Radiology

ORIGINAL RESEARCH • BREAST IMAGING

Texture Analysis with 3.0-T MRI for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer

Na Lae Eun, MD • Daesung Kang, PhD • Eun Ju Son, MD • Jeong Seon Park, MD • Ji Hyun Youk, MD • Jeong-Ah Kim, MD • Hye Mi Gweon, MD

From the Department of Radiology, Gangnam Severance Hospital, Yonsei University College of Medicine, 211 Eonju-ro, Gangnam-gu, 06273 Seoul, Republic of Korea (N.L.E., E.J.S., J.H.Y., J.A.K., H.M.G.); Department of Radiology, Hanyang University, College of Medicine, Seoul, Republic of Korea (N.L.E., J.S.P.); and Department of Healthcare Information Technology, Inje University, Gimhae, Republic of Korea (D.K.). Received November 28, 2018; revision requested January 8, 2019; final revision received August 16; accepted October 7. **Address correspondence to** H.M.G. (e-mail: *hyemig@yuhs.ac*).

Conflicts of interest are listed at the end of this article.

Radiology 2020; 294:31–41 • https://doi.org/10.1148/radiol.2019182718 • Content code: BR

Background: Previous studies have suggested that texture analysis is a promising tool in the diagnosis, characterization, and assessment of treatment response in various cancer types. Therefore, application of texture analysis may be helpful for early prediction of pathologic response in breast cancer.

Purpose: To investigate whether texture analysis of features from MRI is associated with pathologic complete response (pCR) to neoadjuvant chemotherapy (NAC) in breast cancer.

Materials and Methods: This retrospective study included 136 women (mean age, 47.9 years; range, 31–70 years) who underwent NAC and subsequent surgery for breast cancer between January 2012 and August 2017. Patients were monitored with 3.0-T MRI before (pretreatment) and after (midtreatment) three or four cycles of NAC. Texture analysis was performed at pre- and midtreatment T2-weighted MRI, contrast material–enhanced T1-weighted MRI, diffusion-weighted MRI, and apparent diffusion coefficient (ADC) mapping by using commercial software. A random forest method was applied to build a predictive model for classifying those with pCR with use of texture parameters. Diagnostic performance for predicting pCR was assessed and compared with that of six other machine learning classifiers (adaptive boosting, decision tree, k-nearest neighbor, linear support vector machine, naive Bayes, and linear discriminant analysis) by using the Wald test and DeLong method.

Results: Forty of the 136 patients (29%) achieved pCR after NAC. In the prediction of pCR, the random forest classifier showed the lowest diagnostic performance with pretreatment ADC (area under the receiver operating characteristic curve [AUC], 0.53; 95% confidence interval: 0.44, 0.61) and the highest diagnostic performance with midtreatment contrast-enhanced T1-weighted MRI (AUC, 0.82; 95% confidence interval: 0.74, 0.88) among pre- and midtreatment T2-weighted MRI, contrast-enhanced T1-weighted MRI, diffusion-weighted MRI, and ADC mapping.

Conclusion: Texture parameters using a random forest method of contrast-enhanced T1-weighted MRI at midtreatment of neoadjuvant chemotherapy were valuable and associated with pathologic complete response in breast cancer.

Online supplemental material is available for this article.

Neoadjuvant chemotherapy (NAC) is commonly accepted as the treatment of choice for large operable or locally advanced breast cancers (1,2). NAC enables the downsizing of breast cancers, thus allowing breast-conserving surgery and assessment of response to chemotherapy during treatment. Furthermore, the achievement of pathologic complete response (pCR) is an independent predictor of better disease-free survival; in the future, surgery may not be required for patients showing pCR after NAC (3,4). Therefore, early identification of such patients is crucial for improving and personalizing patient treatment.

Several studies have investigated the value of breast MRI for assessing or predicting treatment response to NAC (5–7). Recent studies showed that changes in specific morphologic and kinetic parameters extracted from dynamic contrast material–enhanced MRI, as well as diffusion-weighted imaging (DWI), could help predict treatment response and pCR (8–10). However, MRI has limitations when used clinically because image interpretation is based on the radiologist's visual assessment.

Texture analysis is a mathematic model that allows evaluation of gray-level intensity and position of pixels, as well as the relationships among voxel intensities (11). Texture analysis in medical imaging is being applied for diagnosing, characterizing, and monitoring treatment response by quantifying lesion heterogeneity and irregularity of tissue components (12–14). Several studies have evaluated the use of texture analysis in MRI for predicting the response of breast cancer to NAC (15–24). However, these studies had relatively small cohorts and did not compare the diagnostic performance among various sequences. Therefore, we explored the application of texture analysis in a larger patient cohort by using pre- and midtreatment (after three or four cycles of NAC) 3.0-T MRI with T2-weighted imaging, dynamic contrast-enhanced T1-weighted imaging, DWI, and apparent diffusion coefficient (ADC) mapping. We hypothesized that texture features at pre- or midtreatment MRI could be associated with pCR after NAC. The aim of this study was to investigate whether texture

Abbreviations

ADC = apparent diffusion coefficient, AUC = area under the receiver operating characteristic curve, DWI = diffusion-weighted imaging, HER2 = human epidermal growth factor 2, NAC = neoadjuvant chemotherapy, NPV = negative predictive value, pCR = pathologic complete response, PPV = positive predictive value, SSF = spatial scale filter

Summary

Texture features of contrast material–enhanced T1-weighted MRI at midtreatment of neoadjuvant chemotherapy showed good diagnostic performance for demonstrating complete pathologic response in breast cancer.

Key Results

- Texture features at midtreatment contrast material–enhanced T1-weighted MRI showed the highest diagnostic performance (area under the receiver operating characteristic curve [AUC], 0.82) for predicting complete pathologic response compared with features at pre- and midtreatment T2-weighted, contrast-enhanced T1-weighted, and diffusion-weighted imaging.
- The random forest model (AUC, 0.82) had better diagnostic performance for showing association with complete pathologic response compared with six other machine learning classifiers (AUCs: adaptive boosting, 0.76; decision tree, 0.70; k-nearest neighbor, 0.80; linear support vector machine, 0.75; naive Bayes, 0.74; linear discriminant analysis, 0.79) at midtreatment contrast-enhanced T1-weighted MRI.

analysis of features from MRI is associated with pCR to NAC in breast cancer and whether changes in these features can help predict pCR after NAC.

Materials and Methods

Study Participants

The institutional review board of Gangnam Severance Hospital approved this retrospective study and waived the requirement to obtain patient approval or written informed consent for the review of medical records or images. Between January 2012 and August 2017, we enrolled 158 women with breast cancer who received NAC. The eligibility criteria were as follows: *(a)* breast cancer was pathologically confirmed with core needle biopsy, *(b)* patients underwent NAC, and *(c)* MRI was performed during NAC. At the completion of chemotherapy, the patients underwent surgery. Of the 158 women, 22 were excluded for the following reasons: *(a)* images were obtained with a 1.5-T MRI system (n = 3), *(b)* MRI was performed at an outside hospital (n = 2), *(c)* no midtreatment MRI data were available (n = 5), *(d)* there was a complete radiologic response at midtreatment MRI with complete resolution of the known cancer (n = 8), and *(e)* nonstandard treatment was used (n = 4) (Fig 1).

NAC was performed according to the standard protocol at our institution (mean number of cycles, seven; range: six to 12 cycles), as follows: *(a)* doxorubicin (50 mg/m^2) with docetaxel (75 mg/m^2) was given intravenously every 3 weeks for six cycles; *(b)* doxorubicin (60 mg/m^2) with cyclophosphamide (600 mg/m^2) was given intravenously every 3 weeks for six cycles; *(c)* doxorubicin (60 mg/m^2) with cyclophosphamide (600 mg/m^2) plus docetaxel (75 mg/m^2) was given intravenously every 3 weeks for four cycles, followed by docetaxel intravenously every 3 weeks for four cycles; or *(d)* docetaxel (75 mg/m^2), carboplatin (area under the plasma drug concentration-time curve, 6), and a combination with trastuzumab (8 mg/kg at cycle 1, 6 mg/kg at cycles 2–6) and pertuzumab (840 mg at cycle 1, 420 mg at cycles 2–6) was given intravenously every 3 weeks for six cycles.

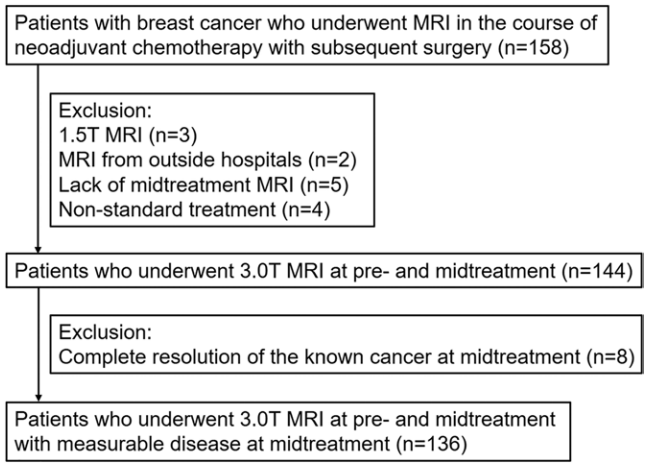

Figure 1: Flowchart shows study population and exclusion criteria.

MRI Technique

All MRI examinations were performed by using a 3.0-T system (Achieva, Philips Medical Systems, Best, the Netherlands; Discovery MR750, GE Medical Systems, Milwaukee, Wis) with a dedicated four-channel breast coil. All images were obtained with bilateral axial views. The routine protocol was composed of turbo spin-echo T1-weighted (repetition time msec/echo time msec, 505/10; matrix, 564 × 338; field of view, 20–34 cm; slice thickness, 3 mm) and T2-weighted fat-suppressed spin-echo (5506/70; matrix, 564 × 261; field of view, 20–34 cm; slice thickness, 3 mm) sequences. Echo-planar DWI was performed before contrast material enhancement (12 500/78; matrix, 96 × 126; field of view, 32 cm; slice thickness, 3 mm) with b values of 0 and 1000 sec/mm^2. Dynamic contrast-enhanced MRI was performed with one precontrast and five postcontrast series by using a T1-weighted gradient-echo sequence (5/2.5; matrix, 340 × 274; flip angle, 12°; field of view, 34 cm; slice thickness, 2 mm). Image subtraction was performed at all postcontrast phases. Dynamic contrast-enhanced MRI was performed after injecting gadopentetate dimeglumine (Bonogen I; Central Medical Service, Seoul, South Korea) or gadobutrol (Gadovist; Bayer Healthcare, Berlin, Germany) at a dose of 0.1 mmol/kg by using an automated injector (Nemoto; Nemoto Kyorindo, Tokyo, Japan) at a rate of 2 mL/sec, followed by a 20-mL saline flush.

Image Analysis

Images were interpreted by two radiologists (N.L.E. and H.M.G., with 7 and 12 years of experience in breast imaging). Maximum tumor diameter was measured on contrast-enhanced T1-weighted images. Tumor appearance at MRI (mass, nonmass

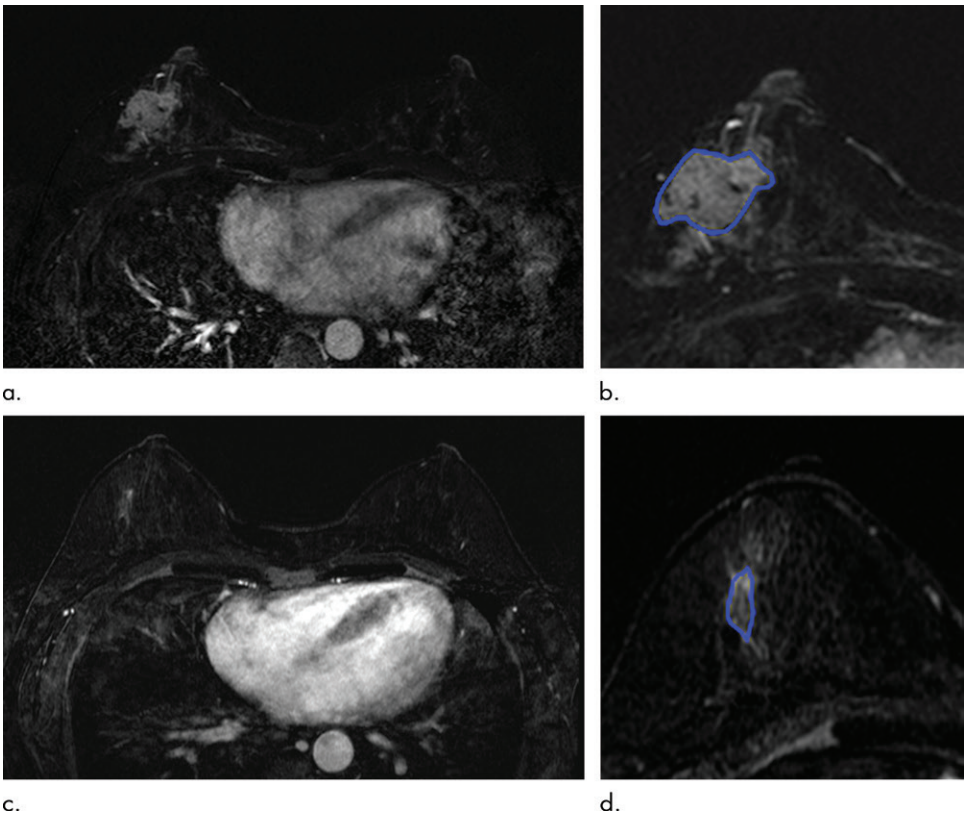

Figure 2: Texture analysis at contrast-enhanced T1-weighted MRI in a 51-year-old woman with cancer of the right breast. The patient achieved pathologic complete response (pCR) after neoadjuvant chemotherapy (NAC). **(a)** Image obtained at pretreatment MRI shows irregular, heterogeneously enhancing cancer in upper outer quadrant of right breast. **(b)** Tumor was manually delineated at pretreatment MRI, and texture analysis was performed. **(c)** Image obtained at midtreatment MRI shows that the tumor has decreased in size. **(d)** Tumor was manually delineated at midtreatment MRI, and texture analysis was performed. Random forest model at midtreatment MRI predicted pCR, which was confirmed at final surgical-pathologic evaluation after NAC.

enhancement, mass and nonmass enhancement) was visually assessed by the two radiologists in consensus.

Texture analysis of T2-weighted images, contrast-enhanced T1-weighted images, DWI scans, and ADC maps was performed by using a commercial research software algorithm (TexRAD; Feedback Medical, Somerset, England). The two radiologists manually drew a region of interest around the whole visible tumor in consensus (Figs 2, 3). All images were aligned by using the first dynamic contrast-enhanced T1-weighted images as a reference. A two-dimensional region of interest enclosing the largest cross-sectional area of the tumor was delineated on T2-weighted images, contrast-enhanced T1-weighted images, DWI scans, and ADC maps. If the cancer was multifocal or multicentric, the region of interest was measured at the tumor with the largest size. On T2-weighted images, the tumor area was defined as areas with low or intermediate signal intensity compared with normal breast parenchyma. On dynamic contrast-enhanced images, an area of high signal intensity with homogeneous or heterogeneous enhancement that differed from that of normal background parenchyma at the first dynamic contrast-enhanced series was considered a tumor. On images obtained at DWI, high signal intensity suggestive of the diffusion-restricted areas with low signal intensity on ADC maps was regarded as a tumoral region. If the lesion did not appear on T2-weighted images, DWI scans, and ADC maps, the region of interest was defined by referencing the lesion location on a contrast-enhanced T1-weighted image.

After tumors were segmented, the following six texture parameters were calculated on the basis of the gray-level intensity histogram: mean pixel intensity, standard deviation, mean proportion of positive pixels, entropy (irregularity of gray-level distribution), skewness (asymmetry of the histogram), and kurtosis (peakedness of the histogram). The TexRAD software displays the imaging texture parameters through an initial filtration step called a spatial scale factor (SSF). An SSF of 2 highlights a fine-texture feature with a radius of 2 mm, an SSF of 3 or 4 highlights a medium-texture feature with a radius of 3–4 mm, and an SSF of 5 or 6 highlights a coarse-texture feature with a radius of 5–6 mm. The effect of filtration, by highlighting larger pixels, has been hypothesized to accentuate the contribution of the vasculature to texture features (15,25). In our study, images were analyzed by simultaneously choosing SSFs of 2 (highlighting fine-texture features), 4 (highlighting medium-texture features), and 6 (highlighting coarse-texture features).

Histopathologic Assessment

One pathologist with 30 years of experience in breast pathology assessed the histologic results. The histologic type and grade were evaluated from histopathologic reports of US-guided core biopsies performed before NAC. The expression of estrogen receptor, progesterone receptor, and human epidermal growth factor 2 (HER2) was assessed by using the standard avidin-biotin complex immunohistochemical staining method. Estrogen receptor and progesterone receptor positivity were assessed with the Allred score, which rates the proportion of positive cells (on a scale of 0–5) and the staining intensity (on a scale of 0–3). Tumors were considered estrogen receptor or progesterone receptor positive if the Allred score exceeded 3. Tumors were considered HER2 positive if they had a score of 3+ at immunohistochemical examination. If HER2 status was equivocal (score, 2+) at immuno-

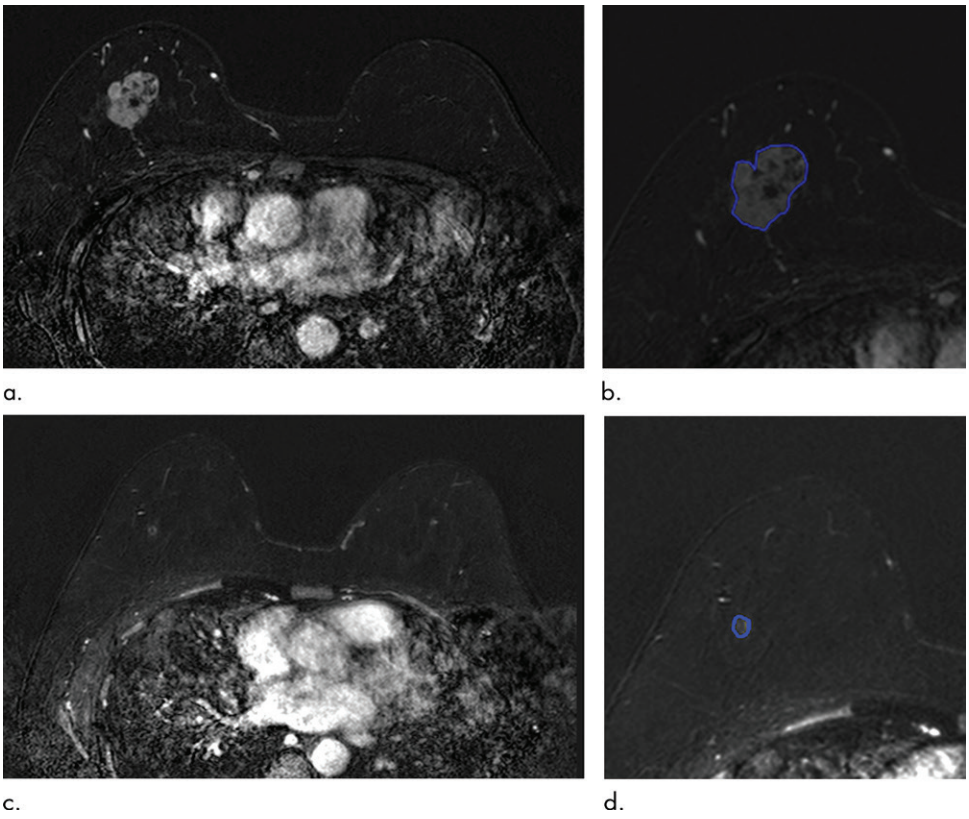

Figure 3: Texture analysis at contrast-enhanced T1-weighted MRI in a 48-year-old woman with cancer of the right breast. The patient achieved pathologic complete response (pCR) after neoadjuvant chemotherapy (NAC). **(a)** Image obtained at pretreatment MRI shows irregular heterogeneously enhancing cancer in upper outer quadrant of right breast. **(b)** Tumor was manually delineated at pretreatment MRI, and texture analysis was performed. **(c)** Image obtained at midtreatment MRI shows that the tumor has decreased in size, demonstrating focal nonmass enhancement. **(d)** Tumor was manually delineated at midtreatment MRI. Random forest model at midtreatment MRI predicted nonpathologic complete response; however, final surgical-pathologic evaluation after NAC revealed pCR.

histochemical examination, fluorescence in situ hybridization analysis was performed to confirm the diagnosis.

After surgery, pCR was defined as the absence of residual invasive cancer in the breast surgical specimen (ductal carcinoma in situ could be present) with the absence of axillary lymph node involvement. The isolated tumor cell of the lymph node was considered axillary pCR.

Statistical Analysis

Clinical-pathologic characteristics and pretreatment MRI findings, including maximum tumor diameter and type of lesion, were collected and compared according to pathologic response by using the *t* test, χ^2 test, or Fisher exact test.

For analyzing texture parameters at pretreatment and midtreatment MRI and the difference of texture parameters between mid- and pretreatment MRI, a random forest method was applied to build a predictive model to classify pCR responders. This method combines the results of many individual decision trees to reduce overfitting and to improve generalization. Random forest method grows the decision trees in the ensemble by using bootstrap samples of the data and selects a random subset of predictors to use at each decision split, which also enables an automatic measure of feature importance (25–27). The measure of feature importance was based on the percentage increase in prediction error when values of that feature were permuted randomly compared with the prediction error with that value of that feature left intact. This measure was computed for every tree, then averaged over the entire ensemble and divided by the standard deviation over the entire ensemble (26). The random forest classifier was optimized for the number of trees (51, 251, 501, 751, 1001, 2001) with repeated (n = 100) and fivefold cross-validation. To train the random forest model, repeated cross-validation was used for the less biased estimates compared to use of out-of-bag estimates (27,28). For each cross-validation, the 18 features extracted from TexRad were centered and scaled because the range of feature values varied widely. For each combination of MRI type (contrast-enhanced T1-weighted imaging, T2-weighted imaging, DWI, ADC mapping) and feature type (pretreatment, midtreatment, and difference between mid- and pretreatment), we computed the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and area under the receiver operating characteristic curve (AUC) from the receiver operating characteristic for predicting pCR.

For comparison between the random forest model and other machine learning methods, we investigated six machine learning algorithms: k-nearest neighbor, naive Bayes classifier, decision tree, linear discriminant analysis, adaptive boosting, and linear support vector machine. Details on the individual classifiers are provided in Appendix E1 (online). To train the six machine learning algorithms, repeated (n = 100) and fivefold cross-validation was applied. Each classifier adopted different types of hyperparameters to obtain reasonable results.

The χ^2 test was used to compare the diagnostic performance (ie, sensitivity, specificity, PPV, NPV, and accuracy) by using SPSS software (version 25.0; IBM, Armonk, NY). The Wald test was performed to compare AUCs in independent data sets, and the DeLong method was used to compare AUCs in independent data sets by using MedCalc software (version 18.9.1; MedCalc, Ostend, Belgium). The false discovery rate correction was also performed for multiple comparison (29). $P < .05$

Table 1: Patient Characteristics

Characteristics	Non-pCR Group (n = 96)	pCR Group (n = 40)	P Value
Age (y)*	48 ± 10	48 ± 9	.66
Sex			
Men	0 (0)	0 (0)	
Women	96 (100)	40 (100)	
Interval between pretreatment and midtreatment MRI (d)*	73 ± 9	70 ± 11	.09
Mean time between pretreatment MRI and surgery (d)*	158 ± 20	150 ± 32	.06
MRI features			.28
Mass	66 (69)	26 (65)	
Nonmass enhancement	8 (8)	1 (3)	
Mass and nonmass enhancement	22 (23)	13 (32)	
Clinical stage			.13
II	30 (31)	8 (20)	
III	66 (69)	32 (80)	
Tumor histologic type			.53
Ductal	89 (93)	39 (98)	
Lobular	1 (1)	0 (0)	
Other	6 (6)	1 (2)	
ER status			.01
Negative	39 (41)	29 (72)	
Positive	57 (59)	11 (28)	
PR status			<.001
Negative	59 (62)	38 (95)	
Positive	37 (38)	2 (5)	
HER2 status			.04
Negative	63 (66)	19 (48)	
Positive	33 (34)	21 (52)	
Triple-negative cancer			
Yes	25 (26)	14 (35)	.31
No	71 (74)	26 (65)	
Chemotherapy regimen			.001
AT	13 (13)	5 (12)	
AC	5 (5)	0 (0)	
AC-T	61 (64)	15 (38)	
TCHP	17 (18)	20 (50)	
Surgery type			.51
Total mastectomy	49 (51)	21 (52)	
Breast-conserving surgery	47 (49)	19 (48)	

Note.—Unless otherwise noted, values are numbers of patients, with percentages in parentheses. AC = doxorubicin with cyclophosphamide, AC-T = doxorubicin with cyclophosphamide plus docetaxel, AT = doxorubicin with docetaxel, ER = estrogen receptor, HER2 = human epidermal growth factor 2, pCR = pathologic complete response, PR = progesterone receptor, TCHP = docetaxel and carboplatin with trastuzumab and pertuzumab.

* Numbers are means ± standard deviations.

Table 2: Median AUCs of 12 Random Forest Models in the Prediction of Pathologic Complete Response

Image and Feature	AUC	P Value*
T1-weighted MRI		
PRE	0.57 (0.48, 0.65)	<.001
MID	0.82 (0.74, 0.88)	…
MID–PRE	0.69 (0.60, 0.77)	.07
T2-weighted MRI		
PRE	0.53 (0.44, 0.61)	<.001
MID	0.58 (0.49, 0.66)	<.001
MID–PRE	0.67 (0.58, 0.75)	.03
DWI		
PRE	0.57 (0.48, 0.66)	<.001
MID	0.73 (0.65, 0.80)	.19
MID–PRE	0.65 (0.56, 0.73)	.02
ADC mapping		
PRE	0.52 (0.44, 0.61)	<.001
MID	0.69 (0.60, 0.77)	.07
MID–PRE	0.67 (0.58, 0.75)	.04

Note.—Numbers in parentheses are 95% confidence intervals. ADC = apparent diffusion coefficient, AUC = area under the receiver operating characteristic curve, DWI = diffusion-weighted imaging, MID = features at midtreatment MRI, MID–PRE = difference between features at midtreatment and pretreatment MRI, PRE = features at pretreatment MRI.

* P values are for comparison with midtreatment T1-weighted MRI.

was considered to indicate a statistically significant difference. Statistical analyses and comparison of all machine learning classifiers were performed by using open-source R software (version 3.5.1; R Foundation for Statistical Computing, Vienna, Austria) and commercially available Matlab R2018a (MathWorks, Natick, Mass) on Windows 10.

Results

Patient Characteristics

One hundred thirty-six women comprised the final study group (mean age, 47.9 years; range, 31–70 years). Patient characteristics are listed in Table 1. The median maximum tumor diameter at pretreatment MRI was 4.2 cm (range, 1–10.7 cm). Of the 136 patients, 40 (29%) achieved pCR at final pathologic evaluation. The pCR group showed a higher proportion of estrogen receptor negativity (29 of 40 patients [72%]; $P < .01$), progesterone receptor negativity (38 of 40 patients [95%]; $P < .01$), and HER2 positivity (21 of 40 patients [52%]; $P = .04$) than did the non-pCR group. There were no differences in age ($P = .66$), MRI features ($P = .28$), clinical stage ($P = .13$), tumor histologic type ($P = .53$), or surgery type ($P = .51$) between patients with and patients without pCR.

Texture Analysis

For repeated cross-validation (n = 100), the histogram (20 bins) and ranges of 100 AUCs (number of trees, 251) of 12

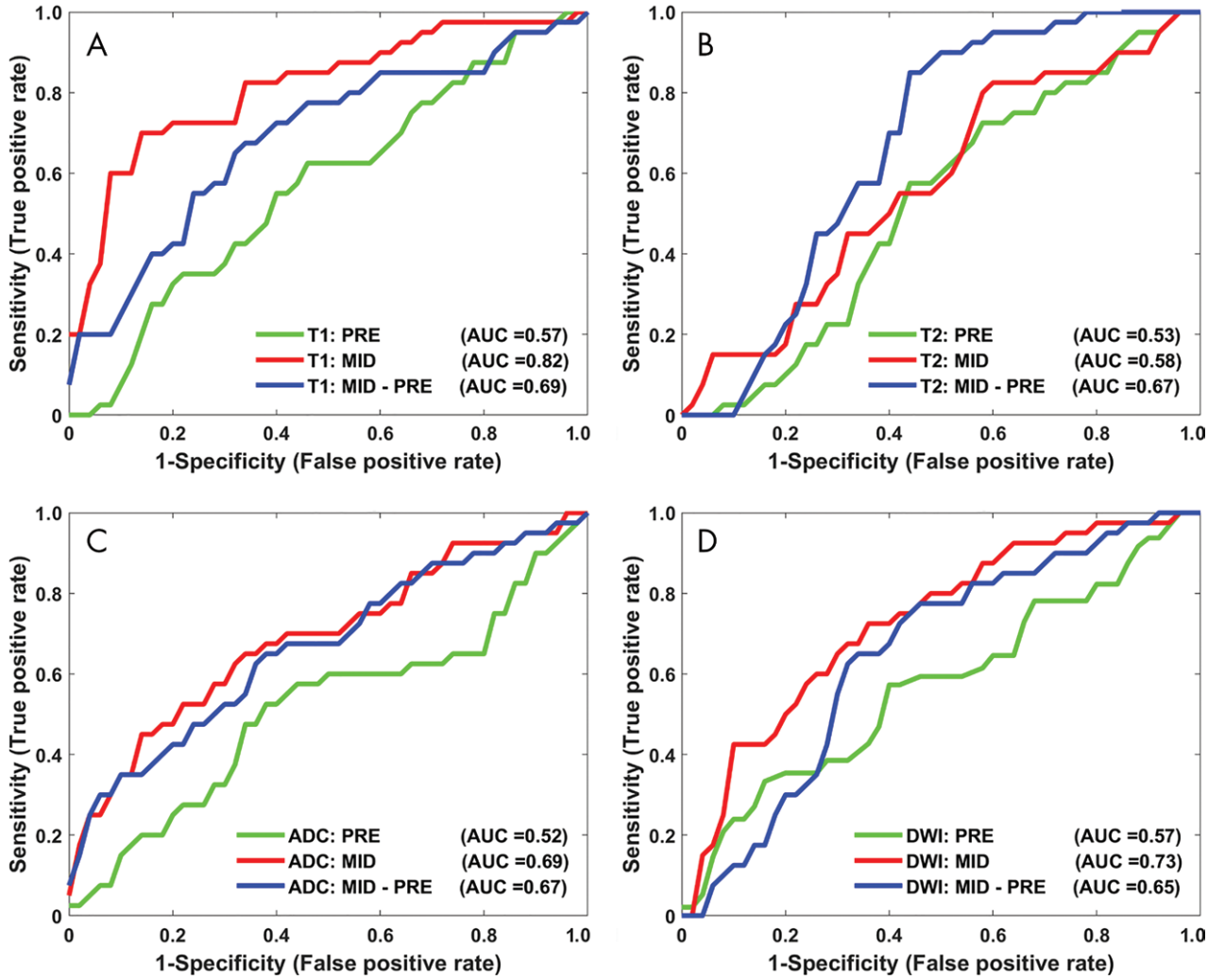

Figure 4: Graphs show area under the receiver operating characteristic curve (AUC) comparison of pretreatment features (PRE), midtreatment features (MID), and difference between midtreatment and pretreatment features (MID – PRE) at, *A*, T1-weighed (T1) MRI, *B*, T2-weighted (T2) MRI, *C*, apparent diffusion coefficient (ADC) mapping, and, *D*, diffusion-weighted imaging (DWI) in prediction of pathologic complete response.

random forest classifiers are depicted in Figure E1 (online). The minimum and maximum AUCs obtained with pretreatment ADC mapping and midtreatment dynamic contrast-enhanced T1-weighted MRI were 0.5 and 0.85, respectively. Midtreatment dynamic contrast-enhanced T1-weighted MRI showed the highest diagnostic performance as well as robustness based on AUC ranges.

Among the 100 AUCs, we selected median AUCs as a representative AUC of the random forest models. The median values are given in Table 2, and the receiver operating characteristic curves based on feature type (pretreatment features, midtreatment features, and difference between midtreatment and pretreatment features) and image type (T1-weighted image, T2-weighted image, ADC map, DWI scan) are shown in Figures 4 and 5, respectively. Among the 12 median AUCs, midtreatment dynamic contrast-enhanced T1-weighted MRI achieved the best diagnostic performance (AUC, 0.82; 95% confidence interval: 0.74, 0.88). Compared with the remaining 11 median AUCs, the median AUC with midtreatment dynamic contrast-enhanced T1-weighted MRI was significantly higher than the others except for difference between midtreatment and pretreatment features at T1-weighted imaging (AUC, 0.69; P = .07), midtreatment DWI (AUC, 0.73; P = .19), and midtreatment ADC (AUC, 0.69; P = .07).

The diagnostic performance for each image and feature type in the prediction of pCR is summarized in Table 3. All diagnostic performance measures for midtreatment T1-weighted MRI were higher than those for pretreatment T2-weighted MRI (sensitivity: 62.5% vs 7.5%, respectively, $P < .001$; PPV: 75.8% vs 15%, $P < .001$; NPV: 85.4% vs 68.1%, P = .01; accuracy, 83.1% vs 60.3%, $P < .001$), midtreatment T2-weighted MRI (sensitivity: 62.5% vs 15%, $P < .001$; PPV: 75.8% vs 25%, P = .01; NPV: 85.4% vs 69.6%, P = .02; accuracy: 83.1% vs 61.8%, $P < .001$), and the difference between midtreatment and pretreatment features at T2-weighted MRI (sensitivity: 62.5% vs 32.5%, P = .01; specificity: 91.7% vs 75%, P = .02; PPV: 75.8% vs 35.1%, P = .02; NPV: 85.4% vs 72.7%, P = .04; accuracy, 83.1% vs 62.5%, $P < .001$) except the specificity with pretreatment T2-weighted MRI (91.7% vs 82.3%, P = .11)

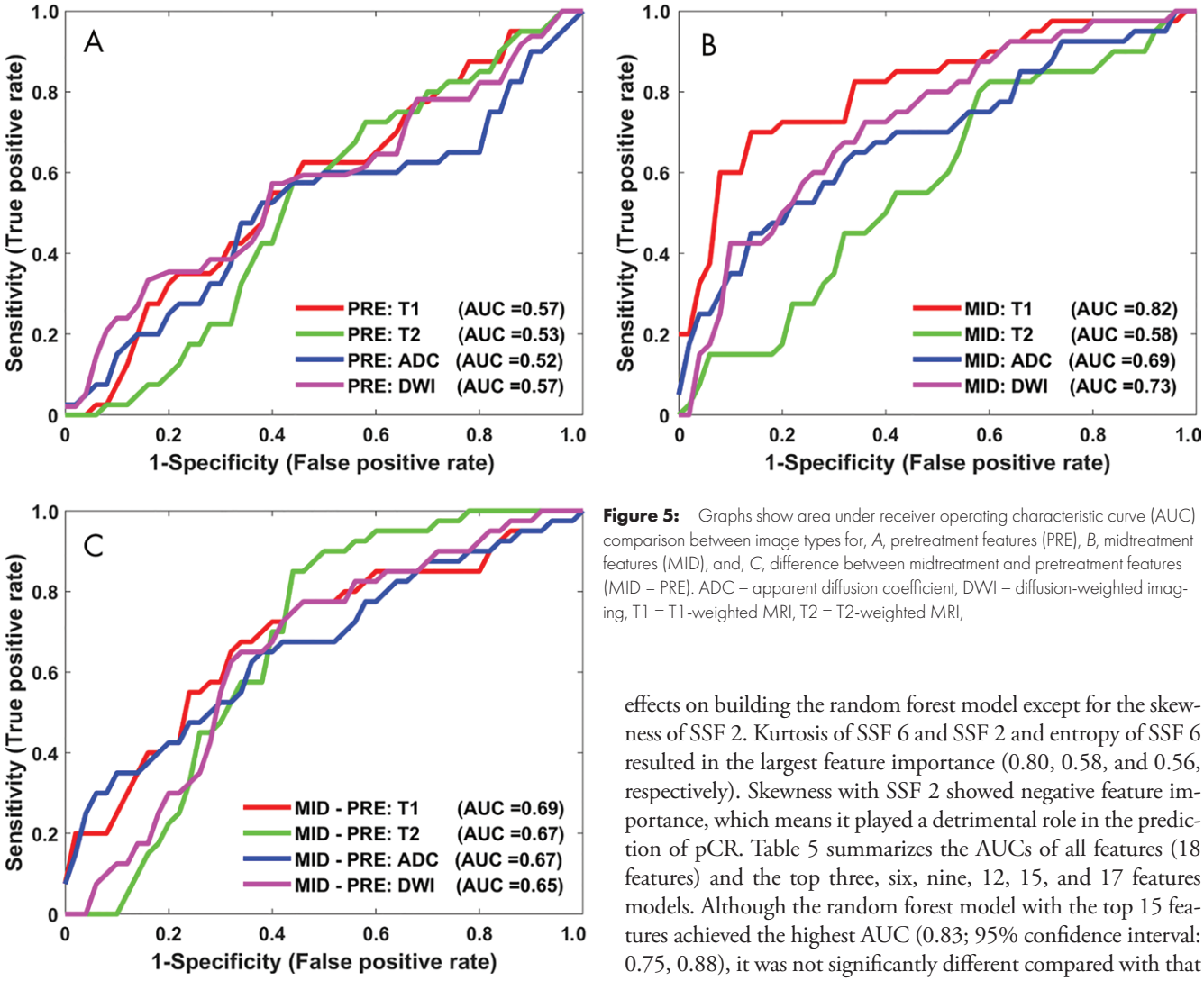

Figure 5: Graphs show area under receiver operating characteristic curve (AUC) comparison between image types for, *A*, pretreatment features (PRE), *B*, midtreatment features (MID), and, *C*, difference between midtreatment and pretreatment features (MID – PRE). ADC = apparent diffusion coefficient, DWI = diffusion-weighted imaging, T1 = T1-weighted MRI, T2 = T2-weighted MRI,

and midtreatment T2-weighted MRI (91.7% vs 81.3%, P = .09). Although the AUCs for midtreatment T1-weighted MRI were not significantly different than those for the difference between midtreatment and pretreatment features at T1-weighted imaging and midtreatment DWI (0.82 vs 0.69 [P = .07] and 0.73, respectively; P = .19), the PPV with midtreatment T1-weighted imaging (75.8%) was higher than that for the difference between midtreatment and pretreatment features at T1-weighted imaging (50%, P = .04). In addition, the accuracy with midtreatment T1-weighted MRI (83.1%) was higher than that with the difference between midtreatment and pretreatment features at T1-weighted imaging (70.6%, P = .02) and midtreatment DWI (72.1%, P = .02). The diagnostic performance of midtreatment T1-weighted imaging was higher than that with midtreatment ADC (AUC: 0.82 vs 0.69, respectively, P = .08; sensitivity: 62.5% vs 45%, P = .12; specificity: 91.7% vs 86.5%, P = .3; PPV: 75.8% vs 58.1%, P = .13; NPV: 85.4% vs 79%, P = .23; accuracy: 83.1% vs 74.3%, P = .08); however, the difference was not significant.

The importance of features at midtreatment T1-weighted imaging is shown in Table 4. Of 18 features, all showed positive effects on building the random forest model except for the skewness of SSF 2. Kurtosis of SSF 6 and SSF 2 and entropy of SSF 6 resulted in the largest feature importance (0.80, 0.58, and 0.56, respectively). Skewness with SSF 2 showed negative feature importance, which means it played a detrimental role in the prediction of pCR. Table 5 summarizes the AUCs of all features (18 features) and the top three, six, nine, 12, 15, and 17 features models. Although the random forest model with the top 15 features achieved the highest AUC (0.83; 95% confidence interval: 0.75, 0.88), it was not significantly different compared with that of the full feature model (P = .36).

Figure 6 is a boxplot showing the diagnostic performance of the random forest model and six other machine learning classifiers in the prediction of pCR with midtreatment T1-weighted imaging. The diagnostic performance with the random forest model was better than that of the other six machine learning classifiers at midtreatment contrast-enhanced T1-weighted MRI (Table 6).

Discussion

Our study showed that the texture parameters at 3-T MRI with random forest modeling have the potential to enable early prediction of pathologic complete response (pCR) after neoadjuvant chemotherapy (NAC). The pCR is associated with better disease-free survival and overall survival in patients with breast cancer undergoing NAC. Therefore, early identification of optimal patients for NAC with MRI may be pivotal. Because texture can provide information for pCR at pre- and midtreatment MRI, it may be helpful for precision medicine in breast cancer. Several studies have investigated the value of texture analysis in MRI for predicting response after NAC in breast cancer; however, they had relatively small study populations and did not analyze various sequences.

Table 3: Diagnostic Performance of Image and Feature Types in the Prediction of Pathologic Complete Response

Type of Image and Feature	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
Contrast-enhanced T1-weighted MRI					
PRE	10 (0.7, 19.3) [<.001]*	88.5 (82.2, 94.9) [.47]	26.7 (4.3, 49) [.01]*	70.2 (62.1, 78.4) [.02]*	65.4 (57.4, 73.4) [.01]*
MID	62.5 (47.5, 77.5)	91.7 (86.1, 97.2)	75.8 (61.1, 90.4)	85.4 (78.6, 92.2)	83.1 (76.8, 89.4)
MID–PRE	40 (24.8, 55.2) [.05]	83.3 (75.9, 90.8) [.11]	50 (32.7, 67.3) [.04]*	76.9 (68.8, 85) [.14]	70.6 (62.9, 78.2) [.02]*
T2-weighted MRI					
PRE	7.5 (0, 15.7) [<.001]*	82.3 (74.7, 89.9) [.11]	15 (0, 30.6) [<.001]*	68.1 (59.6, 76.6) [.01]*	60.3 (52.1, 68.5) [<.001]*
MID	15 (3.9, 26.1) [<.001]*	81.3 (73.4, 89.1) [.09]	25 (7.7, 42.3) [.01]*	69.6 (61.1, 78.2) [.02]*	61.8 (53.6, 69.9) [<.001]*
MID–PRE	32.5 (18, 47) [.01]*	75 (66.3, 83.7) [.02]*	35.1 (19.8, 50.5) [.01]*	72.7 (64, 81.5) [.04]*	62.5 (54.4, 70.6) [<.001]*
DWI					
PRE	5 (0, 11.8) [<.001]*	83.3 (75.9, 90.8) [.11]	11.1 (0, 25.6) [<.001]*	67.8 (59.4, 76.2) [.01]*	60.3 (52.1, 68.5) [<.001]*
MID	45 (29.6, 60.4) [.12]	83.3 (75.9, 90.8) [.11]	52.9 (36.2, 69.7) [.06]	78.4 (70.4, 86.4) [.21]	72.1 (64.5, 79.6) [.03]*
MID–PRE	36.4 (20, 52.8) [.01]*	72.8 (64.2, 81.4) [.05]*	30 (15.8, 44.2) [.01]*	78.1 (69.9, 86.4) [.04]*	64 (55.9, 72) [<.001]*
ADC mapping					
PRE	22.5 (9.6, 35.4) [<.001]*	81.2 (73.4, 89.1) [.10]	33.3 (15.6, 51.1) [.01]*	71.6 (63.1, 80) [.03]*	64 (55.9, 72) [.01]*
MID	45 (29.6, 60.4) [.12]	86.5 (79.6, 93.3) [.30]	58.1 (40.7, 75.4) [.13]	79 (71.3, 86.8) [.23]	74.3 (66.9, 81.6) [.08]
MID–PRE	35 (20.2, 49.8) [.02]*	87.5 (80.9, 94.1) [.38]	53.8 (34.7, 73) [.09]	76.4 (68.4, 84.3) [.13]	72.1 (64.5, 79.6) [.03]*

Note.—Data are percentages, with 95% confidence intervals in parentheses and *P* values in brackets. ADC = apparent diffusion coefficient, DWI = diffusion-weighted imaging, MID = features at midtreatment MRI, MID–PRE = difference between features at midtreatment and pretreatment MRI, NPV = negative predictive value, PPV = positive predictive value, PRE = features at pretreatment MRI.

* Adjusted $P < .05$, compared with contrast-enhanced midtreatment T1-weighted MRI.

In our study, texture features at midtreatment contrast-enhanced T1-weighted MRI with random forest modeling had better diagnostic performance for showing an association with pCR, with the highest AUC compared with those at T2-weighted MRI (0.82 vs 0.58; $P < .001$). Our analysis also showed that texture parameters obtained with midtreatment contrast-enhanced T1-weighted imaging (AUC, 0.82) showed better diagnostic performance than those with DWI (AUC, 0.73; $P = .19$) and ADC mapping (AUC, 0.69; $P = .07$), albeit the difference was not significant. Among the texture parameters consisting of the random forest models of midtreatment contrast-enhanced T1-weighted imaging, kurtosis for SSF 6 showed the highest feature importance, followed by kurtosis for SSF 2 and entropy for SSF 6. In addition, we compared the performance of the random forest classifier to that of six other machine learning algorithms by using features from midtreatment contrast-enhanced T1-weighted MRI and demonstrated that the random forest model has better diagnostic performance for showing association with pCR.

Previous studies with the same commercial software showed AUCs of 0.74–0.84 for predicting pCR in breast cancer (15,16), which are similar to those in our study. However, those studies used limited MRI sequences. Several previous studies using quantitative features extracted from dynamic contrast-enhanced MRI combining multiparametric data have shown AUCs of 0.68–0.91, which are similar to those in our study (18,20,22,24). This highlights that our data were based only on basic texture features. In contrast to our study, a previous study showed that the diagnostic performance of texture parameter changes on a T2-weighted image using entropy and uniformity was better than that on a contrast-enhanced T1-weighted image (15). Chamming's et al (19) reported that kurtosis on T2-weighted images was independently associated with pCR in non–triple-negative breast cancer. Although their studies used the same software application as ours, these conflicting results may be explained in part by the different magnetic field strengths used (1.5 T vs 3 T) or by differences in the MRI protocol. A previous study found that signal-to-noise ratio, field strength, and intensity normalization algorithms have an effect on texture analysis (30). In addition, other studies have suggested that spatial resolution is an important factor when performing texture analysis and that the sequence with higher spatial resolution resulted in better classification compared with the sequence with lower spatial resolution (31).

Table 4: Importance of Features at Midtreatment Contrast-enhanced T1-weighted Imaging in the Prediction of Pathologic Complete Response

Feature and Texture Parameter	Importance	Cumulative Importance Ratio (%)*
Kurtosis, SSF 6	0.80	14.89
Kurtosis, SSF 2	0.58	25.65
Entropy, SSF 6	0.56	36.14
Entropy, SSF 4	0.47	44.93
Standard deviation, SSF 4	0.32	50.89
MPP, SSF 6	0.31	56.73
MPP, SSF 4	0.31	62.45
Standard deviation, SSF 2	0.28	67.60
Skewness, SSF 6	0.26	72.55
Mean pixel intensity, SSF 4	0.26	77.36
Entropy, SSF 2	0.22	81.55
Mean pixel intensity, SSF 6	0.20	85.37
MPP, SSF 2	0.19	88.90
Mean pixel intensity, SSF 2	0.17	92.01
Standard deviation, SSF 6	0.16	95.06
Skewness, SSF 4	0.14	97.65
Kurtosis, SSF 4	0.13	100
Skewness, SSF 2	−0.08	…

Note.—MPP = mean proportion of positive pixels, SSF = spatial scaling factor, SSF 2 = highlighting fine-texture features, SSF 4 = highlighting medium-texture features, SSF 6 = highlighting coarse-texture features.

* Cumulative importance ratio except negative importance.

Table 5: AUCs according to Number of Texture Features at Midtreatment Contrast-enhanced T1-weighted MRI in the Prediction of Pathologic Complete Response

Feature	AUC	Cumulative Importance Ratio (%)*	*P* Value	Adjusted *P* Value
All (18 features)	0.82 (0.74, 0.88)	…	…	
Top three features	0.80 (0.72, 0.86)	36.14	.69	.93
Top six features	0.81 (0.74, 0.87)	56.73	.93	.93
Top nine features	0.82 (0.74, 0.88)	72.55	.73	.93
Top 12 features	0.82 (0.75, 0.88)	85.37	.65	.93
Top 15 features	0.83 (0.75, 0.88)	95.06	.36	.93
Top 17 features	0.82 (0.74, 0.88)	100	.89	.93

Note.—Numbers in parentheses are the 95% confidence interval. AUC = area under the receiver operating characteristic curve.

* Cumulative importance ratio except negative importance.

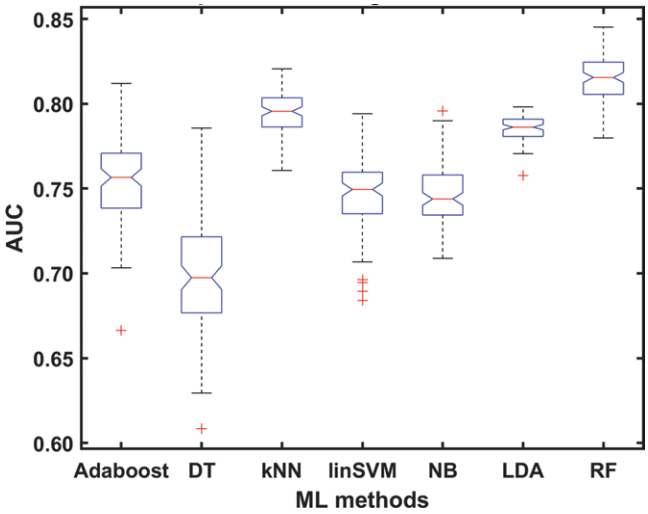

Figure 6: Box and whisker plot illustrates performance of various machine learning (ML) algorithms in the prediction of pathologic complete response. Adaboost = adaptive boosting, AUC = area under the receiver operating characteristic curve, DT = decision tree, kNN = k-nearest neighbor, LDA = linear discriminant analysis, linSVM = linear support vector machine, NB = naive Bayes, RF = random forest.

Table 6: AUCs according to Various Machine Learning Classifiers of Midtreatment Contrast-enhanced T1-weighted MRI in the Prediction of Pathologic Complete Response

Classifier	AUC
Adaptive boosting	0.76 (0.68, 0.83)
Decision tree	0.70 (0.61, 0.77)
kNN	0.80 (0.72, 0.86)
LinSVM	0.75 (0.67, 0.82)
Naive Bayes	0.74 (0.65, 0.84)
LDA	0.79 (0.87, 0.70)
Random forest	0.82 (0.74, 0.88)

Note.—Numbers in parentheses are the 95% confidence interval. AUC = area under the receiver operating characteristic curve, kNN = k-nearest neighbor, LDA = linear discriminant analysis, LinSVM = linear support vector machine.

To our knowledge, this is the first study to evaluate the diagnostic performances of random forest models with use of texture parameters and their changes at different sequences and timing, including T2-weighted images, contrast-enhanced T1-weighted images, DWI scans, and ADC maps at 3-T MRI before and in the middle of treatment. Yoon et al (32) investigated DWI for predicting response to NAC and found different ADCs between the responding and the nonresponding group. However, unlike in our study, there were no data for diagnostic performance and no results for comparison between other sequences. Several previous studies have also investigated whether texture parameter changes during NAC might help predict pCR in breast cancer (15,16,19,23); however, their data were solely based on one or two MRI sequences.

Our results are in accordance with those from a recent study showing that kurtosis at pretreatment MRI was associated with pCR after NAC in breast cancer (19). Kurtosis reflects the peakedness of the histogram analysis of the pixel signal intensity, which is considered to provide information about tissue microstructure organization (33,34). Horvat et al (27) investigated the association between texture parameters and pathologic response after NAC in rectal cancer; they also found that kurtosis was associated with pCR. We then evaluated the feature importance for

accurate diagnosis using a random forest model. When analyzed with several numbers of top features, the AUC of all features showed no significant difference with those of three, six, nine, 12, 15, and 17 features. This result indicates that the three top texture parameters showing the highest feature importance, such as kurtosis for SSF 6 and SSF 2 and entropy for SSF 6, might give us enough information to predict pCR. Our study differed from other studies in that the results were analyzed from random forest models that measure feature importance and build decision trees to reduce overfitting and diagnostic error.

In addition, we compared the performance of a random forest classifier to that of six other machine learning algorithms by using features from midtreatment contrast-enhanced T1-weighted MRI. We demonstrated that the diagnostic performance of the random forest model for showing an association with pCR was better than that of the machine learning algorithms (AUC: adaptive boosting, 0.76; decision tree, 0.70; k-nearest neighbor, 0.80; linear support vector machine, 0.75; naive Bayes, 0.74; linear discriminant analysis, 0.79). Tahmassebi et al (35) investigated machine learning in multiparametric MRI using qualitative and quantitative features from all MRI sequences for early prediction of pCR. Although they did not assess the texture features, machine learning based on an extreme gradient boosting classifier model enabled prediction of pCR, with a median AUC of 0.86.

Our study has several limitations. First, this was a retrospective study in a single institution. This may have caused selection bias. Second, the heterogeneous nature of the histopathologic cancer subtype included in our study led to the use of different chemotherapy regimens, which reflects clinical reality. However, this may affect pathologic response to NAC and cause selection bias. Third, the software we used could analyze only two-dimensional rather than three-dimensional MRI scans. Therefore, it may not have represented the texture features of the entire tumor. Finally, the region of interest was drawn by two radiologists in consensus.

In conclusion, texture features of contrast-enhanced T1-weighted images at midtreatment MRI showed the most valuable results for association with pathologic complete response to neoadjuvant chemotherapy (NAC) in patients with breast cancer among texture parameters from our mathematic modeling analysis. However, further investigation in a larger and independent external validation data set is warranted before this approach can be used for actual clinical decision making. After validation, this analysis may become an important monitoring tool for predicting response to NAC for breast cancer.

Author contributions: Guarantor of integrity of entire study, H.M.G.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, N.L.E., E.J.S., J.S.P., J.A.K., H.M.G.; clinical studies, N.L.E., E.J.S., J.S.P., J.H.Y., J.A.K., H.M.G.; experimental studies, D.S.K., H.M.G.; statistical analysis, N.L.E., D.S.K.; and manuscript editing, D.S.K., E.J.S., H.M.G.

Disclosures of Conflicts of Interest: **N.L.E.** disclosed no relevant relationships. **D.K.** disclosed no relevant relationships. **E.J.S.** disclosed no relevant relationships. **J.S.P.** disclosed no relevant relationships. **J.H.Y.** disclosed no relevant relationships. **J.A.K.** disclosed no relevant relationships. **H.M.G.** disclosed no relevant relationships.

References

1. Kaufmann M, von Minckwitz G, Bear HD, et al. Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006. Ann Oncol 2007;18(12):1927–1934.
2. Wolmark N, Wang J, Mamounas E, Bryant J, Fisher B. Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18. J Natl Cancer Inst Monogr 2001;(30):96–102.
3. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet 2014;384(9938):164–172.
4. Rauch GM, Kuerer HM, Adrada B, et al. Biopsy feasibility trial for breast cancer pathologic complete response detection after neoadjuvant chemotherapy: imaging assessment and correlation endpoints. Ann Surg Oncol 2018;25(7):1953–1960.
5. Uematsu T, Kasami M, Yuen S. Neoadjuvant chemotherapy for breast cancer: correlation between the baseline MR imaging findings and responses to therapy. Eur Radiol 2010;20(10):2315–2322.
6. Hylton NM, Blume JD, Bernreuter WK, et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy—results from ACRIN 6657/I-SPY TRIAL. Radiology 2012;263(3):663–672.
7. Loo CE, Straver ME, Rodenhuis S, et al. Magnetic resonance imaging response monitoring of breast cancer during neoadjuvant chemotherapy: relevance of breast cancer subtype. J Clin Oncol 2011;29(6):660–666.
8. Cho N, Im SA, Park IA, et al. Breast cancer: early prediction of response to neoadjuvant chemotherapy using parametric response maps for MR imaging. Radiology 2014;272(2):385–396.
9. Li X, Abramson RG, Arlinghaus LR, et al. Multiparametric magnetic resonance imaging for predicting pathological response after the first cycle of neoadjuvant chemotherapy in breast cancer. Invest Radiol 2015;50(4):195–204.
10. Fangberget A, Nilsen LB, Hole KH, et al. Neoadjuvant chemotherapy in breast cancer-response evaluation and prediction of response to treatment using dynamic contrast-enhanced and diffusion-weighted MR imaging. Eur Radiol 2011;21(6):1188–1199.
11. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. Clin Radiol 2004;59(12):1061–1069.
12. Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging 2012;3(6):573–589.
13. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. Radiology 2013;269(1):8–15.
14. Chitalia RD, Kontos D. Role of texture analysis in breast MRI as a cancer biomarker: A review. J Magn Reson Imaging 2019;49(4):927–938.
15. Parikh J, Selmi M, Charles-Edwards G, et al. Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. Radiology 2014;272(1):100–112.
16. Henderson S, Purdie C, Michie C, et al. Interim heterogeneity changes measured using entropy texture features on T2-weighted MRI at 3.0 T are associated with pathological response to neoadjuvant chemotherapy in primary breast cancer. Eur Radiol 2017;27(11):4602–4611.
17. Ahmed A, Gibbs P, Pickles M, Turnbull L. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. J Magn Reson Imaging 2013;38(1):89–101.
18. Teruel JR, Heldahl MG, Goa PE, et al. Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. NMR Biomed 2014;27(8):887–896.
19. Chamming's F, Ueno Y, Ferré R, et al. Features from computerized texture analysis of breast cancers at pretreatment MR imaging are associated with response to neoadjuvant chemotherapy. Radiology 2018;286(2):412–420.
20. Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L. Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. Eur J Radiol 2017;94:140–147.
21. Thibault G, Tudorica A, Afzal A, et al. DCE-MRI texture features for early prediction of breast cancer therapy response. Tomography 2017;3(1):23–32.
22. Machireddy A, Thibault G, Tudorica A, et al. Early prediction of breast cancer therapy response using multiresolution fractal analysis of DCE-MRI parametric maps. Tomography 2019;5(1):90–98.
23. Wu J, Gong G, Cui Y, Li R. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. J Magn Reson Imaging 2016;44(5):1107–1115.
24. Braman NM, Etesami M, Prasanna 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(1):57 [Published correction appears in Breast Cancer Res 2017;19(1):80.].
25. Ueno Y, Forghani B, Forghani R, et al. Endometrial carcinoma: MR imaging-based texture model for preoperative risk stratification-a preliminary analysis. Radiology 2017;284(3):748–757.
26. Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
27. Horvat N, Veeraraghavan H, Khan M, et al. MR imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. Radiology 2018;287(3):833–843.
28. Mitchell MW. Bias of the random forest out-of-bag (OOB) error for certain input parameters. Open J Stat 2011;1(3):205–211.

29. Chen W, Samuelson FW, Gallas BD, Kang L, Sahiner B, Petrick N. On the assessment of the added value of new predictive biomarkers. BMC Med Res Methodol 2013;13(1):98.
30. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging 2004;22(1):81–91.
31. Waugh SA, Lerski RA, Bidaut L, Thompson AM. The influence of field strength and different clinical breast MRI protocols on the outcome of texture analysis using foam phantoms. Med Phys 2011;38(9):5058–5066.
32. Yoon HJ, Kim Y, Chung J, Kim BS. Predicting neo-adjuvant chemotherapy response and progression-free survival of locally advanced breast cancer using textural features of intratumoral heterogeneity on F-18 FDG PET/CT and diffusion-weighted MR imaging. Breast J 2019;25(3):373–380.
33. Just N. Improving tumour heterogeneity MRI assessment with histograms. Br J Cancer 2014;111(12):2205–2213.
34. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magn Reson Med 2005;53(6):1432–1440.
35. 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 chemotherapy and survival outcomes in breast cancer patients. Invest Radiol 2019;54(2):110–117.