155 (58%) | 75 (56%) | 80 (60%) | 89 (59%) | 44 (58%) | 45 (60%) |
| Post | 129 (45%) | 67 (47%) | 62 (43%) | 113 (42%) | 60 (44%) | 53 (40%) | 62 (41%) | 32 (42%) | 30 (40%) |
| NA | 1 (0%) | 0 (0%) | 1 (1%) | 1 (0%) | 0 (0%) | 1 (1%) | 0 (0%) | 0 (0%) | 0 (0%) |
| ER status | | | | | | | | | |
| Positive | 179 (62%) | 88 (61%) | 91 (63%) | 160 (59%) | 79 (59%) | 81 (60%) | 46 (30%) | 22 (29%) | 24 (32%) |
| Negative | 109 (38%) | 56 (39%) | 53 (37%) | 109 (41%) | 56 (41%) | 53 (40%) | 105 (70%) | 54 (71%) | 51 (68%) |
| PR status | | | | | | | | | |
| Positive | 146 (51%) | 71 (49%) | 75 (52%) | 128 (48%) | 62 (46%) | 66 (49%) | 36 (24%) | 18 (24%) | 18 (24%) |
| Negative | 142 (49%) | 73 (51%) | 69 (48%) | 141 (52%) | 73 (54%) | 68 (51%) | 115 (76%) | 58 (76%) | 57 (76%) |
| HER2 status | | | | | | | | | |
| Positive | 73 (25%) | 34 (24%) | 39 (27%) | 72 (27%) | 33 (24%) | 39 (29%) | 73 (48%) | 34 (45%) | 39 (52%) |
| Negative | 215 (75%) | 110 (76%) | 105 (73%) | 197 (73%) | 102 (76%) | 95 (71%) | 78 (52%) | 42 (55%) | 36 (48%) |
| Nottingham grade | | | | | | | | | |
| 1 | 19 (7%) | 9 (6%) | 10 (7%) | 16 (6%) | 9 (7%) | 7 (5%) | 3 (2%) | 1 (1%) | 2 (3%) |
| 2 | 135 (47%) | 74 (51%) | 61 (42%) | 121 (45%) | 66 (49%) | 55 (41%) | 58 (38%) | 33 (43%) | 25 (33%) |
| 3 | 132 (46%) | 61 (42%) | 71 (49%) | 130 (48%) | 60 (44%) | 70 (52%) | 88 (58%) | 42 (55%) | 46 (61%) |
| NA | 2 (1%) | 0 (0%) | 2 (1%) | 2 (1%) | 0 (0%) | 2 (1%) | 2 (1%) | 0 (0%) | 2 (3%) |
| Field strength | | | | | | | | | |
| 1.5T | 133 (46%) | 63 (44%) | 70 (49%) | 124 (46%) | 60 (44%) | 64 (48%) | 67 (44%) | 33 (43%) | 34 (45%) |
| 3T | 155 (54%) | 81 (56%) | 74 (51%) | 145 (54%) | 75 (56%) | 70 (52%) | 84 (56%) | 43 (57%) | 41 (55%) |
| MRI model | | | | | | | | | |
| Avanto[a] | 67 (23%) | 30 (21%) | 37 (26%) | 63 (23%) | 28 (21%) | 35 (26%) | 39 (26%) | 18 (24%) | 21 (28%) |
| OptimaMR450w[b] | 25 (9%) | 15 (10%) | 10 (7%) | 23 (9%) | 15 (11%) | 8 (6%) | 10 (7%) | 6 (8%) | 4 (5%) |
| Signa excite[b] | 1 (0%) | 0 (0%) | 1 (1%) | 1 (0%) | 0 (0%) | 1 (1%) | 1 (1%) | 0 (0%) | 1 (1%) |
| Signa HDx[b] | 91 (32%) | 51 (35%) | 40 (28%) | 85 (32%) | 49 (36%) | 36 (27%) | 42 (28%) | 25 (33%) | 17 (23%) |
| Signa HDxt[b] | 57 (20%) | 23 (16%) | 34 (24%) | 54 (20%) | 21 (16%) | 33 (25%) | 34 (23%) | 17 (22%) | 17 (23%) |
| Skyra[b] | 24 (8%) | 12 (8%) | 12 (8%) | 24 (9%) | 12 (9%) | 12 (9%) | 18 (12%) | 8 (11%) | 10 (13%) |
| Trio[a] | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| TrioTim[a] | 23 (8%) | 13 (9%) | 10 (7%) | 19 (7%) | 10 (7%) | 9 (7%) | 7 (5%) | 2 (3%) | 5 (7%) |
| Contrast | | | | | | | | | |
| Gadavist[c] | 1 (0%) | 1 (1%) | 0 (0%) | 1 (0%) | 1 (1%) | 0 (0%) | 1 (1%) | 1 (1%) | 0 (0%) |
| Magnevist[c] | 185 (64%) | 92 (64%) | 93 (65%) | 175 (65%) | 87 (64%) | 88 (66%) | 98 (65%) | 51 (67%) | 47 (63%) |
| Multihance[d] | 85 (30%) | 46 (32%) | 39 (27%) | 79 (29%) | 43 (32%) | 36 (27%) | 45 (30%) | 22 (29%) | 23 (31%) |
| Unknown | 17 (6%) | 5 (3%) | 12 (8%) | 14 (5%) | 4 (3%) | 10 (7%) | 7 (5%) | 2 (3%) | 5 (7%) |

*NA* information was not available, *Grade* nottingham grade

[a]Siemens, Munich, Germany

[b]GE Healthcare, Little Chalfont, UK

[c]Bayer Healthcare, Berlin, Germany

[d]Bracco, Milan, Italy; Age measured in years

**Table 3** Distribution of pCR and pNR in different subsets of the study population

| | All patients with NAT | All patients with NACT | TN/HER2$^+$ patients with NAT |
|---|---|---|---|
| Training set | | | |
| pCR | 26 | 26 | 22 |
| pNR | 118 | 109 | 54 |
| Test set | | | |
| pCR | 38 | 38 | 28 |
| pNR | 106 | 96 | 47 |
| Total | 288 | 269 | 151 |

showed three features that were independently prognostic of pCR in the test set. For each of our subpopulations, features extracted from tumor and FGT were selected to be included in our models and were combined to form a comprehensive set of features. Although Fan et al. used a similar combination of tumor and background parenchymal features to perform analyses, our study used the clinically preferred pathologic assessment to determine response rather than the RECIST criteria based on imaging [21].

While several studies use radiomic MRI features to *monitor* response [15, 18, 35, 36], our goal was to use MRI to *predict* tumor response. Published studies with this goal

**Table 4** Performance of the multivariate models in the test set: areas under the ROC, 95% confidence interval, and *p*-value

| | Neoadjuvant therapy | Neoadjuvant chemotherapy | Neoadjuvant therapy in TN/HER2$^+$ |
|---|---|---|---|
| Number of features selected for multivariate modeling | 2 | 7 | 4 |
| Logistic regression-based classifier | 0.658 (0.556–0.760, $p=0.0106$) | 0.589 (0.484–0.694, $p=0.140$) | 0.707 (0.582–0.832, $p<0.002$) |
| Support vector machine | 0.593 (0.482–0.703, $p=0.136$) | 0.593 (0.489–0.694, $p=0.127$) | 0.705 (0.581–0.829, $p<0.002$) |

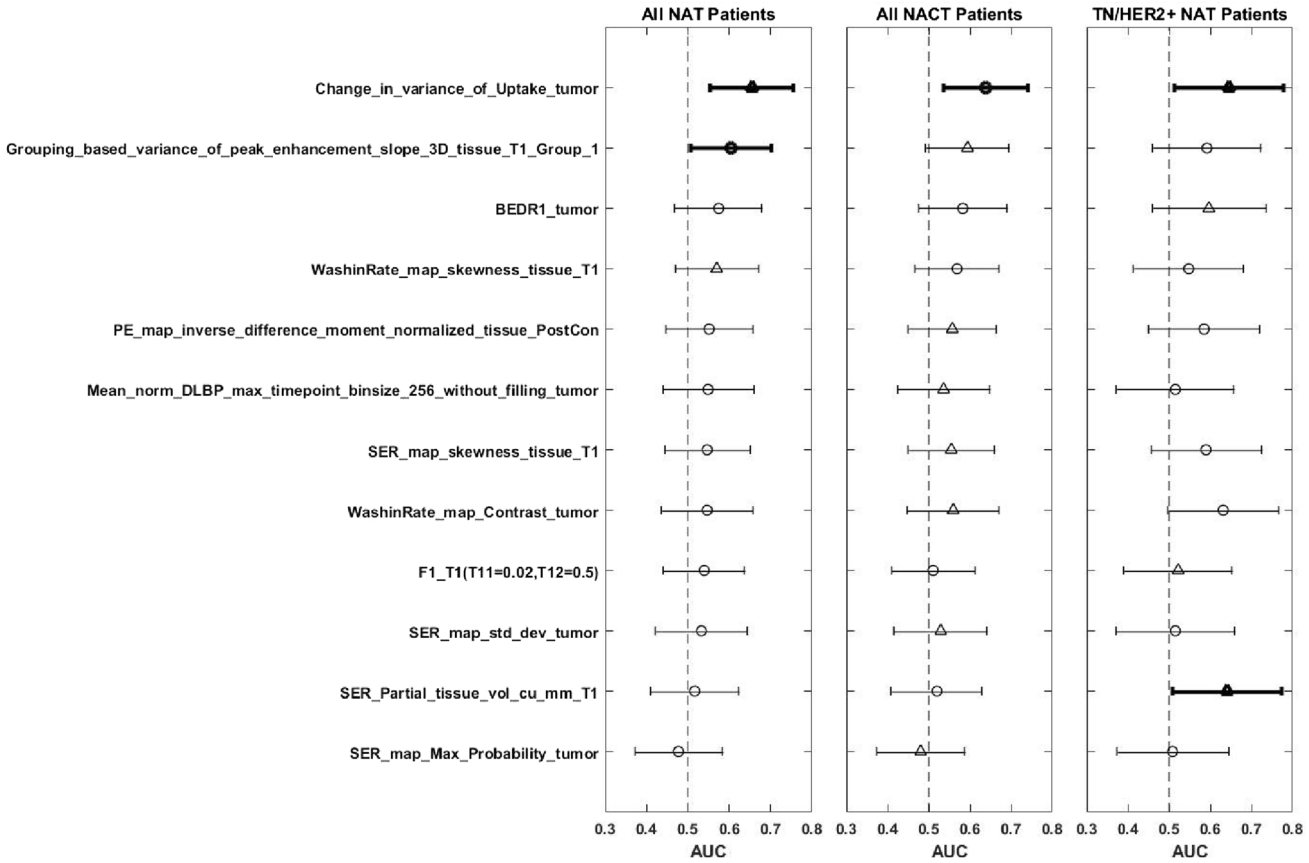


**Fig. 2** AUC and CI of individual features selected for at least one of the four subgroups used for predicting pCR in the test set. The triangular marker indicates that a particular feature was selected in the training set for the subgroup. Bold indicates that a feature's confidence interval does not include 0.5

are limited by small cohort sizes [18, 20–22, 24, 25]. Furthermore, most of these studies used RECIST to measure response [18, 21, 24, 25], even though pCR is highly correlated with event-free survival and overall survival [1].

Studies using pathologic assessment to determine response have varying definitions of pCR [20, 22]. In some of these studies patients with positive lymph nodes [20, 22], residual DCIS [20, 22], or even residual invasive carcinoma in the breast [22] are categorized as having pCR. In an attempt to differentiate between NAT patients in a clinically relevant manner, we defined pCR as having no residual invasive or in situ carcinoma in the breast or lymph nodes [1].

Our study is the first of this type to incorporate a broader definition of NAT. Our study included patients who received endocrine and anti-HER2neu therapy in the neoadjuvant setting, whether or not these patients received neoadjuvant chemotherapy (anthracycline, platinum, or taxane-based chemotherapy). However, in addition to the more inclusive analysis of all NAT patients ($n = 288$), we separately analyzed the sub-cohort of patients who received NACT ($n = 269$).

Due to our larger cohort, we were able to evaluate the specific group of TN/HER2+ patients who received NAT (only one patient in this cohort did not receive NACT, which is expected [37–43]) separately. Furthermore, we were able to validate our findings for this subpopulation using an independent test set, which past studies [20] were unable to do. We found a significant association between our multivariate models and pCR in TN/HER2+ patients. These findings are important because TN/HER2+ patients achieve higher rates of pCR compared to HR+/HER2- cancers and have improved disease free and overall survival after a pCR [28, 38, 44, 45]. While there is no current method to predict pCR on an individual basis for TN/HER2+ 2+ patients, our study shows promise that pre-treatment MRI might serve this purpose.

## Conclusion

While breast cancer patients who achieve pathologic complete response to neoadjuvant therapy have significant improvements in both event-free survival and overall survival, only 10–50% of patients achieve this level of response to NAT [46]. While patients not having a complete response may still derive benefit, their treatment is, by definition, sub-optimal. It is prudent to find effective means of predicting response to neoadjuvant therapy prior to initiating this treatment. Such a predictive tool would allow for early identification of non-responders, allowing for more personalized medical care. Our study used machine learning models that combined MRI-based features for this purpose. We found this method successfully predicted pCR in TN/HER2+ patients who had NAT. Our findings highlight the potential use of pre-treatment breast MRI as a means to stratify breast cancer patients to more appropriate treatment regimens, allowing for more personalized care and improved quality of life.

**Acknowledgements** Funding was provided by North Carolina Biotechnology Center (Grant No. 2016-BIG-6520), National Institutes of Health (Grant No. 1R01EB021360).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. von Minckwitz G, Untch M, Blohmer J-U et al (2012) Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol 30(15):1796–1804. https://doi.org/10.1200/JCO.2011.38.8595
2. Kaufmann M, von Minckwitz G, Bear HD et al (2007) Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006. Ann Oncol 18(12):1927–1934. https://doi.org/10.1093/annonc/mdm201
3. Kaufmann M, von Minckwitz G, Smith R et al (2003) International expert panel on the use of primary (preoperative) systemic treatment of operable breast cancer: review and recommendations. J Clin Oncol 21(13):2600–2608. https://doi.org/10.1200/JCO.2003.01.136
4. Heys SD, Hutcheon AW, Sarkar TK et al (2002) Neoadjuvant docetaxel in breast cancer: 3-year survival results from the Aberdeen trial. Clin Breast Cancer 3:S69–S74. https://doi.org/10.3816/CBC.2002.s.015
5. van der Hage JH, van de Velde CC, Mieog SJ (2007) Preoperative chemotherapy for women with operable breast cancer. In: Mieog SJ (ed) Cochrane database of systematic reviews. John Wiley & Sons, Ltd, Chichester. https://doi.org/10.1002/14651858.CD005002.pub2
6. Fisher B, Bryant J, Wolmark N et al (1998) Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 16(8):2672–2685. https://doi.org/10.1200/JCO.1998.16.8.2672
7. van der Hage JA, van de Velde CJ, Julien JP et al (2001) Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. J Clin Oncol 19(22):4224–4237. https://doi.org/10.1200/JCO.2001.19.22.4224
8. Michoux N, Van den Broeck S, Lacoste L et al (2015) Texture analysis on MR images helps predicting non-response to NAC in breast cancer. BMC Cancer 15(1):574. https://doi.org/10.1186/s12885-015-1563-8
9. Barbi GP, Marroni P, Bruzzi P, Nicolò G, Paganuzzi M, Ferrara GB (1987) Correlation between steroid hormone receptors and prognostic factors in human breast cancer. Oncology 44(5):265–269. https://doi.org/10.1159/000226492
10. von Minckwitz G, Sinn H-P, Raab G et al (2008) Clinical response after two cycles compared to HER2, Ki-67, p53, and bcl-2 in independently predicting a pathological complete response after preoperative chemotherapy in patients with

operable carcinoma of the breast. Breast Cancer Res 10(2):R30. https://doi.org/10.1186/bcr1989
11. Esserman L, Kaplan E, Partridge S et al (2001) MRI phenotype is associated with response to doxorubicin and cyclophosphamide neoadjuvant chemotherapy in stage III breast cancer. Ann Surg Oncol 8(6):549–559. https://doi.org/10.1007/s10434-001-0549-8
12. Nishimura R, Osako T, Okumura Y, Hayashi M, Arima N (2010) Clinical significance of Ki-67 in neoadjuvant chemotherapy for primary breast cancer as a predictor for chemosensitivity and for prognosis. Breast Cancer 17(4):269–275. https://doi.org/10.1007/s12282-009-0161-5
13. Fangberget A, Nilsen LB, Hole KH et al (2011) Neoadjuvant chemotherapy in breast cancer-response evaluation and prediction of response to treatment using dynamic contrast-enhanced and diffusion-weighted MR imaging. Eur Radiol 21(6):1188–1199. https://doi.org/10.1007/s00330-010-2020-3
14. Uematsu T, Kasami M, Yuen S (2010) Neoadjuvant chemotherapy for breast cancer: correlation between the baseline MR imaging findings and responses to therapy. Eur Radiol 20(10):2315–2322. https://doi.org/10.1007/s00330-010-1813-8
15. Pickles MD, Manton DJ, Lowry MTL (2009) Prognostic value of pre-treatment DCE-MRI parameters in predicting disease free and overall survival for breast cancer patients undergoing neoadjuvant chemotherapy. Eur J Radiol 71(3):498–505. https://doi.org/10.1016/J.EJRAD.2008.05.007
16. Craciunescu OI, Blackwell KL, Jones EL et al (2009) DCE-MRI parameters have potential to predict response of locally advanced breast cancer patients to neoadjuvant chemotherapy and hyperthermia: A pilot study. Int J Hyperth 25(6):405–415. https://doi.org/10.1080/02656730903022700
17. Smith IC, Heys SD, Hutcheon AW et al (2002) Neoadjuvant chemotherapy in breast cancer: significantly enhanced response with docetaxel. J Clin Oncol 20(6):1456–1466. https://doi.org/10.1200/JCO.2002.20.6.1456
18. Kawamura M, Satake H, Ishigaki S, Nishio A, Sawaki M, Naganawa S (2011) Early prediction of response to neoadjuvant chemotherapy for locally advanced breast cancer using MRI. Nagoya J Med Sci 73(3–4):147–156
19. Goorts B, van Nijnatten TJA, de Munck L et al (2017) Clinical tumor stage is the most important predictor of pathological complete response rate after neoadjuvant chemotherapy in breast cancer patients. Breast Cancer Res Treat 163(1):83–91. https://doi.org/10.1007/s10549-017-4155-2
20. Braman NM, Etesami M, Prasanna P et al (2017) Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res 19(1):57. https://doi.org/10.1186/s13058-017-0846-1
21. Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L (2017) Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. Eur J Radiol 94:140–147. https://doi.org/10.1016/J.EJRAD.2017.06.019
22. Eom H-J, Cha JH, Choi WJ, Chae EY, Shin HJ, Kim HH (2017) Predictive clinicopathologic and dynamic contrast-enhanced MRI findings for tumor response to neoadjuvant chemotherapy in triple-negative breast cancer. Am J Roentgenol 208(6):W225–W230. https://doi.org/10.2214/AJR.16.17125
23. Chamming's F, Ueno Y, Ferré R et al (2017) Features from computerized texture analysis of breast cancers at pretreatment MR imaging are associated with response to neoadjuvant chemotherapy. Radiology. https://doi.org/10.1148/radiol.2017170143
24. Aghaei F, Tan M, Hollingsworth AB, Qian W, Liu H, Zheng B (2015) Computer-aided breast MR image feature analysis for prediction of tumor response to chemotherapy. Med Phys 42(11):6520–6528. https://doi.org/10.1118/1.4933198
25. Aghaei F, Tan M, Hollingsworth AB, Zheng B, Cheng S (2016) Computer-aided global breast MR image feature analysis for prediction of tumor response to chemotherapy: performance assessment. In: Tourassi GD, Armato SG (eds) International society for optics and photonics. https://doi.org/10.1117/12.2216326
26. Ahmed A, Gibbs P, Pickles M, Turnbull L (2013) Texture analysis in assessment and prediction of chemotherapy response in breast cancer. J Magn Reson Imaging 38(1):89–101. https://doi.org/10.1002/jmri.23971
27. Nilsen L, Fangberget A, Geier O, Olsen DR, Seierstad T (2010) Diffusion-weighted magnetic resonance imaging for pretreatment prediction and monitoring of treatment response of patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. Acta Oncol (Madr) 49(3):354–360. https://doi.org/10.3109/02841861003610184
28. Tudorica A, Oh KY, Chui SY-C et al (2016) Early prediction and evaluation of breast cancer response to neoadjuvant chemotherapy using quantitative DCE-MRI 1. Transl Oncol 9:8–17. https://doi.org/10.1016/j.tranon.2015.11.016
29. New response evaluation criteria in solid tumours (2009) Revised RECIST guideline (version 1.1). Eur J Cancer 45(2):228–247. https://doi.org/10.1016/J.EJCA.2008.10.026
30. Bezdek JC (2013) Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media, New York
31. Saha A, Yu X, Sahoo D, Mazurowski MA (2017) Effects of MRI scanner parameters on breast cancer radiomics. Expert Syst Appl 87:384–391. https://doi.org/10.1016/j.eswa.2017.06.029
32. Team RC (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna 2014.
33. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44(3):837. https://doi.org/10.2307/2531595
34. Guyon I, Elisseeff A, De AM (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
35. Dave RV, Millican-Slater R, Dodwell D, Horgan K, Sharma N (2017) Neoadjuvant chemotherapy with MRI monitoring for breast cancer. Br J Surg 104(9):1177–1187. https://doi.org/10.1002/bjs.10544
36. Sharma U, Danishad KKA, Seenu V, Jagannathan NR (2009) Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. NMR Biomed 22(1):104–113. https://doi.org/10.1002/nbm.1245
37. Cleator S, Heller W, Coombes RC (2007) Triple-negative breast cancer: therapeutic options. Lancet Oncol 8(3):235–244. https://doi.org/10.1016/S1470-2045(07)70074-8
38. Liedtke C, Mazouni C, Hess KR et al (2008) Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol 26(8):1275–1281. https://doi.org/10.1200/JCO.2007.14.4147
39. Carey LA, Dees EC, Sawyer L et al (2007) The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res 13(8):2329–2334. https://doi.org/10.1158/1078-0432.CCR-06-1109
40. Gianni L, Eiermann W, Semiglazov V et al (2010) Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trial with a parallel HER. Lancet 375(9712):377–384. https://doi.org/10.1016/S0140-6736(09)61964-4
41. Buzdar AU, Suman VJ, Meric-Bernstam F et al (2013) Fluorouracil, epirubicin, and cyclophosphamide (FEC-75) followed by paclitaxel plus trastuzumab versus paclitaxel plus trastuzumab followed by FEC-75 plus trastuzumab as neoadjuvant treatment

for patients with HER2-positive breast cancer (Z1041): a random. Lancet Oncol 14(13):1317–1325. https://doi.org/10.1016/S1470-2045(13)70502-3

42. Robidoux A, Tang G, Rastogi P et al (2013) Lapatinib as a component of neoadjuvant therapy for HER2-positive operable breast cancer (NSABP protocol B-41): an open-label, randomised phase 3 trial. Lancet Oncol 14(12):1183–1192. https://doi.org/10.1016/S1470-2045(13)70411-X
43. Untch M, Loibl S, Bischoff J et al (2012) Lapatinib versus trastuzumab in combination with neoadjuvant anthracycline-taxane-based chemotherapy (GeparQuinto, GBG 44): a randomised phase 3 trial. Lancet Oncol 13(2):135–144. https://doi.org/10.1016/S1470-2045(11)70397-7
44. Cortazar P, Zhang L, Untch M et al (2014) Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet 384(9938):164–172. https://doi.org/10.1016/S0140-6736(13)62422-8
45. Chen X, Ye G, Zhang C et al (2013) Superior outcome after neoadjuvant chemotherapy with docetaxel, anthracycline, and cyclophosphamide versus docetaxel plus cyclophosphamide: results from the NATT trial in triple negative or HER2 positive breast cancer. Breast Cancer Res Treat 142(3):549–558. https://doi.org/10.1007/s10549-013-2761-1
46. Earl H, Provenzano E, Abraham J et al (2015) Neoadjuvant trials in early breast cancer: pathological response at surgery and correlation to longer term outcomes—what does it all mean? BMC Med 13(1):234. https://doi.org/10.1186/s12916-015-0472-7