Radiology

ORIGINAL RESEARCH • BREAST IMAGING

MRI-based Quantification of Intratumoral Heterogeneity for Predicting Treatment Response to Neoadjuvant Chemotherapy in Breast Cancer

*Zhenwei Shi, PhD** • *Xiaomei Huang, MD** • *Ziliang Cheng, MD* • *Zeyan Xu, MD* • *Huan Lin, MD* • *Chen Liu, MD* • *Xiaobo Chen, MD* • *Chunling Liu, MD* • *Changhong Liang, MD* • *Cheng Lu, PhD* • *Yanfen Cui, MD* • *Chu Han, PhD* • *Jinrong Qu, MD* • *Jun Shen, MD*** • *Zaiyi Liu, MD***

From the Department of Radiology (Z.S., Z.X., H.L., Chen Liu, X.C., Chunling Liu, C. Liang, C. Lu, Y.C., C.H., Z.L.), Guangdong Cardiovascular Institute (Z.S., Y.C.), and Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (Z.S., Z.X., H.L., Chen Liu, X.C., C. Liang, C. Lu, Y.C., C.H., Z.L.), Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, No. 106 Zhongshan Er Road, Guangzhou 510080, Guangdong, China; Department of Medical Imaging, Nanfang Hospital, Southern Medical University, Guangzhou, China (X.H.); Department of Radiology, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China (Z.C., J.S.); School of Medicine, South China University of Technology, Guangzhou, China (Z.X., H.L.); The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China (Chen Liu, X.C.); Department of Radiology, Shanxi Cancer Hospital, Shanxi Medical University, Taiyuan, China (Y.C.); and Department of Radiology, The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou, China (J.Q.). Received November 5, 2022; revision requested November 23; revision received May 15, 2023; accepted May 19. **Address correspondence to** Z.L. (email: *liuzaiyi@gdph.org.cn*).

This work was supported by the Key-Area Research and Development Program of Guangdong Province (2021B0101420006), Regional Innovation and Development Joint Fund of National Natural Science Foundation of China (U22A20345), National Key Research and Development Program of China (2021YFF1201003), National Science Fund for Distinguished Young Scholars (81925023), Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (2022B1212010011), National Natural Science Foundation of China (82102034, 62102103, 82001789, 82171920, 12126610, 82171996, U1801681), China Postdoctoral Science Foundation (2022M710843, 2021M700897), Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2017), The Natural Science Foundation for Distinguished Young Scholars of Guangdong Province (2023B1515020043), and Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) (GML-KF-22-30).

* Z.S. and X.H. contributed equally to this work.

** J.S. and Z.L. are co-senior authors.

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

See also the editorial by Rauch in this issue.

Radiology 2023; 308(1):e222830 • https://doi.org/10.1148/radiol.222830 • Content codes: BR MR

Background: Breast cancer is highly heterogeneous, resulting in different treatment responses to neoadjuvant chemotherapy (NAC) among patients. A noninvasive quantitative measure of intratumoral heterogeneity (ITH) may be valuable for predicting treatment response.

Purpose: To develop a quantitative measure of ITH on pretreatment MRI scans and test its performance for predicting pathologic complete response (pCR) after NAC in patients with breast cancer.

Materials and Methods: Pretreatment MRI scans were retrospectively acquired in patients with breast cancer who received NAC followed by surgery at multiple centers from January 2000 to September 2020. Conventional radiomics (hereafter, C-radiomics) and intratumoral ecological diversity features were extracted from the MRI scans, and output probabilities of imaging-based decision tree models were used to generate a C-radiomics score and ITH index. Multivariable logistic regression analysis was used to identify variables associated with pCR, and significant variables, including clinicopathologic variables, C-radiomics score, and ITH index, were combined into a predictive model for which performance was assessed using the area under the receiver operating characteristic curve (AUC).

Results: The training data set was comprised of 335 patients (median age, 48 years [IQR, 42–54 years]) from centers A and B, and 590, 280, and 384 patients (median age, 48 years [IQR, 41–55 years]) were included in the three external test data sets. Molecular subtype (odds ratio [OR] range, 4.76–8.39 [95% CI: 1.79, 24.21]; all $P < .01$), ITH index (OR, 30.05 [95% CI: 8.43, 122.64]; $P < .001$), and C-radiomics score (OR, 29.90 [95% CI: 12.04, 81.70]; $P < .001$) were independently associated with the odds of achieving pCR. The combined model showed good performance for predicting pCR to NAC in the training data set (AUC, 0.90) and external test data sets (AUC range, 0.83–0.87).

Conclusion: A model that combined an index created from pretreatment MRI-based imaging features quantitating ITH, C-radiomics score, and clinicopathologic variables showed good performance for predicting pCR to NAC in patients with breast cancer.

Supplemental material is available for this article.

Neoadjuvant chemotherapy (NAC) has been widely used as a first-line treatment in patients with locally advanced breast cancer (1). However, breast tumor response to NAC varies greatly among patients, with an estimated 19%–30% of patients experiencing pathologic complete response (pCR) and 5%–20% exhibiting disease progression (2–4). Patients with pCR following NAC receive de-escalation surgery and have a more favorable outcome than patients without pCR (5,6). Currently, there are no standard methods or imaging biomarkers

Abbreviations

AUC = area under the receiver operating characteristic curve, ITH = intratumoral heterogeneity, NAC = neoadjuvant chemotherapy, OR = odds ratio, pCR = pathologic complete response

Summary

A model combining clinicopathologic variables, conventional radiomics, and a quantitative MRI-based measure of intratumoral heterogeneity showed good performance for predicting pathologic complete response to neoadjuvant chemotherapy in patients with breast cancer.

Key Results

- In this retrospective study, pretreatment MRI scans from 335 patients with breast cancer were used to develop a quantitative measure of intratumoral heterogeneity (ITH) that achieved reasonable performance for predicting pathologic complete response (pCR) to neoadjuvant chemotherapy in three external test data sets that included a total of 1254 patients (area under the receiver operating characteristic curve [AUC] range, 0.74–0.76).
- A model combining clinicopathologic variables, conventional radiomics features, and the ITH index achieved good performance for predicting pCR in the three external test data sets (AUC range, 0.83–0.87).

available in clinical practice for accurately predicting pCR to NAC in patients with breast cancer.

Breast tumor heterogeneity introduces challenges in clinical treatment decision-making and follow-up management (7,8). This heterogeneity may also account for differences in breast tumor responses to NAC, especially in patients with otherwise similar clinical characteristics (eg, clinical stage, molecular subtype). Consequently, tumor biomarkers that reflect characteristics of the whole breast tumor but neglect intratumoral heterogeneity (ITH) have shown limited predictive value for treatment response (9). Thus, a quantitative measure of ITH could be a valuable biomarker for prediction of clinical outcomes.

MRI is recognized as the most sensitive imaging modality for assessing breast cancer treatment response to NAC (10). Extracting radiomic features from MRI can provide additional information that has been shown to improve diagnosis, treatment, and prognosis of breast cancer (11–13). Indeed, radiomic analysis of dynamic contrast-enhanced MRI in 95 patients with primary invasive breast cancer enabled categorization of patients based on low, medium, and high ITH (13). While this categorization demonstrated prognostic value, the measure of ITH used was not quantitative. Another study used histopathologic images to develop a quantitative measure of breast tumor microenvironment heterogeneity, which was associated with clinical outcomes (14), but whether a similar technique could be applied to MRI is unknown.

Therefore, the aim of this study was to develop a quantitative measure of ITH using radiomic analysis of tumor subregions on MRI scans and to test whether a model combining this quantitative measure with clinicopathologic variables and conventional radiomics (hereafter, C-radiomics) could predict pCR to NAC in breast cancer prior to treatment.

Materials and Methods

Study Sample

This retrospective multicenter study was approved by the institutional review committee at each participating center and the need for patient written informed consent was waived. This study was designed using the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines and represents a type III study (15).

Data sets used in this study included pretreatment MRI scans in female patients with breast cancer who received NAC followed by surgery from March 2014 to September 2020 at centers A, B, and C (all three are academic medical centers). Two publicly available data sets from The Cancer Imaging Archive (16) were also used in this analysis, which were the Duke Breast Cancer MRI study conducted from January 2000 to March 2014 (data set D) (17) and the I-SPY2 trial conducted from May 2010 to November 2016 (data set E) (18). The inclusion criteria for all were as follows: *(a)* biopsy-confirmed invasive breast cancer, *(b)* available pretreatment MRI studies within a month before NAC, and *(c)* postoperative pathologic evaluation of treatment response to NAC. Exclusion criteria were *(a)* missing or inadequate quality of MRI, *(b)* prior history of breast cancer, *(c)* incompletion of NAC, *(d)* special histologic type or lack of complete pre-NAC histopathologic data, *(e)* not receiving NAC, and *(f)* missing pathologic response information (Fig 1). The training data set was comprised of patients from centers A and B and the external test data sets were comprised of patients from center C and the publicly available data sets.

The patients received a NAC regimen involving either taxane-based, alkylator-based, and anthracycline-based chemotherapy alone, or in combination with neoadjuvant endocrine and/or anti–human epidermal growth factor receptor 2 targeted treatment based on molecular subtypes.

Histopathologic Assessment and Definition of pCR

For data sets from centers A, B, and C, histopathologic assessment was performed independently by two site pathologists with more than 5 years of experience in breast pathology who were blinded to MRI findings. Disagreements were reviewed by the two pathologists and a consensus was reached by discussion. The details of clinicopathologic variables are described in Appendix S1. For data sets D and E, the histopathologic evaluation was also performed by site pathologists and details have been reported in prior studies (17,18). In the current study, pCR was defined as no residual invasive cancer in both the breast and axillary lymph nodes, while cancer in situ was allowed in some cases (ypT0/Tis ypN0).

MRI Procedure and Image Preprocessing

Acquisition parameters for MRI are shown in Appendix S1 and Table S1. The phase with peak tumor enhancement at pretreatment dynamic contrast-enhanced MRI was used, and image quality was assessed using the open-source tool MRQy *(https://github.com/ccipd/MRQy)* (Appendix S1, Fig S1) (19). Image preprocessing included N4 bias field correction (20), resampling

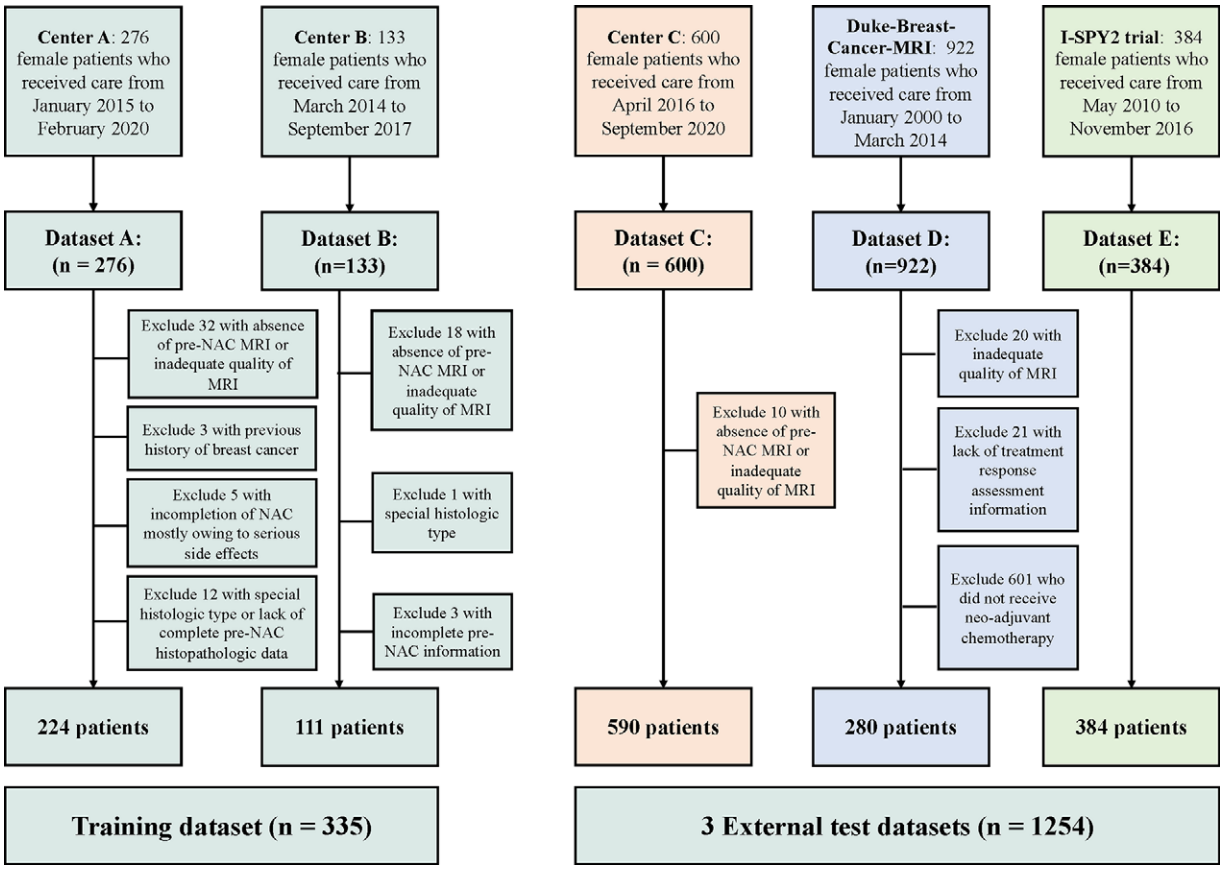

Figure 1: Flowchart shows patient exclusion for each data set. The Duke Breast Cancer MRI and I-SPY2 trial data sets are publicly available and were downloaded from The Cancer Imaging Archive. NAC = neoadjuvant chemotherapy.

isotropic voxels of 1 mm³ using B-spline interpolation, and histogram standardization of intensity values on MRI scans (21).

Intratumoral Subregion Segmentation and Ecological Diversity Feature Generation

The three-dimensional tumor region was manually delineated by four radiologists (X.H., Z.C., Y.C., and Chunling Liu, with 5, 5, 8, and 10 years of experience, respectively, in breast MRI) using ITK-SNAP (version 3.8.0; *http://www.itksnap.org/pmwiki/pmwiki.php*) software. To extract intratumoral subregions objectively, a simple linear interactive clustering method was used (Fig 2, Appendix S1, Fig S2). A total of 105 C-radiomics features (18 histogram-based, 14 shape-based, and 73 texture-based) complying with the image biomarker standardization initiative (22) were extracted from the whole tumor region and intratumoral subregions using PyRadiomics (version 3.1; *https://pyradiomics.readthedocs.io*) (23). Afterward, a Gaussian mixture model was used to cluster intratumoral subregions with similar C-radiomics features. The optimal number of clusters representing the diversity of the tumor ecosystem was identified by using the Bayesian information criteria in an unbiased manner, with a cluster range from 1 to 5 (14).

Feature Dimension Reduction and Prediction Model Development

The dimension of imaging features was reduced using the feature selection method for minimum redundancy and maximum relevancy (Appendix S1, Table S2). For intratumoral ecological diversity and C-radiomics features, interobserver reproducibility was assessed first using intraclass correlation coefficients. For this analysis, a tool was built in-house and used to generate five pseudo masks that imitate human delineation, then imaging features were extracted from the pseudo masks for reproducibility assessment (Appendix S1). Intratumoral ecological diversity and C-radiomics features were included in predictive models if they achieved an interclass correlation coefficient greater than 0.90 and greater than 0.75, respectively. Second, the variance of each feature value was evaluated and the features with low variance (<75%) were removed. Third, the Pearson correlation coefficient was calculated to eliminate features that were highly correlated with one another (>0.75). Fourth, the *t* test or Mann-Whitney *U* test was used to compare positive and negative samples for each feature. Ultimately, five intratumoral ecological diversity and six C-radiomics features with a *P* value less than .05 were selected for model development. The feature importance of intratumoral ecological diversity and C-radiomics features are shown in Appendix S1 and Tables S3 and S4.

For clinicopathologic variables, univariable logistic regression was used to evaluate their association with pCR, and variables with $P < .05$ were fed into the multivariable logistic regression analysis by using backward stepwise selection with the Akaike information criterion. Also, a heat map was generated to visualize

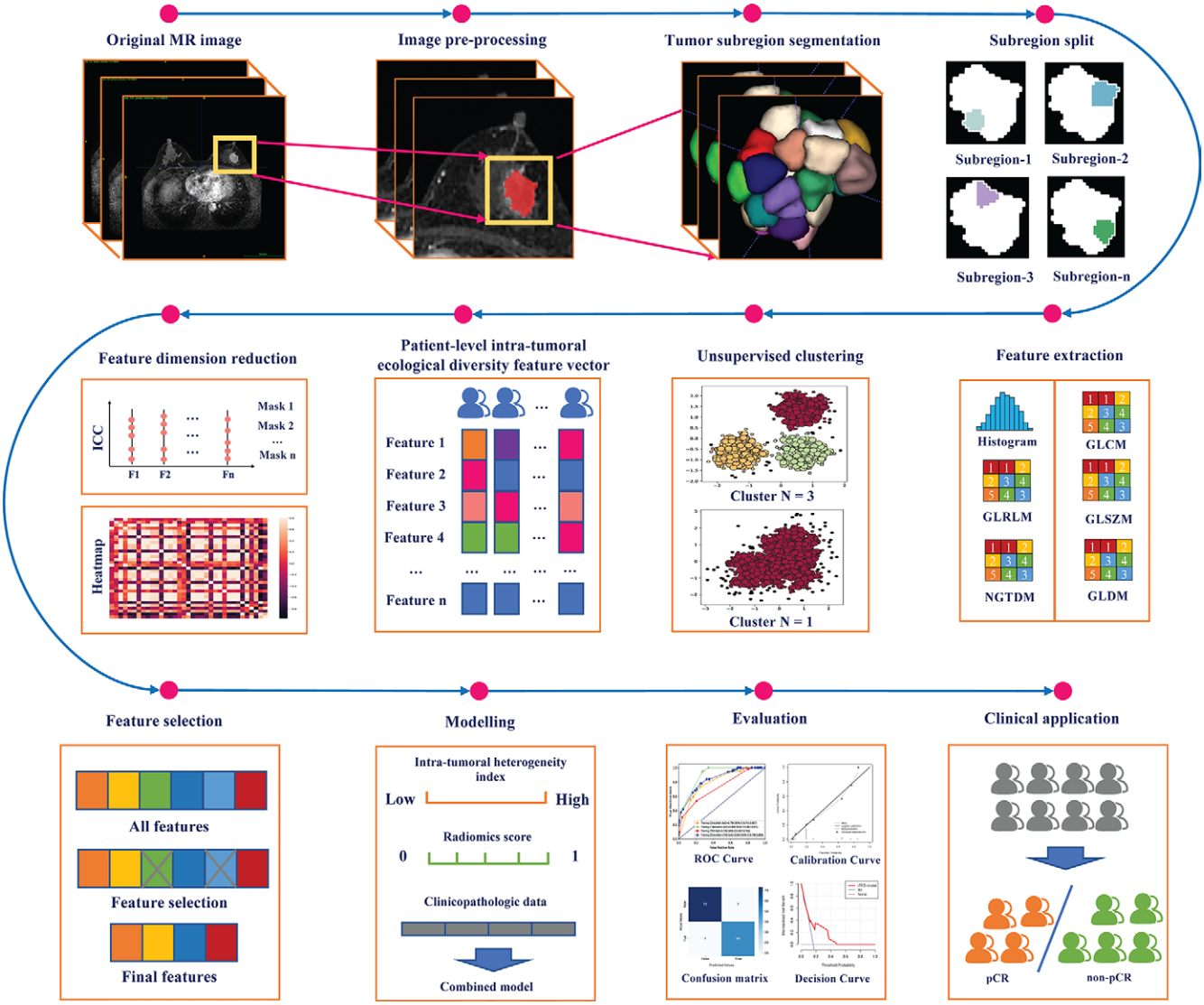

Figure 2: Schematic shows the workflow of the proposed breast cancer intratumoral heterogeneity (ITH) assessment method used in this study. First, the intratumoral subregions were segmented from manually delineated tumor regions. Second, an intratumoral ecological diversity feature vector was generated per patient. Third, feature selection processes were performed to select the features with predictive power. Finally, a combined model constructed using the ITH index, conventional radiomics score, and clinicopathologic data was evaluated using multiple external test data sets. GLCM = gray-level co-occurrence matrix, GLDM = gray-level dependence matrix, GLRLM = gray-level run-length matrix, GLSZM = gray-level size-zone matrix, ICC = intraclass correlation coefficient, NGTDM = neighboring gray-tone difference matrix, pCR = pathologic complete response, ROC = receiver operating characteristic.

the association of these significant variables with pCR for each patient across all data sets.

Ultimately, seven model types were developed using *(a)* clinicopathologic variables (clinicopath model), *(b)* intratumoral ecological diversity features (ITH model), *(c)* C-radiomics features (C-radiomics model), *(d)* intratumoral ecological diversity and C-radiomics features (ITH-radiomics model), *(e)* clinicopathologic variables and intratumoral ecological diversity features (clinicopath-ITH model), *(f)* clinicopathologic variables and C-radiomics features (clinicopath-radiomics model), and *(g)* clinicopathologic variables, intratumoral ecological diversity, and C-radiomics features (combined model). A multivariable logistic regression and a decision tree (version 1.0.2; scikit-learn) package in Python were used to construct prediction models. The analysis codes are available in a GitHub repository *(https://github.com/zhenweishi/QMITH)*.

Statistical Analysis

Continuous and categorical variables are reported as medians with IQRs and frequencies with percentages, respectively. The Fisher exact test or Pearson χ^2 test and Kruskal-Wallis test were used to compare patient characteristics in the training and external test data sets. Univariable and multivariable logistic regression analyses were performed to assess the association between different patient characteristics and the odds of pCR.

The performance of the models for predicting pCR to NAC was evaluated using the area under the receiver operating characteristic curve (AUC) and 95% CI. Calibration curves, assessed using the Hosmer-Lemeshow goodness-of-fit

Table 1: Comparison of Clinicopathologic Characteristics between the Training and External Validation Data Sets

Characteristic	Training Data Set (*n* = 335)	External Test Data Set C (*n* = 590)	External Test Data Set D (*n* = 280)	External Test Data Set E (*n* = 384)	*P* Value
Age (y)*	48 (42–54)	47 (40–54)	47 (42–55)	49 (41–56)	.14[†]
Menstrual status[‡]					.29[§]
Premenopausal	217 (65)	367 (62)	160 (57)	191 (50)	
Postmenopausal	118 (35)	223 (38)	119 (42)	114 (30)	
Perimenopausal	0 (0)	0 (0)	0 (0)	18 (5)	
Not available	0 (0)	0 (0)	1 (1)	61 (15)	
Pre-NAC clinical T stage[‡]					<.001[§‖]
cT1–2	218 (65)	272 (46)	65 (23)	0 (0)	
cT3–4	117 (35)	318 (54)	213 (76)	0 (0)	
Not available	0 (0)	0 (0)	2 (1)	384 (100)	
Pre-NAC clinical N stage[‡]					<.001[§‖]
cN0	0 (0)	98 (16)	110 (39)	0 (0)	
cN1	220 (66)	412 (70)	120 (43)	0 (0)	
cN2–3	115 (34)	80 (14)	49 (17)	0 (0)	
Not available	0 (0)	0 (0)	1 (1)	384 (100)	
HR status					<.001[§‖]
Positive	225 (67)	492 (83)	170 (61)	222 (58)	
Negative	110 (33)	98 (17)	110 (39)	162 (42)	
HER2 status					<.001[§‖]
Negative	180 (54)	349 (59)	200 (71)	294 (77)	
Positive	155 (46)	241 (41)	80 (29)	90 (23)	
Ki-67 index[‡]					.26[§]
Low proliferation (<20%)	35 (10)	78 (13)	0 (0)	0 (0)	
High proliferation (≥20%)	300 (90)	512 (87)	0 (0)	0 (0)	
Not available	0 (0)	0 (0)	280 (100)	384 (100)	
Molecular subtype					<.001[§‖]
HR positive, HER2 negative	136 (40)	308 (52)	118 (42)	162 (42)	
HR positive, HER2 positive	89 (27)	184 (31)	52 (19)	60 (16)	
HR negative, HER2 positive	66 (20)	57 (10)	28 (10)	30 (8)	
HR negative, HER2 negative	44 (13)	41 (7)	82 (29)	132 (34)	
Treatment response					.01[§‖]
pCR	71 (21)	123 (21)	64 (23)	113 (29)	
Non-pCR	264 (79)	467 (79)	216 (77)	271 (71)	

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. *P* values represent the comparison of clinicopathologic variables across all data sets. HER2 = human epidermal growth factor receptor 2, HR = hormone receptor, NAC = neoadjuvant chemotherapy, pCR = pathologic complete response.

* Data are medians, with IQRs in parentheses.

[†] Kruskal-Wallis test.

[‡] Only patients with available data were used in the analysis.

[§] Fisher exact test or Pearson χ^2 test.

[‖] Indicates statistical significance; $P < .05$.

test, and decision curves were used to evaluate the usefulness of the predictive models. Sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were also calculated. The performance of the combined model was also assessed in prespecified patient subgroups stratified according to age (≤45 years, >45 years), pre-NAC clinical T and N stages, and menstrual status. The output probabilities of the ITH and C-radiomics models were used to generate a patient-level ITH index and C-radiomics score, respectively. A binary ITH index (low vs high) was determined using the median probability (0.55) as the cutoff. To demonstrate that the ITH index is an independent predictor, multivariable logistic regression analysis without backward selection was used to analyze the variables with a *P* value less than .05 from the univariable analysis. Shapley additive explanations analysis (24) was performed to quantify the impact of each intratumoral ecological diversity and C-radiomics feature on the model prediction.

Statistical analyses were carried out by X.H. and Z.S. using R (version 3.6.0; The R Foundation) and SciPy (version 1.8.0; Python Software Foundation). Two-tailed $P < .05$ was considered indicative of a statistically significant difference.

Table 2: Univariable Logistic Regression Analysis for the Association between pCR to NAC and Patient Characteristics

	Training Data Set (n = 335)		
Characteristic	Odds Ratio	95% CI	P Value
Age	0.99	0.96, 1.02	.53
Pre-NAC clinical T stage			
cT1–2	1.77	1.00, 3.26	.06
cT3–4	Reference		
Pre-NAC clinical N stage			
cN1	Reference		
cN2–3	0.90	0.51, 1.55	.70
HR status			
Negative	3.86	2.25, 6.71	<.001*
Positive	Reference		
HER2 status			
Negative	Reference		
Positive	2.62	1.53, 4.57	<.001*
Ki-67 index			
Low proliferation (<20%)	Reference		
High proliferation (≥20%)	1.34	0.57, 3.69	.54
Molecular subtype			
HR positive, HER2 negative	Reference		
HR positive, HER2 positive	3.08	1.41, 7.06	.01*
HR negative, HER2 positive	7.87	3.67, 17.93	<.001*
HR negative, HER2 negative	5.30	2.20, 13.12	<.001*
Binary ITH index			
High	Reference		
Low	6.52	3.73, 11.67	<.001*
C-radiomics score	32.51	14.68, 78.15	<.001*

Note.—P values represent the comparison between patients with pCR and those without pCR after NAC in the training data set. C-radiomics = conventional radiomics, HER2 = human epidermal growth factor receptor 2, HR = hormone receptor, ITH = intratumoral heterogeneity, NAC = neoadjuvant chemotherapy, pCR = pathologic complete response.

* Indicates statistical significance; $P < .05$.

Results

Patient Characteristics

A total of 2315 MRI examinations were initially acquired from three centers and two publicly available data sets. Patients were excluded for poor quality or missing pre-NAC MRI (n = 80), previous history of breast cancer (n = 3), noncompletion of NAC (n = 5), incomplete histologic data or special histologic subtype (n = 13), incomplete pre-NAC information (n = 3), lack of treatment response information (n = 21), or no NAC treatment (n = 601) (Fig 1). The final training data set included 335 patients (center A [n = 224], center B [n = 111]; median age, 48 years [IQR, 42–54 years]). The three external test data sets were comprised of 590 patients from center C (median age, 47 years [IQR, 40–54 years]), 280 patients from data set D (median age, 47 years [IQR, 42–55 years]), and 384 patients from data set E (median age, 49 years [IQR, 41–56 years]). Differences in treatment response rate were observed across data sets (P = .01) (Table 1).

Variables Associated with pCR in the Training Data Set

Univariable analysis showed that hormone receptor status (odds ratio [OR], 3.86 [95% CI: 2.25, 6.71]; $P < .001$), human epidermal growth factor receptor 2 status (OR, 2.62 [95% CI: 1.53, 4.57]; $P < .001$), molecular subtype (OR range, 3.08–7.87; $P \leq .01$), binary ITH index (OR, 6.52 [95% CI: 3.73, 11.67]; $P < .001$), and C-radiomics score (OR, 32.51 [95% CI: 14.68, 78.15]; $P < .001$) were all associated with pCR in the training data set (Table 2). Multivariable analysis demonstrated that molecular subtype, binary ITH index, and C-radiomics score were independent predictors for pCR (Table 3).

After adjustment of the multivariable model without backward selection for the variables with a P value less than .05 in the univariable analysis (Table 2), the ITH index remained an independent predictor for pCR (OR, 5.03 [95% CI: 2.52, 10.37]; $P < .001$).

Performance Evaluation of Prediction Models

For discriminating pCR to NAC, the ITH model achieved AUCs of 0.77 (95% CI: 0.71, 0.84) in the training data set and

Table 3: Multivariable Analysis of Backward Stepwise–selected Patient Variables for their Association with pCR to NAC

	Training Data Set (n = 335)		
Variable	β Coefficient	Odds Ratio	P Value
Intercept	−5.76		<.001*
Molecular subtype			
HR positive, HER2 negative	Reference		
HR positive, HER2 positive	1.56	4.76 (1.79, 13.46)	.01*
HR negative, HER2 positive	2.13	8.39 (3.13, 24.21)	<.001*
HR negative, HER2 negative	1.70	5.51 (1.85, 17.17)	.01*
Binary ITH index			
High	Reference		
Low	3.40	30.05 (8.43, 122.64)	<.001*
C-radiomics score	3.40	29.90 (12.04, 81.70)	<.001*

Note.—Data in parentheses are 95% CIs. C-radiomics = conventional radiomics, HER2 = human epidermal growth factor receptor 2, HR = hormone receptor, ITH = intratumoral heterogeneity, NAC = neoadjuvant chemotherapy pCR = pathologic complete response.

* Indicates statistical significance; $P < .05$.

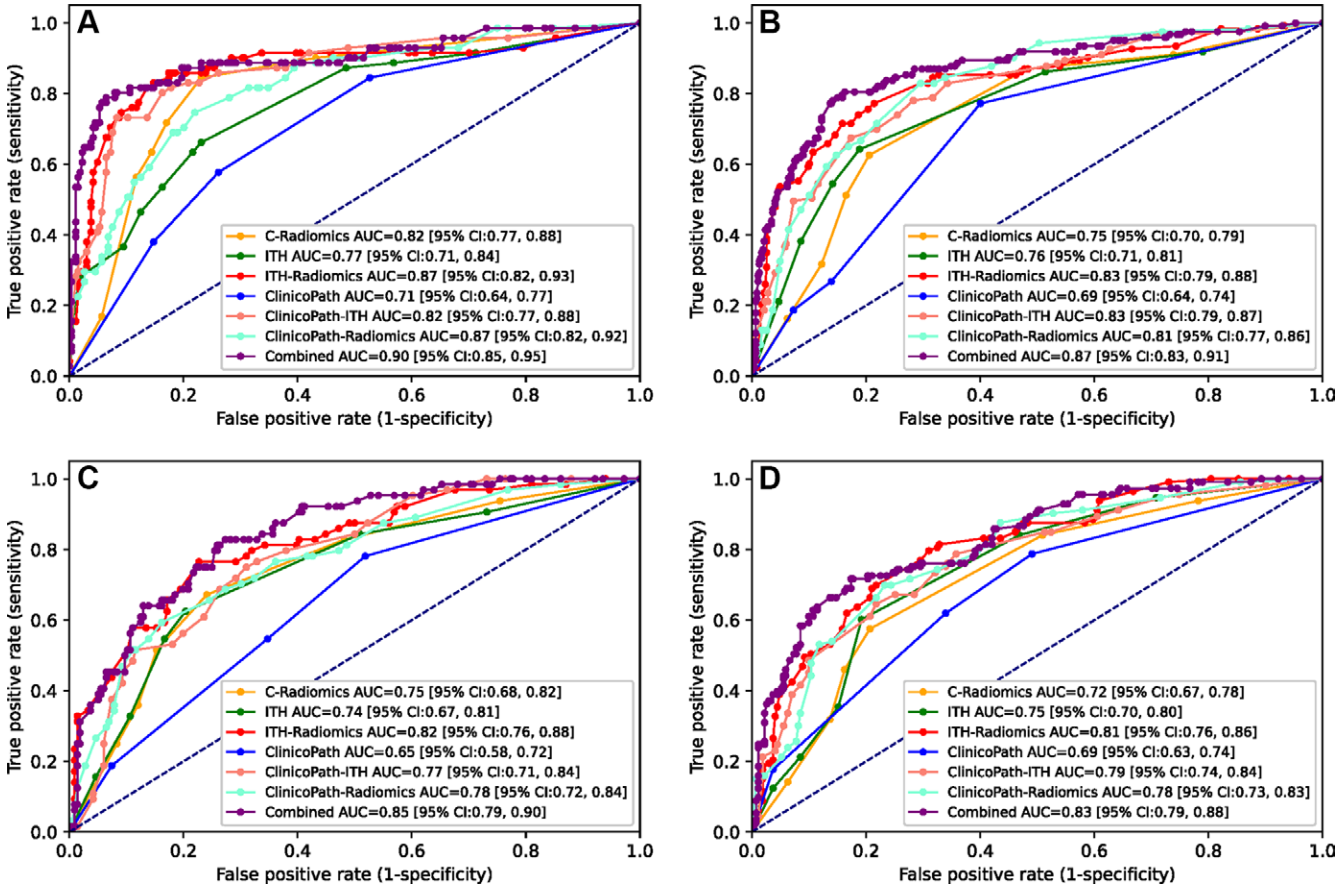

Figure 3: Performance of the seven developed models for predicting pathologic complete response of breast cancer to neoadjuvant therapy in all data sets. Receiver operating characteristic curves of the **(A)** training data set, **(B)** external test data set C, **(C)** external test data set D, and **(D)** external test data set E show that the model combining clinicopathologic variables, conventional radiomics (C-radiomics), and the intratumoral heterogeneity (ITH) index (purple line) achieved the highest area under the receiver operating characteristic curve (AUC) values of all models tested, with AUCs of 0.90, 0.87, 0.85, and 0.83, respectively.

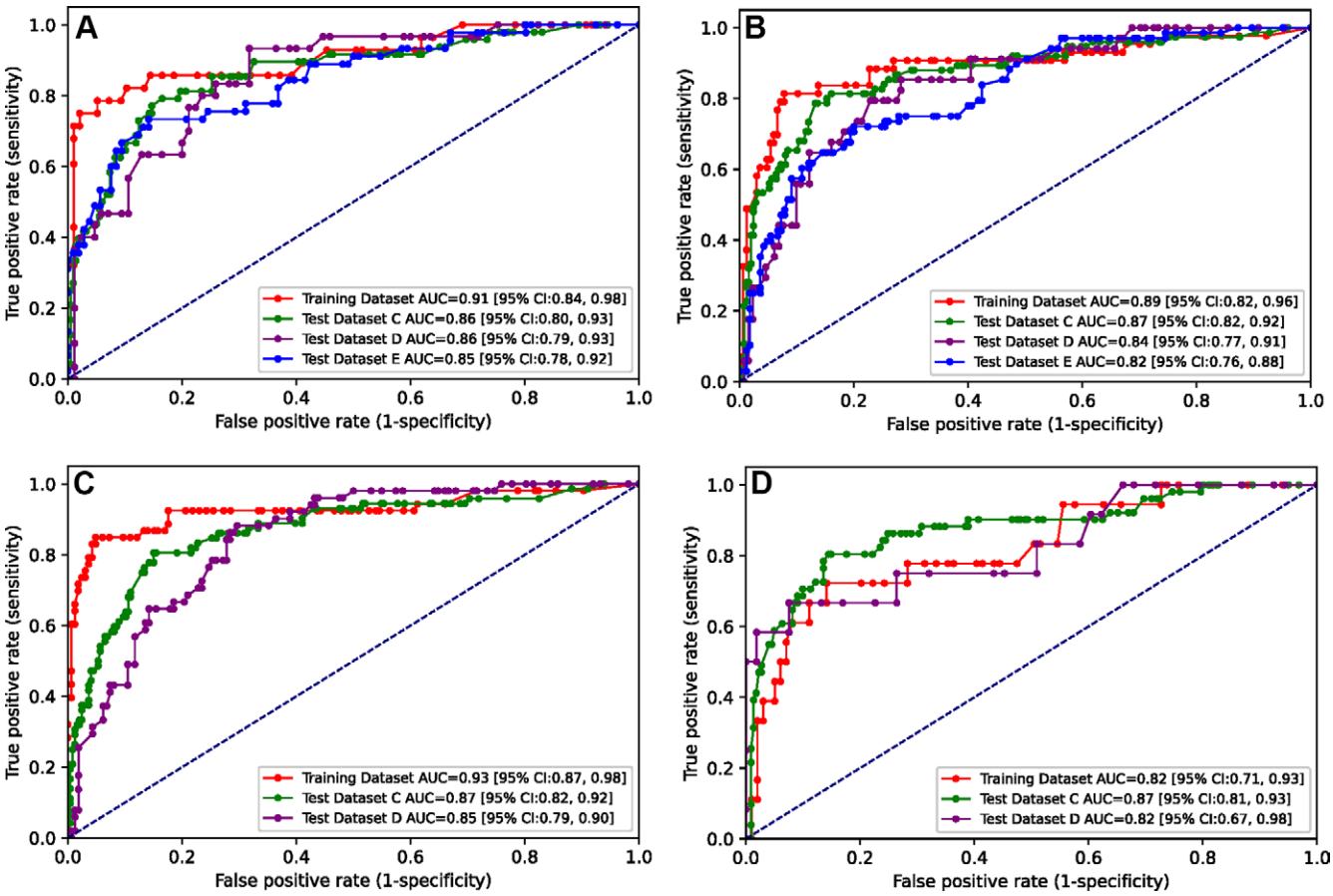

Figure 4: Performance of the combined model in patient subgroups stratified according to patient age, clinical T and N stages before neoadjuvant chemotherapy (NAC), and menstrual status. Receiver operating characteristic curves for **(A)** age less than or equal to 45 years, **(B)** age greater than 45 years, **(C)** pre-NAC clinical T stage 1–2, and **(D)** pre-NAC clinical T stage 3–4 in the training data set (red line) and external test data sets C (green line), D (purple line), and E (blue line) show that the model combining clinicopathologic variables, conventional radiomics, and the intratumoral heterogeneity index achieved moderate performance in all patient subgroups. AUC = area under the receiver operating characteristic curve *(Fig 4 continues)*.

0.76 (95% CI: 0.71, 0.81), 0.74 (95% CI: 0.67, 0.81), and 0.75 (95% CI: 0.70, 0.80) in the test data sets C–E, respectively (Fig 3). Additionally, by integrating clinicopathologic variables with the ITH index and C-radiomics score, the combined model achieved improved predictive performance, with AUCs of 0.90 (95% CI: 0.85, 0.95) in the training data set and 0.87 (95% CI: 0.83, 0.91), 0.85 (95% CI: 0.79, 0.90), and 0.83 (95% CI: 0.79, 0.88) in the external test data sets.

Other diagnostic metrics for the seven models, including sensitivity, specificity, accuracy, positive predictive value, and negative predictive value are shown in Appendix S1 and Table S5. For the combined model, calibration curve analysis showed that the predicted probabilities and the actual probabilities of pCR had reasonable agreement in the external test data sets (Appendix S1, Fig S8). Furthermore, the decision curves suggest that if the threshold probability is in the range of 5%–88%, 5%–87%, 7%–79%, and 9%–85%, the combined model can achieve higher net benefits than the clinicopath model in the training and external test data sets C–E, respectively; if the threshold probability is in the range of 7%–88%, 7%–85%, 9%–35% (and 42%–79%), and 19%–85%, the combined model can achieve higher net benefits than the clinicopath-radiomics model in the training and external test data sets C–E, respectively (Appendix S1, Fig S9).

Evaluation of Model Performance in Patient Subgroups

The number of patients in each category of age, pre-NAC clinical T and N stages, and menstrual status for the training data set and the external test sets is shown in Table 1. Use of the combined model for prediction of pCR to NAC in the external test sets achieved AUCs ranging 0.85–0.86 for patients 45 years or younger, 0.82–0.87 for patients older than 45 years, 0.85–0.87 for patients with a pre-NAC clinical T stage 1–2 and 0.82–0.87 for stage 3–4, 0.85–0.88 for patients with a pre-NAC clinical N stage 0–1 and 0.81–0.82 for stage 2–3, 0.84–0.86 for premenopausal patients, and 0.78–0.88 for postmenopausal patients (Fig 4). Additional subgroup analyses based on Ki-67 index and therapy regimen are shown in Appendix S1 and Figure S3.

Important Model Features and Example Model Interpretation

The impact of intratumoral ecological diversity and C-radiomics features on model predictions of pCR to NAC are visualized in Appendix S1 and Figures S4–S6. The Pearson correlation coeffcients for the generated imaging features and clinicopathologic variables are shown in Appendix S1 and Figure S7. The heat map in Figure 5 suggests that patients with

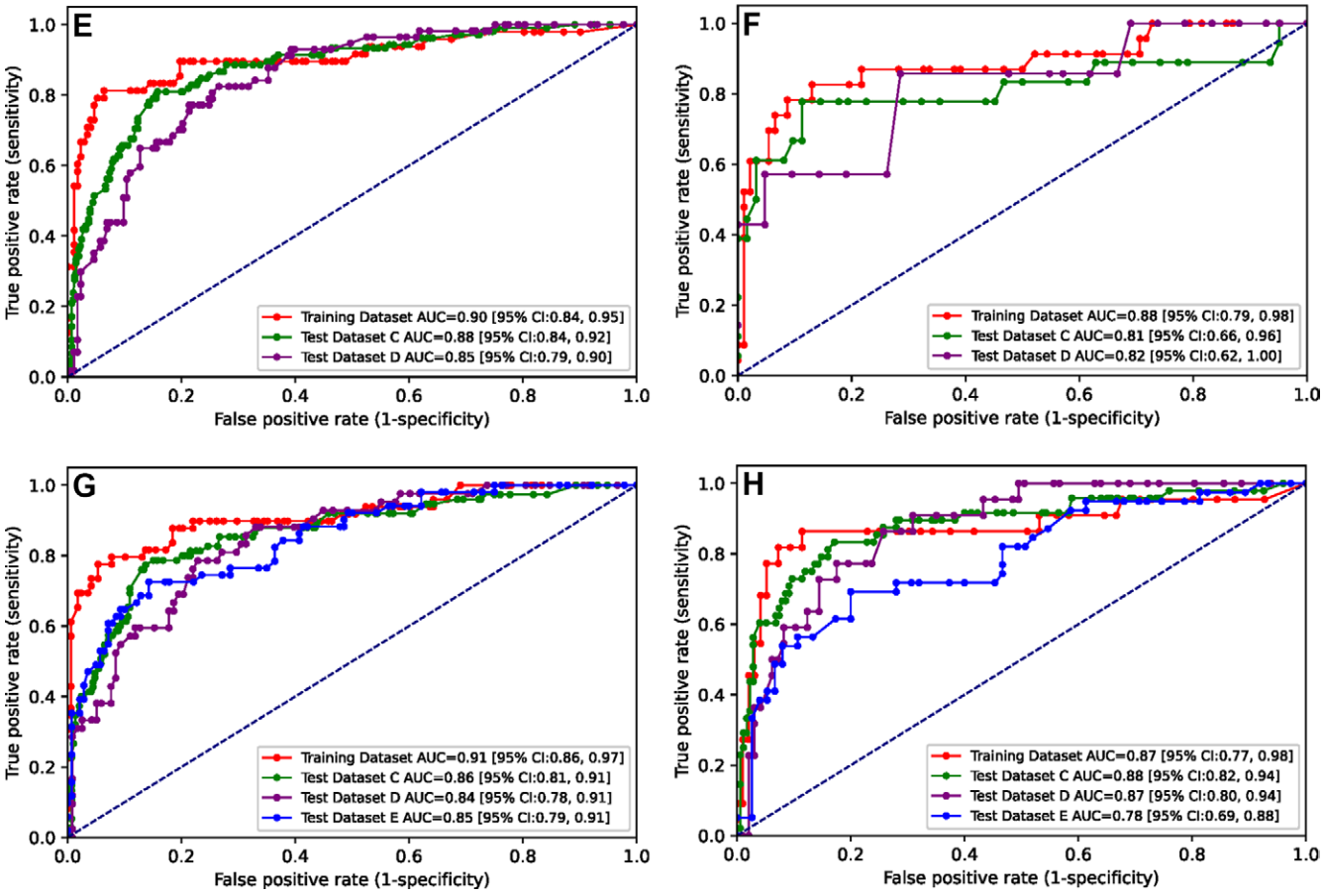

Figure 4 *(continued)*: Performance of the combined model in patient subgroups stratified according to patient age, clinical T and N stages before neoadjuvant chemotherapy (NAC), and menstrual status. Receiver operating characteristic curves for **(E)** pre-NAC clinical N stage 0–1, **(F)** pre-NAC clinical N stage 2–3, **(G)** premenopausal status, and **(H)** postmenopausal status in the training data set (red line) and external test data sets C (green line), D (purple line), and E (blue line) show that the model combining clinicopathologic variables, conventional radiomics, and the intratumoral heterogeneity index achieved moderate performance in all patient subgroups. AUC = area under the receiver operating characteristic curve.

human epidermal growth factor receptor 2–positive tumor, low binary ITH index, low continuous ITH index, and high C-radiomics score tend to have pCR after NAC treatment. Figure 6 shows two patients who had similar clinical characteristics but experienced different NAC treatment outcomes. Patient A had a low ITH index and pCR after NAC treatment, whereas patient B had a high ITH index and did not have pCR.

Discussion

The predictive model using intratumoral ecological diversity features showed good performance for predicting pathologic complete response (pCR) to neoadjuvant chemotherapy (NAC) in breast cancer in the three external test sets (area under the receiver operating characteristic curve [AUC] range, 0.74–0.76). The highest performance for predicting pCR to NAC in the external test sets was achieved when intratumoral ecological diversity features were combined with clinicopathologic variables and C-radiomics (AUC range, 0.83–0.87).

Several previous studies have used imaging features and clinicopathologic variables to construct models for predicting pCR to NAC. One pCR prediction model developed by Liu et al (12) integrated a multiparametric MRI signature and clinicopathologic variables, which achieved AUCs ranging 0.71–0.80 in the three external test data sets. Peng et al (25) used radiomics and deep learning techniques to extract imaging features from MRI and combined these features with kinetic and molecular information. In a validation data set, the radiomics-based model and deep learning–based model achieved AUCs of 0.78 and 0.83, respectively, for predicting pCR. The developed clinicopath-radiomics model in the current study achieved AUCs of 0.81, 0.78, and 0.78 in the external test data sets, which are comparable to the previous studies. When the ITH index was added to this model, the resulting combined model showed the highest performance of all models tested, achieving AUCs of 0.87, 0.85, and 0.83 in the three external test data sets. Additionally, decision curve analysis demonstrated that the combined model could achieve higher net benefit than the clinicopath model and clinicopath-radiomics model in the three external test data sets in a wide range of potential thresholds, indicating that adding the ITH index to prediction models for pCR could be clinically useful.

Despite the known association between ITH and treatment response (7,8,26), previous quantitative imaging analysis

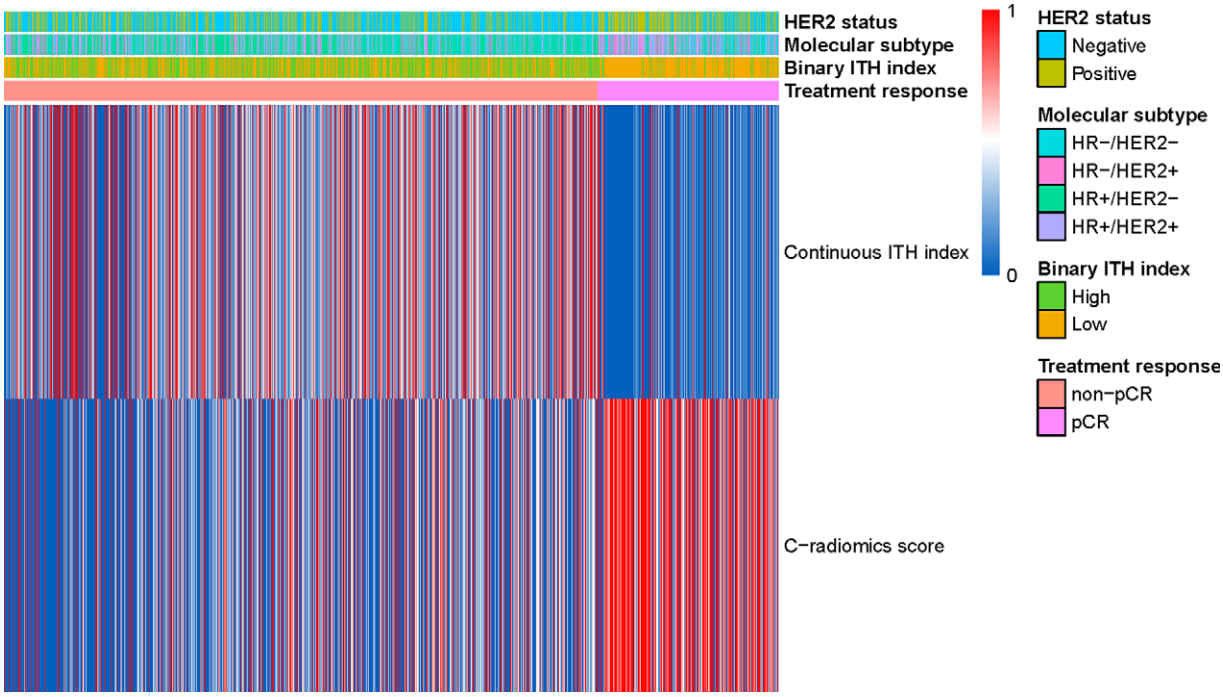

Figure 5: Heat map shows the association of pathologic complete response (pCR) of breast cancer to neoadjuvant chemotherapy (NAC; responder vs nonresponder) with clinicopathologic variables (human epidermal growth factor receptor 2 [HER2] status, molecular subtype) and MRI features (binary intratumoral heterogeneity [ITH] index, continuous ITH index, conventional radiomics [C-radiomics] score) in all data sets. Patients with positive human epidermal growth factor receptor 2 status, hormone receptor (HR)–negative and human epidermal growth factor receptor 2–positive molecular subtype, low binary ITH index or low continuous ITH index values, or high C-radiomics score tend to have pCR after NAC treatment. The color bar indicates continuous ITH index and C-radiomics score values, with a minimum of 0 and maximum of 1.

approaches have not fully assessed the relationships among C-radiomics features, ITH, and clinical outcomes. For example, Wu et al (9) characterized intratumoral spatial heterogeneity at perfusion MRI by quantifying multiple spatially distinct subregions of tumors partitioned by four kinetic maps at dynamic contrast-enhanced MRI. However, the transferability of this method to other modalities (eg, CT, PET) is limited. In the current study, we performed a two-step method for intratumor subregion partitioning that included presegmentation and clustering based on C-radiomics features, which can be transferred to other modalities (Tables S6, S7).

Intratumoral ecological diversity features reflecting ITH were proposed in this study to develop the prediction models for treatment response to NAC, which not only showed good feature reproducibility (intraclass correlation coefficient range, 0.87–0.95), but also achieved reasonable discrimination and generalization in the external test data sets (AUC range, 0.74–0.76). Also, the generalization of the proposed combined model (AUC range, 0.83–0.87) was validated in three external test data sets with a large number of patients (n = 1254) scanned with different scanners and imaging protocols, which was comparable to the AUC ranges reported by both Liu et al (12) (0.71–0.80) and Peng et al (25) (0.78–0.83).

This study had some limitations. First, it was a retrospective analysis with potential bias in population selection, although external validation was performed to improve reliability. To investigate the MRI-based quantitative measure of breast tumor heterogeneity, further prospective analysis should be performed. Second, only using the phase with peak tumor enhancement at pretreatment dynamic contrast-enhanced MRI cannot comprehensively quantitate the ITH of breast cancer, which may introduce analysis bias influencing generalizability in clinical application. Finally, manual tumor delineation by different readers could have influenced the stability of radiomic features. Despite our efforts to address this issue by selecting features with intraclass correlation coefficients greater than 0.75, it is imperative to employ an automated and accurate tumor segmentation method, which could not only enhance the efficiency of quantitative imaging analysis but also ensure stability and consistency in the process.

In conclusion, the model developed in this study, which integrated clinicopathologic information, intratumoral ecological diversity features, and C-radiomics features, achieved the best performance for predicting pathologic complete response to neoadjuvant chemotherapy in patients with breast cancer in three external test data sets. Future research should explore the underlying biologic mechanisms of imaging-based methods for quantifying heterogeneity in breast cancer by incorporating multimodal data.

Author contributions: Guarantors of integrity of entire study, **Z.S., X.H., Z.C., X.C., Chunling Liu, C. Liang, J.S.**; 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

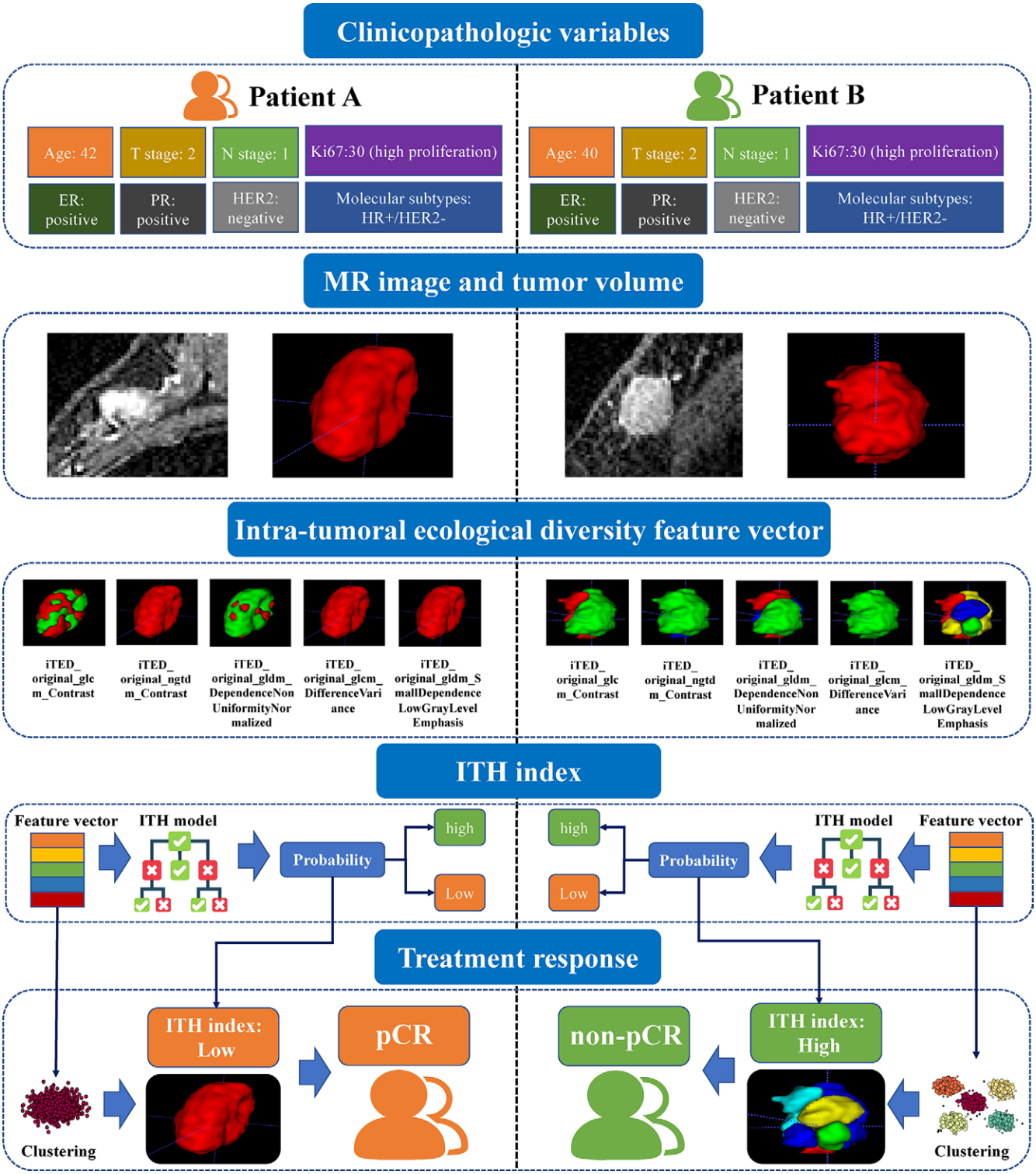

Figure 6: Schematic shows example implementation of the breast cancer intratumoral heterogeneity (ITH) index in two patients with breast cancer who have similar clinical characteristics. Patient A and patient B both had the same clinical T and N stages prior to neoadjuvant chemotherapy (NAC), Ki-67 index category (high proliferation), estrogen receptor (ER) status, progesterone receptor (PR) status, human epidermal growth factor receptor 2 (HER2) status, molecular subtype, and approximate age (42 and 40 years). MRI bounding boxes, including tumor regions, are shown for both patients. The Gaussian mixture model was used to cluster intratumoral subregions color-labeled using the same imaging properties, which leads to an intratumoral ecological diversity (iTED) feature vector for each patient. This feature vector was then used as input to generate a binary ITH index (low vs high). The optimal number of clusters was 1 and 5 for patients A and B, respectively. Patient A had a low ITH index of 0.34 and pathologic complete response (pCR) after NAC treatment. Patient B had a high ITH index of 0.85 and did not experience pCR after NAC treatment. glcm = gray-level co-occurrence matrix, gldm = gray-level dependence matrix, HR = hormone receptor, ngtdm = neighboring gray-tone difference matrix.

work are appropriately resolved, all authors; literature research, **Z.S., X.H., Z.C., Z.X., H.L., X.C., Chunling Liu, C. Liang, Y.C., C.H., J.Q., Z.L.**; clinical studies, **Z.S., X.H., Z.C., Z.X., Chen Liu, X.C., Chunling Liu, J.Q., J.S., Z.L.**; experimental studies, **Z.S., Z.C., X.C., Chunling Liu, Z.L.**; statistical analysis, **Z.S., X.H., X.C., Y.C.**; and manuscript editing, **Z.S., X.H., Z.X., Chen Liu, X.C., C. Liang, C. Lu, J.Q., Z.L.**

Disclosures of conflicts of interest: **Z.S.** No relevant relationships. **X.H.** No relevant relationships. **Z.C.** No relevant relationships. **Z.X.** No relevant relationships. **H.L.** No relevant relationships. **Chen Liu** No relevant relationships. **X.C.** No relevant relationships. **Chunling Liu** No relevant relationships. **C. Liang** No relevant relationships. **C. Lu** No relevant relationships. **Y.C.** No relevant relationships. **C.H.** No relevant relationships. **J.Q.** No relevant relationships. **J.S.** No relevant relationships. **Z.L.** No relevant relationships.

References

1. Korde LA, Somerfield MR, Carey LA, et al. Neoadjuvant Chemotherapy, Endocrine Therapy, and Targeted Therapy for Breast Cancer: ASCO Guideline. J Clin Oncol 2021;39(13):1485–1505.
2. Haque W, Verma V, Hatch S, Suzanne Klimberg V, Brian Butler E, Teh BS. Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. Breast Cancer Res Treat 2018;170(3):559–567.
3. Fayanju OM, Ren Y, Thomas SM, et al. The Clinical Significance of Breast-only and Node-only Pathologic Complete Response (pCR) After Neoadjuvant Chemotherapy (NACT): A Review of 20,000 Breast Cancer Patients in the National Cancer Data Base (NCDB). Ann Surg 2018;268(4):591–601.
4. Romeo V, Accardo G, Perillo T, et al. Assessment and Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer: A Comparison of Imaging Modalities and Future Perspectives. Cancers (Basel) 2021;13(14):3521.
5. Spring LM, Fell G, Arfe A, et al. Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-analysis. Clin Cancer Res 2020;26(12):2838–2848.
6. I-SPY2 Trial Consortium; Yee D, DeMichele AM, et al. Association of Event-Free and Distant Recurrence-Free Survival With Individual-Level Pathologic Complete Response in Neoadjuvant Treatment of Stages 2 and 3 Breast Cancer: Three-Year Follow-up Analysis for the I-SPY2 Adaptively Randomized Clinical Trial. JAMA Oncol 2020;6(9):1355–1362.
7. Zardavas D, Irrthum A, Swanton C, Piccart M. Clinical management of breast cancer heterogeneity. Nat Rev Clin Oncol 2015;12(7):381–394.
8. Lüönd F, Tiede S, Christofori G. Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. Br J Cancer 2021;125(2):164–175.
9. 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.
10. Expert Panel on Breast Imaging; Slanetz PJ, Moy L, et al. ACR Appropriateness Criteria® Monitoring Response to Neoadjuvant Systemic Therapy for Breast Cancer. J Am Coll Radiol 2017;14(11S):S462–S475.
11. Yu Y, He Z, Ouyang J, et al. Magnetic resonance imaging radiomics predicts preoperative axillary lymph node metastasis to support surgical decisions and is associated with tumor microenvironment in invasive breast cancer: A machine learning, multicenter study. EBioMedicine 2021;69:103460.
12. Liu Z, Li Z, Qu J, et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. Clin Cancer Res 2019;25(12):3538–3547.
13. Chitalia RD, Rowland J, McDonald ES, et al. Imaging Phenotypes of Breast Cancer Heterogeneity in Preoperative Breast Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) Scans Predict 10-Year Recurrence. Clin Cancer Res 2020;26(4):862–869.
14. Natrajan R, Sailem H, Mardakheh FK, et al. Microenvironmental Heterogeneity Parallels Breast Cancer Progression: A Histology-Genomic Integration Analysis. PLoS Med 2016;13(2):e1001961.
15. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015;102(3):148–158.
16. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013;26(6):1045–1057.
17. 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):455–463.
18. Li W, Newitt DC, Gibbs J, et al. Predicting breast cancer response to neoadjuvant treatment using multi-feature MRI: results from the I-SPY 2 TRIAL. NPJ Breast Cancer 2020;6(1):63.
19. Sadri AR, Janowczyk A, Zhou R, et al. Technical Note: MRQy - An open-source tool for quality control of MR imaging data. Med Phys 2020;47(12):6029–6038.
20. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 2010;29(6):1310–1320.
21. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. IEEE Trans Med Imaging 2000;19(2):143–150.
22. 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):328–338.
23. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77(21):e104–e107.
24. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Programs Biomed 2022;214:106584.
25. Peng Y, Cheng Z, Gong C, et al. Pretreatment DCE-MRI-Based Deep Learning Outperforms Radiomics Analysis in Predicting Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer. Front Oncol 2022;12:846775.
26. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol 2018;15(2):81–94.