Title

MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer.

Authors

Bitencourt, Almir

Gibbs, Peter

Rossi Saccarelli, Carolina

et al.

Publication Date

2020-11-01

DOI

10.1016/j.ebiom.2020.103042

EBioMedicine 61 (2020) 103042

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer

Almir G.V. Bitencourt[a,b], Peter Gibbs[a,c], Carolina Rossi Saccarelli[a,d], Isaac Daimiel[a], Roberto Lo Gullo[a], Michael J. Fox[c], Sunitha Thakur[a,c], Katja Pinker[a,e,*], Elizabeth A. Morris[a], Monica Morrow[f], Maxine S. Jochelson[a]

[a] Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[b] Department of Imaging, A.C. Camargo Cancer Center, Sao Paulo, SP, Brazil
[c] Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[d] Department of Radiology, Hospital Sírio-Libanês, São Paulo, SP, Brazil
[e] Department of Biomedical Imaging and Image-guided Therapy, Division of Molecular and Gender Imaging, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria
[f] Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

ARTICLE INFO

Article History:
Received 21 April 2020
Revised 4 September 2020
Accepted 21 September 2020
Available online xxx

Keywords:
Magnetic resonance imaging
Breast invasive ductal carcinoma
HER2
ErbB-2 receptor
Neoadjuvant therapy
Machine learning

ABSTRACT

Background: To use clinical and MRI radiomic features coupled with machine learning to assess HER2 expression level and predict pathologic response (pCR) in HER2 overexpressing breast cancer patients receiving neoadjuvant chemotherapy (NAC).
Methods: This retrospective study included 311 patients. pCR was defined as no residual invasive carcinoma in the breast or axillary lymph nodes (ypT0/isN0). Radiomics/statistical analysis was performed using MATLAB and CERR software. After ROC and correlation analysis, selected radiomics parameters were advanced to machine learning modelling alongside clinical MRI-based parameters (lesion type, multifocality, size, nodal status). For predicting pCR, the data was split into a training and test set (80:20).
Findings: The overall pCR rate was 60.5% (188/311). The final model to predict HER2 heterogeneity utilised three MRI parameters (two clinical, one radiomic) for a sensitivity of 99.3% (277/279), specificity of 81.3% (26/32), and diagnostic accuracy of 97.4% (303/311). The final model to predict pCR included six MRI parameters (two clinical, four radiomic) for a sensitivity of 86.5% (32/37), specificity of 80.0% (20/25), and diagnostic accuracy of 83.9% (52/62) (test set); these results were independent of age and ER status, and outperformed the best model developed using clinical parameters only (p=0.029, comparison of proportion Chi-squared test).
Interpretation: The machine learning models, including both clinical and radiomics MRI features, can be used to assess HER2 expression level and can predict pCR after NAC in HER2 overexpressing breast cancer patients.
Funding: NIH/NCI (P30CA008748), Susan G. Komen Foundation, Breast Cancer Research Foundation, Spanish Foundation Alfonso Martin Escudero, European School of Radiology.

1. Introduction

The human epidermal growth factor receptor 2 (HER2) gene is located on human chromosome 17q21 and is a member of ErbB family of receptor tyrosine kinases. Overexpression of the HER2 gene and its protein is observed in 20−30% of breast cancer cases, is a prognostic factor, and informs treatment selection [1]. Protein overexpression detected by immunohistochemistry (IHC) or amplification of the HER2 gene analysed by fluorescence in situ hybridisation (FISH) are the two main methods used to determine HER2 status in clinical practice. The American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines provide recommendations to interpret and assign HER2 status [2].

However, tumours may contain multiple clones with distinct HER2 amplification characteristics within the tumour, resulting in HER2 intratumour heterogeneity. Recently, studies have suggested

* Corresponding author at: Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, USA
E-mail address: pinkerdk@mskcc.org (K. Pinker).

https://doi.org/10.1016/j.ebiom.2020.103042

Research in Context

Evidence before this study

Overexpression of the HER2 gene and its protein is an important prognostic factor in breast cancer patients and affects treatment strategy. Prior studies have suggested that intratumour heterogeneity based on HER2 expression levels can affect response to neoadjuvant chemotherapy (NAC). Breast magnetic resonance imaging (MRI) is the most accurate imaging method to evaluate the breast cancer at presentation and to predict pathologic response in HER2-positive breast cancer after NAC.

Added value of this study

Our results showed that MRI features can be used to predict HER2 expression and pathological response after NAC in HER2 overexpressing breast cancer patients. The best accuracy was achieved when the machine learning models included radiomics parameters alongside clinical MRI-based parameters.

Implications of all the available evidence

Machine learning models, including both clinical and radiomics-based MRI features, may be used to better select patients who could benefit from anti-HER2 treatment after validation in larger multi-institutional studies.

that intratumour heterogeneity based on HER2 expression levels can affect response to neoadjuvant chemotherapy (NAC) [3–7].

The assessment of HER2 status is usually based on the analysis from a single biopsy of a potentially heterogenous tumour, which can capture only a snapshot of the tumour tissue and is subject to selection bias. Therefore, the development of non-invasive biomarkers to evaluate whole-tumour heterogeneity are needed. Breast magnetic resonance imaging (MRI) is the most accurate imaging method to evaluate the breast cancer at presentation and to predict pathologic response in HER2-positive breast cancer after NAC [8]. Radiomic analysis of pretreatment breast MRI has been shown to enable pCR prediction [9]. Imaging intratumour heterogeneity with MRI delivers additional predictive value beyond current clinical-pathologic and biologic predictors in breast cancer, and could potentially be used to stratify patients for individualised therapy [10].

The aim of this study was to use MRI-based clinical and radiomic features coupled with machine learning to assess HER2 expression level in HER2 overexpressing breast cancer patients receiving NAC, and to correlate these findings with pathologic response.

2. Methods

This retrospective single-centre Health Insurance Portability and Accountability Act (HIPAA)-compliant study was approved by the Institutional Review Board and the need for written informed consent was waived. For this study, the datasets generated and/or analysed are available from the corresponding author on reasonable request. Inclusion criteria were: HER2 overexpressing breast cancer patients who underwent NAC and pretreatment state-of-the-art contrast-enhanced breast MRI [11,12]. Of 445 consecutive patients with HER2 overexpressing invasive breast carcinoma treated with NAC and surgical resection at our institution between July 2014 and May 2019, 70 were excluded because they did not have pretreatment breast MRI and 64 patients with outside images were excluded because of poor image quality. Of 311 patients included in the final analysis, 168 (54%) had MRI studies performed at our institution and 143 (46%) had studies performed elsewhere (Fig. 1). The neoadjuvant treatment regimens included AC-THP (Adriamycin and Cyclophosphamide followed by Paclitaxel, Trastuzumab, and Pertuzumab) in 270 patients (86.8%), TCHP (Docetaxel, Carboplatin, Trastuzumab, and Pertuzumab) in 28 (9.0%) patients, THP (Docetaxel, Trastuzumab and Pertuzumab) in seven (2.3%) patients, ACT (Adriamycin, Cyclophosphamide and Paclitaxel) in five (1.6%) patients, and AC (Adriamycin and Cyclophosphamide) in 1 (0.3%) patient.

Oestrogen receptor (ER) and HER2 expression were evaluated on percutaneous biopsy according to the 2018 ASCO/CAP guidelines [2,13]. All pathologic results from outside biopsies were reviewed at our institution. ER expression above 1% on IHC was considered positive [13]. Patients were classified into two groups based on HER2 expression level. The first group (IHC group) comprised patients with tumours that showed HER2 protein overexpression on IHC (IHC 3+). The second group (FISH group) comprised patients with tumours that showed HER2 gene amplification detected by FISH in the absence of protein overexpression on IHC (IHC 2+ or 1+ to 2+). Pathologic complete response (pCR) was defined as no residual invasive carcinoma in the breast or axillary lymph nodes (ypT0/isN0) at surgical resection, similar to prior studies [3–5,7]. Breast MRIs were assessed according to the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) lexicon [12]. Clinical MRI features included tumour size, lesion type (mass, non-mass enhancement [NME] or both), presence of multifocality, and suspicious axillary lymph nodes.

A dedicated breast radiologist (AB) with seven years of experience reviewed the MRI exams and performed 3D segmentations of the whole tumour in the first post-contrast non-subtracted sequence using ITK-SNAP software. All included MR images used fat suppression. Susceptibility artefacts related to post-biopsy changes, when present, were excluded from segmentation and only the largest lesion was segmented in multifocal tumours (Fig. 2). Enhancement maps were calculated as the percentage increase in signal from the pre-contrast image to the first post-contrast image. Radiomics and statistical analysis were performed using MATLAB and publicly available CERR (Computational Environment for Radiological Research) software [14]. CERR has recently been shown to conform to the Image Biomarker Standardisation Initiative (IBSI) guidelines [15]. For analysis, data was reduced to a fixed bin number of 16 grey levels and only an interpixel distance of one was considered. CERR analysis resulted in 102 texture parameters sub-divided into six categories – 22 first order statistics, 26 statistics based on grey level cooccurrence matrices, 16 statistics based on run length matrices, 16 statistics based on size zone matrices, 17 statistics based on neighborhood grey level dependence matrices, and finally five statistics based on neighborhood grey tone difference matrices. Features were computed from

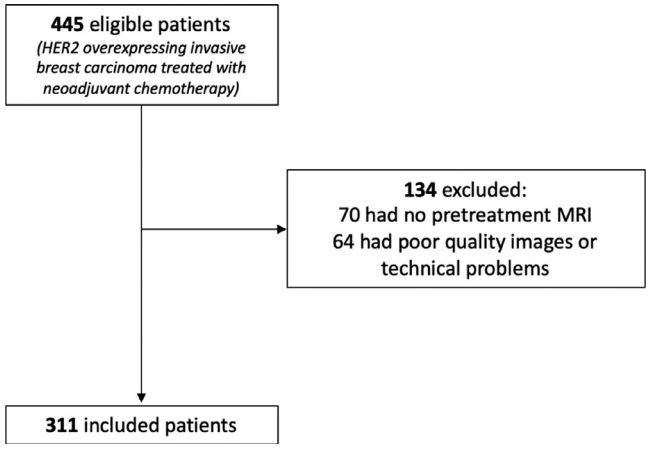

Fig. 1. Flowchart of inclusion and exclusion criteria for the study.

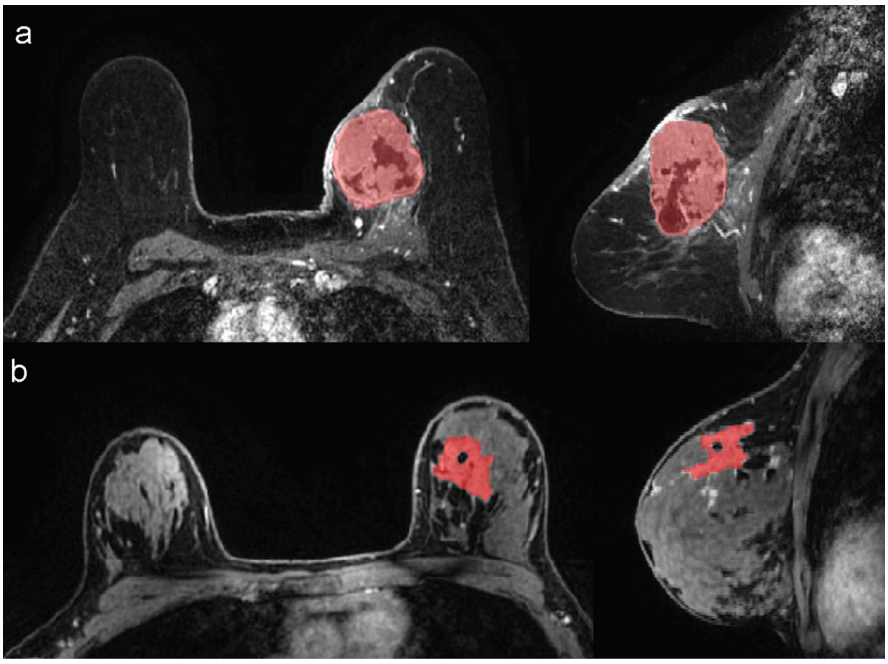

Fig. 2. Examples of 3D segmentations of HER2 overexpressing breast carcinomas in the breast MRI's first minute post-contrast sequence. a) Unifocal tumour (whole tumour was segmented). b) Multifocal tumour (only the largest lesion was segmented).

each 2D directional matrix and averaged over 2D directions and slices, since data was not isotropic. As patients were scanned at different sites, Combat harmonisation was performed to remove the centre effect (in our case local vs. foreign scans) while retaining the pathophysiologic information (in our case either HER2 expression or pathologic response). The harmonisation employed Bayes estimates to account for both additive and multiplicative scanner effects [16].

Initially, univariate analysis was performed to identify significant parameters. Continuous variables were described as mean, standard deviation (SD), and range. The two-tailed Mann–Whitney U test for two independent samples was used to determine significant differences between groups. Correlation analysis was then employed to remove redundant parameters from advancement to model development. If a highly positive (> 0.9) or highly negative (< -0.9) correlation was noted, the parameter with the lowest area under the receiver operating curve (AUROC) was removed. After ROC and correlation analysis, radiomics parameters that demonstrated significant differences between groups were retained and advanced to modelling alongside clinical MRI-based parameters, including tumour size, lesion type, multifocality, and nodal status. For the prediction of pathological complete response, the data was split into training and test sets at a ratio of 4:1 (80% training and 20% test), with feature selection performed purely on the training set. Due to the low number of cases in the minority class, this was not feasible for the comparison between the IHC and FISH groups. Machine learning models were developed utilising coarse decision trees and five-fold cross-validation. Coarse decision tree modelling was implemented in MATLAB, with the maximum number of splits set at four and utilising Gini's diversity index as the splitting criterion. For the prediction of pCR, model development was performed on the training set and then validated on the test set.

3. Results

The mean patient age was 49.4 years (SD: 11.6 years; range: 24–78 years) and the mean tumour size by MRI was 4.7 cm (SD: 2.6 cm; range: 0.9–14.8 cm). The index tumour presented as a mass in 147/311 cases (47.3%), non-mass enhancement (NME) in 34/311 (10.9%), and both mass and NME in 130/311 (41.8%). Most tumours were multifocal ($n = 205/311$; 65.9%) and had suspicious axillary lymph nodes on MRI (n = 233/311; 74.9%). ER was positive in 188/311 tumours (60.5%). There were 279/311 tumours (89.7%) with 3+ IHC expression and 32/311 (10.3%) with amplification by FISH. The overall pCR rate was 60.5% (188/311). Tables 1 and 2 summarise the clinical MRI features according to HER2 expression level and pCR, respectively.

Prior to data harmonisation, 77 radiomic features were significantly different between the local and foreign scans, and after harmonisation, only seven features were still significantly different.

Tables 3 and 4 describe the significant radiomics features related to HER2 expression level and pathologic response on initial univariate analysis, respectively. After observing the correlations between radiomics parameters, two radiomics parameters were advanced for both HER2 expression level and four for pathologic response analysis models. Six clinical parameters, including four MRI features (size, multifocality, lesion type, and nodal status), ER status, and age were

Table 1
Clinical and MRI features of the included patients according to the HER2 expression ($n = 311$).

Clinical/MRI feature	HER2 expression		*p*-value
	IHC	FISH	
Age	49.7 ± 11.6 years	47.3 ± 11.9 years	0.302[a]
Estrogen receptor			
- Negative	112 (91.1%)	11 (8.9%)	0.572[b]
- Positive	167 (88.8%)	21 (11.2%)	
Lesion type			
- Mass	113 (86.9%)	17 (13.1%)	0.204[b]
- NME	133 (90.5%)	14 (9.5%)	
- Both	33 (97.1%)	1 (2.9%)	
Tumour size	4.7 ± 2.5 cm	4.4 ± 2.8 cm	0.312[a]
Multifocality			
- Absent	94 (88.7%)	12 (11.3%)	0.696[b]
- Present	185 (90.2%)	20 (9.8%)	
Nodal status			
- Negative	71 (91.0%)	7 (9.0%)	0.830[b]
- Positive	208 (89.3%)	25 (10.7%)	

[a] Mann–Whitney U-test.
[b] Chi-squared test.

Table 2
Clinical and MRI features of the included patients according to the pathologic response (n = 311).

Clinical/MRI feature	Pathologic response		*p*-value
	No pCR	pCR	
Age	48.7 ± 12.6 years	50.0 ± 10.9 years	0.183[a]
Estrogen receptor			
- Negative	28 (22.8%)	95 (77.2%)	< 0.001[b]
- Positive	95 (50.5%)	93 (49.5%)	
HER2 expression			
- IHC	98 (35.1%)	181 (64.9%)	< 0.001[b]
- FISH	25 (78.1%)	7 (21.9%)	
Lesion type			
- Mass	44 (33.8%)	86 (66.2%)	0.119[b]
- NME	67 (45.6%)	80 (54.4%)	
- Both	12 (35.3%)	22 (64.7%)	
Tumour size	4.3 ± 2.4 cm	4.9 ± 2.6 cm	0.037[a]
Multifocality			
- Absent	46 (43.4%)	60 (56.6%)	0.330[b]
- Present	77 (37.6%)	128 (62.4%)	
Nodal status			
- Negative	37 (47.4%)	41 (52.6%)	0.110[b]
- Positive	86 (36.9%)	147 (63.1%)	

[a] Mann–Whitney U-test.
[b] Chi-squared test.

also advanced to modelling. The processing of selecting radiomic features for final model development is illustrated using correlation and area under the curve analysis for pCR in Table 5.

The final model to predict HER2 intratumour expression levels (IHC vs. FISH) utilised three MRI parameters (two clinical parameters [lesion type and multifocality] and one radiomic parameter [large zone emphasis]) for a sensitivity of 99.3% (277/279), specificity of 81.3% (26/32), positive predictive value of 97.9% (277/283), negative predictive value of 92.9% (26/28), and diagnostic accuracy of 97.4% (303/311) – Table 6.

Random splitting of the data resulted in 249 cases (150 pCR, 99 no pCR) in the training set and 62 cases (38 pCR, 24 no pCR) in the test set. The model to predict pCR included six MRI parameters (two clinical parameters [lesion type and size] and four radiomic parameters [variance, first order entropy, 90th percentile and zone length variance]) for a sensitivity of 87.4% (95% CI: 81.1–92.3%), specificity of 81.6% (95% CI: 72.5–88.7%), positive predictive value of 88.0% (95% CI: 82.8–91.8%), negative predictive value of 80.8% (95% CI: 73.2–86.6%), and diagnostic accuracy of 85.1% (95% CI: 80.1–89.3%) in the training set – Table 7. For the test set, a sensitivity of 86.5% (95% CI: 71.2–95.5%), specificity of 80.0% (95% CI: 59.3–93.2%), positive predictive value of 86.5% (95% CI: 74.3–93.4%), negative predictive value of 80.0% (95% CI: 63.4–90.2%), and diagnostic accuracy of 83.9% (95% CI: 72.3–92.0%) were obtained – Table 8. Age and ER status were not retained in the final models. For comparison, utilising the six clinical features only and exhaustively testing all 63 possible combinations (2^6-1) the best results obtained were a diagnostic accuracy of 65.9% (95% CI: 59.6–71.7%) for the training set and a diagnostic accuracy of 72.6% (95% CI: 59.8–83.2%) for the test set. Comparison of diagnostic accuracy with the radiomics model developed using both clinical and radiomic features revealed a significant difference (p = 0.029, comparison of proportion Chi-squared test), thus showing the value of including radiomic features.

4. Discussion

This study aimed to evaluate HER2 expression level and pathologic response after NAC in HER2 overexpressing breast cancer patients through the analysis of MRI images. Our results show that MRI features are associated with differences in HER2 expression levels and pathological response after NAC. The best accuracy was achieved when the machine learning models included both clinical and radiomic MRI parameters.

Pathological intratumour heterogeneity with respect to differences in HER2 expression levels has been described in 1–40% of HER2 breast cancers [6]. Many authors have reported that pathological intratumour HER2 heterogeneity is associated with a worse prognosis [3–7]; however, there is no consensus on the best way to evaluate this pathological intratumour heterogeneity in clinical practice. In most studies, HER2 intratumour heterogeneity is evaluated based on both protein expression and gene signal on core needle biopsies obtained at diagnosis. However, new approaches have been proposed. Metzger Filho et al. performed core biopsies from two distinct areas of each tumour (three cores/site) in 164 patients and defined intratumour HER2 heterogeneity as at least one area demonstrating either HER negativity or HER2 positivity by FISH in less than 50% of tumour cells. They found a correlation between intratumour heterogeneity and pathologic response, even when stratifying by ER status and HER2 IHC [4].

However, the HER2 status of a core biopsy specimen, even when obtained from several sites within a tumour, is limited in providing comprehensive information on the pathologic heterogeneity of the tumour in its entirety. In contrast, MRI combined with radiomics analysis allows a non-invasive evaluation of the entire tumour. The model presented here showed 97% diagnostic accuracy to detect differences in HER2 expression, suggesting that such a model could help to select patients with HER2 gene amplification without HER2 protein overexpression who would benefit from anti-HER2 therapy. The only radiomics parameter retained by the model to distinguish HER2 expression levels between IHC and FISH assignment of HER2 status was large zone emphasis. This parameter attempts to quantify heterogeneity via the assessment of zones of similar intensity, and as the name suggests, accentuates the presence of larger zones. Higher values are indicative of coarser texture, i.e. higher radiomics heterogeneity. Based on the data in Table 3, such higher radiomics heterogeneity is associated with lesions with high levels of HER2 protein 3+ overexpression on IHC (IHC group).

Our sample of HER2 overexpressing breast carcinomas was composed of large tumours, with high rates of multifocality as well as suspicious axillary lymph nodes, which is consistent with the literature [17–19]. In our study, the pCR rate among patients determined to be HER2+ by IHC was 65%, compared to 22% among patients in the FISH group, similar to the results of other investigators [4,7]. These findings suggest that the degree of benefit provided by HER2-targeted

Table 3
Significant radiomics features according to HER2 expression level (n = 311). Parameters retained in the final predictive model are italicised.

Parameter	IHC	FISH	*p*-value (Mann–Whitney U test)
Variance	2250 (935–5711)	1149 (−3407 to 2839)	0.0007
Run length variance	0.738 (0.391–1.438)	0.486 (−0.948 to 1.112)	0.0045
Large zone emphasis	*232.2 (−27.9 to 313.2)*	*−93.9 (−688.6 to 106.0)*	*< 0.0001*
Large zone low grey level emphasis	37.56 (−13.75 to 40.00)	−15.31 (−60.98 to 13.32)	< 0.0001
Zone length variance	235.5 (−38.6 to 299.8)	−162.2 (−714.7 to 94.7)	< 0.0001

Abbreviations: IHC, immunohistochemistry; FISH, fluorescence in situ hybridisation.

Table 4
Significant radiomics features according to pathologic response for the training set (n = 249). Parameters retained in the final predictive model are italicised.

Parameter	No pCR	pCR	*p*-value (Mann–Whitney U test)
Mean	114 (90 to 146)	129 (101 to 169)	0.016
Standard deviation	43.9 (32.5 to 63.5)	51.3 (37.2 to 72.4)	0.040
Variance	*1596 (803 to 3361)*	*2989 (1088 to 6079)*	*0.001*
Median	112 (88 to 140)	123 (100 to 161)	0.024
Entropy	3.37 (3.01 to 3.82)	*3.59 (3.19 to 3.84)*	*0.029*
Root mean square	127 (99 to 163)	143 (116 to 184)	0.012
Mean absolute deviation	31.9 (24.8 to 47.9)	38.3 (29.3 to 52.8)	0.043
Median absolute deviation	32.0 (24.6 to 47.5)	37.9 (29.4 to 52.4)	0.042
P90	*171 (131 to 218)*	*192 (150 to 236)*	*0.008*
Long run emphasis	2.38 (1.73 to 3.70)	2.56 (1.96 to 4.07)	0.039
Run length variance	0.475 (0.136 to 1.172)	0.779 (0.352 to 1.510)	0.001
Large zone emphasis	1.0 (-126.0 to 188.8)	245.5 (−35.5 to 245.5)	< 0.0001
Zone length variance	*-8.0 (-136.9 to 166.1)*	*249.3 (−44.6 to 249.3)*	*< 0.0001*

Abbreviations: pCR, pathologic complete response.

therapy may be different in patients with HER2 gene amplification with and without HER2 protein overexpression, which is expected, since these drugs target the HER2 receptor on the cell surface.

Previous studies have shown that MRI-based radiomic signatures enable the assessment of breast cancer molecular subtypes with high diagnostic accuracy [20–22]. Radiomics models using pretreatment MRI features have also been used to predict pCR after NAC. For triple negative and HER2 overexpressing subtypes, different models have shown accuracy rates to predict pCR of 70% to 93% [23–25]. In the present study, the MRI-based model showed an 83.9% diagnostic accuracy to predict pCR among HER2 overexpressing tumours. Three of the radiomics parameters utilised in the determination of pCR, variance, entropy and zone level variance, are measures of the spread of values present within the lesion, with higher values representing greater spread. Lesions with pCR showed higher values for these parameters, suggesting that increased radiomic heterogeneity is advantageous in terms of treatment response. Significantly higher values of the 90th percentile in pCR lesions seems to indicate that lesions with hotspots of enhancement are more likely to respond well to treatment.

In our study, higher levels of HER2 protein overexpression as well favourable treatment response were associated with greater radiomics intratumour heterogeneity. This seems counter-intuitive as prior studies have shown that pathologic intratumour heterogeneity is associated with a worse prognosis [3–7]. Our data suggest that radiomic intratumour heterogeneity does not equal pathologic intratumour heterogeneity but can be used in prediction models as a valuable parameter. This finding is supported by Rauch et al. who also demonstrated that radiomic intratumoural heterogeneity seems to be different from pathological intratumuor heterogeneity, in this case with respect to the level of tumour infiltrating lymphocytes, and was shown to be associated with achievement of pCR [26].

The clinical parameters selected in the final models to predict both HER2 expression level and pCR did not necessarily have a statistically significant association with the outcome in univariate analysis. Larger tumour size was the only clinical parameter retained that showed a statistically significant correlation to pCR in univariate analysis. On the other hand, even though ER status was a significant parameter to predict pCR in univariate analysis, it was not retained in the final model.

This study has some limitations related to its retrospective nature. We included MR images from different sites with different acquisition protocols, which could affect the radiomics analysis despite the

Table 5
Extract of radiomic features determination to forward to model development for prediction of pathological complete response, utilising correlation analysis.

	Spearman rank correlation coefficients						AUROC
	Variance	Entropy	P90	rlv	lze	zlv	
Variance		0.883	0.791	0.534	0.578	0.584	0.697
Entropy	0.883		0.868	0.301	0.290	0.287	0.655
P90	0.791	0.868		0.203	0.216	0.220	0.671
rlv	0.534	0.301	0.203		0.928	0.910	0.693
lze	0.578	0.290	0.216	0.928		0.997	0.749
zlv	0.584	0.287	0.220	0.910	0.997		0.761

Abbreviations: AUROC, area under the receiver operating curve; P90, 90th percentile; rlv, run length variance; lze, large zone emphasis; zlv, zone length variance.

Table 6
Results from the final model to predict HER2 intratumour expression levels (IHC vs. FISH), which utilised three MRI parameters (two clinical parameters [lesion type and multifocality] and one radiomic parameter [large zone emphasis]).

Predictive Model Result	HER2 expression		Total
	IHC	FISH	
IHC	277 (97.9%)	6 (2.1%)	283 (100%)
FISH	2 (7.1%)	26 (92.9%)	28 (100%)
Total	279 (89.7%)	32 (10.3%)	311 (100%)

Abbreviations: IHC, immunohistochemistry; FISH, fluorescence in situ hybridisation.

Table 7
Results from the model to predict pCR included six MRI parameters (two clinical parameters [lesion type and size] and four radiomic parameters [variance, entropy, 90th percentile and zone length variance]) in the training set.

Predictive model result	Pathologic complete response (pCR)		Total
	No pCR	pCR	
No pCR	80 (80.8%)	19 (19.2%)	99 (100%)
pCR	18 (12.0%)	132 (88.0%)	150 (100%)
Total	98 (39.4%)	151 (60.6%)	249 (100%)

Table 8
Results from the model to predict pCR included six MRI parameters (two clinical parameters [lesion type and size] and four radiomic parameters [variance, entropy, 90th percentile and zone length variance]) in the test set.

Predictive model result	Pathologic complete response (pCR)		Total
	No pCR	pCR	
No pCR	20 (80.0%)	5 (20.0%)	25 (100%)
pCR	5 (13.5%)	32 (86.5%)	37 (100%)
Total	25 (40.3%)	37 (59.7%)	62 (100%)

harmonisation process used to reduce this effect. On the other hand, different protocols are commonly seen in clinical practice and including all studies could improve the reproducibility of the method in a multi-institutional setting. Secondly, we used HER2 expression level at percutaneous biopsy to assess intratumour heterogeneity, the approach used in previous studies. While whole-tumour evaluation of HER2 heterogeneity would be the best approach, it was not feasible as all included patients received neoadjuvant chemotherapy. It is also acknowledged that the model developed to assess radiomic features in HER2 IHC and FISH lesions demonstrates associations only rather than predictions, since no test set was utilised due to the small size of the minority class.

In conclusion, machine learning models, including both clinical and radiomics-based MRI features, can be used to assess HER2 expression level and can predict pCR after NAC in HER2 overexpressing breast cancer patients. This information could be used to better select patients who could benefit from anti-HER2 treatment after validation in larger multi-institutional studies.

Declaration of Interests

Katja Pinker received payment for activities not related to the present article including lectures including service on speakers bureaus and for travel/accommodations/meeting expenses unrelated to activities listed from the European Society of Breast Imaging (MRI educational course, annual scientific meeting), IDKD educational course 2019, and Siemens Healthineers. Maxine S Jochelson has received honoraria not related to the present article including an honorarium from GE for speaking, and an honorarium for speaking at the Lynn Sage Breast Cancer Symposium and at MD Anderson. Elizabeth A Morris has received a grant from GRAIL Inc. for research not related to the present article. Monica Morrow has received honoraria from Genomic Health. The rest of the authors report no potential conflicts of interest.

Acknowledgements

This work was supported by the NIH/NCI Cancer Center Support Grant (P30 CA008748), Susan G. Komen Foundation, Breast Cancer Research Foundation, Spanish Foundation Alfonso Martin Escudero, and European School of Radiology. The funding sources had no role in study the design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. The authors thank Joanne Chin, MFA, ELS, for her help in editing the manuscript.

References

[1] Hagemann IS. Molecular testing in breast cancer: a guide to current practices. Arch Pathol Lab Med 2016;140(8):815–24.
[2] Wolff AC, Hammond MEH, Allison KH, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. J Clin Oncol 2018;36(20):2105–22.
[3] Rye IH, Trinh A, Saetersdal AB, et al. Intratumor heterogeneity defines treatment-resistant HER2+ breast tumors. Mol Oncol 2018;12(11):1838–55.
[4] Filho OM, Viale G, Trippa L, et al. HER2 heterogeneity as a predictor of response to neoadjuvant T-DM1 plus pertuzumab: results from a prospective clinical trial. J Clin Oncol 2019;37(15_suppl):502.
[5] Hou Y, Nitta H, Wei L, et al. HER2 intratumoral heterogeneity is independently associated with incomplete response to anti-HER2 neoadjuvant chemotherapy in HER2-positive breast carcinoma. Breast Cancer Res Treat 2017;166(2):447–57.
[6] Muller KE, Marotti JD, Tafe LJ. Pathologic features and clinical implications of breast cancer with HER2 intratumoral genetic heterogeneity. Am J Clin Pathol 2019;152(1):7–16.
[7] Krystel-Whittemore M, Xu J, Brogi E, et al. Pathologic complete response rate according to HER2 detection methods in HER2-positive breast cancer treated with neoadjuvant systemic therapy. Breast Cancer Res Treat 2019;177(1):61–6.
[8] van Ramshorst MS, Loo CE, Groen EJ, et al. MRI predicts pathologic complete response in HER2-positive breast cancer after neoadjuvant chemotherapy. Breast Cancer Res Treat 2017;164(1):99–106.
[9] 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.
[10] Wu J, Cao G, Sun X, et al. Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy. Radiology 2018;288(1):26–35.
[11] Mann RM, Balleyguier C, Baltzer PA, et al. Breast MRI: EUSOBI recommendations for women's information. Eur Radiol 2015;25(12):3669–78.
[12] Morris EA, Comstock CE, Lee CH, et al. ACR BI-RADS® magnetic resonance imaging. ACR BI-RADS® atlas, breast imaging reporting and data system. Reston, VA: American College of Radiology; 2013.
[13] Allison KH, Hammond MEH, Dowsett M, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. J Clin Oncol 2020;38(12):1346–66.
[14] Apte AP, Iyer A, Crispin-Ortuzar M, et al. Technical Note: extension of CERR for computational radiomics: a comprehensive MATLAB platform for reproducible radiomics research. Med Phys 2018;45(8):3713–20.
[15] Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295(2):328–38.
[16] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8(1):118–27.
[17] Bitencourt AG, Pereira NP, França LK, et al. Role of MRI in the staging of breast cancer patients: does histological type and molecular subtype matter? Br J Radiol 2015;88(1055):20150458.
[18] Grimm LJ, Johnson KS, Marcom PK, Baker JA, Soo MS. Can breast cancer molecular subtype help to select patients for preoperative MR imaging? Radiology 2015;274(2):352–8.
[19] Ha R, Jin B, Mango V, et al. Breast cancer molecular subtype as a predictor of the utility of preoperative MRI. AJR Am J Roentgenol 2015;204(6):1354–60.
[20] Leithner D, Bernard-Davila B, Martinez DF, et al. Radiomic signatures derived from diffusion-weighted imaging for the assessment of breast cancer receptor status and molecular subtypes. Mol Imaging Biol 2020;22(2):453–61.
[21] Leithner D, Horvat JV, Marino MA, et al. Radiomic signatures with contrast-enhanced magnetic resonance imaging for the assessment of breast cancer receptor status and molecular subtypes: initial results. Breast Cancer Res 2019;21(1):106.
[22] Fan M, Li H, Wang S, Zheng B, Zhang J, Li L. Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer. PLoS One 2017;12(2):e0171683.
[23] 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(2):455–63.
[24] Liu Z, Li Z, Qu J, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. Clin Cancer Res 2019;25(12):3538–47.
[25] Braman N, Prasanna P, Whitney J, et al. Association of peritumoral radiomics with tumor biology and pathologic response to preoperative targeted therapy for HER2 (ERBB2)-positive breast cancer. JAMA Netw Open 2019;2(4):e192561.
[26] Rauch G, Zhu H, Li H, et al. Abstract PD2-09: Association of quantitative MRI features with tumor infiltrating lymphocytes and treatment response prediction in HER2 positive subtype of breast cancer. Cancer Res 2018;78(4 Supplement):PD2–09.