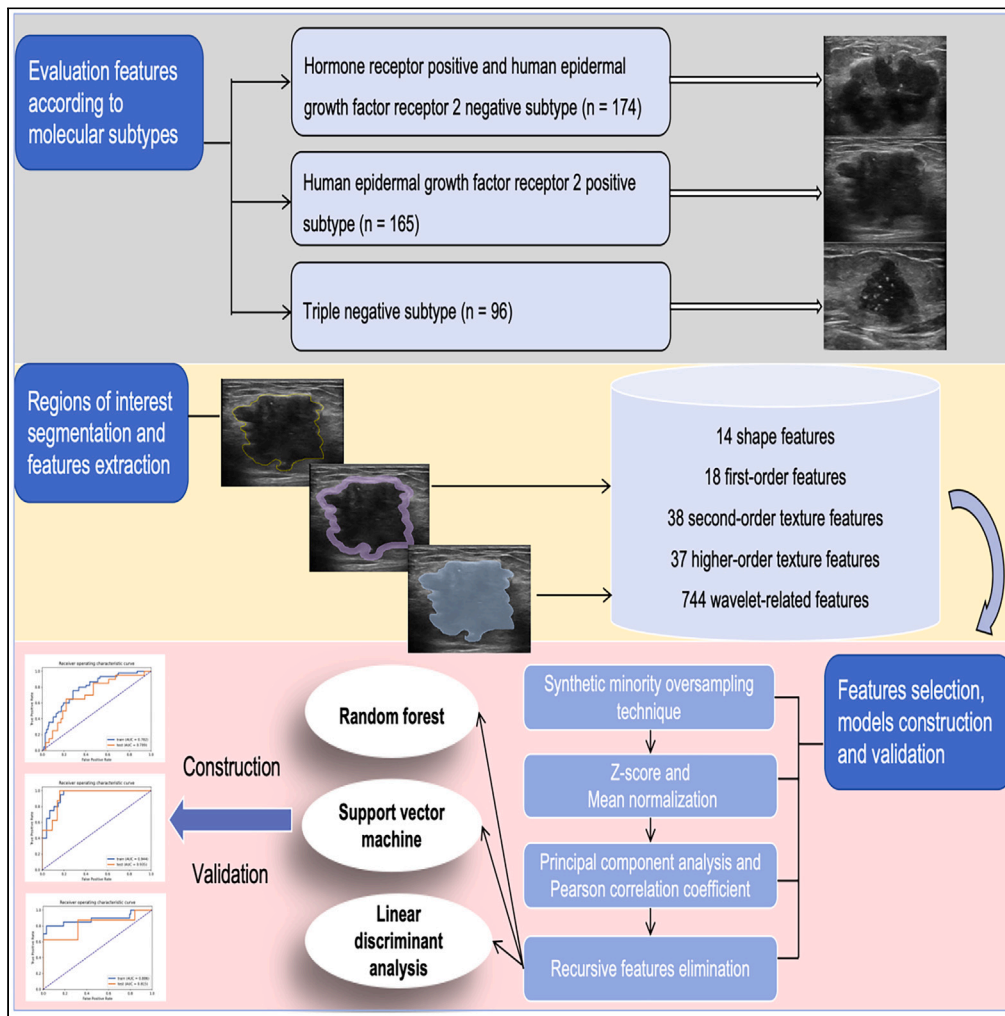**Article**

Predicting axillary response to neoadjuvant chemotherapy using peritumoral and intratumoral ultrasound radiomics in breast cancer subtypes

Jiejie Yao,
Xiaohong Jia, Wei
Zhou, Ying Zhu,
Xiaosong Chen,
Weiwei Zhan,
Jianqiao Zhou

zhousu30@126.com

Highlights

Axillary response to NAC is
an important indicator for
predicting prognosis

This study constructed
PURS, IURS, and the
combined P-IURS
predictive models

The RF-based combined
P-IURS model in the HER2+
subtype achieved the
highest performance

ML-based US radiomics
can be a promising
biomarker to help clinical
strategies

Article

Predicting axillary response to neoadjuvant chemotherapy using peritumoral and intratumoral ultrasound radiomics in breast cancer subtypes

Jiejie Yao,[1,3] Xiaohong Jia,[1,3] Wei Zhou,[1,3] Ying Zhu,[1] Xiaosong Chen,[2] Weiwei Zhan,[1] and Jianqiao Zhou[1,4,*]

SUMMARY

To explore machine learning (ML)-based breast tumor peritumoral (P) and intratumoral ultrasound radiomics signatures (IURS) for predicting axillary response to neoadjuvant chemotherapy (NAC) in patients with breast cancer (BC) with node-positive. A total of 435 patients were divided into hormone receptor (HR)+/human epidermal growth factor receptor (HER)2-, HER2+, and triple-negative (TN) subtypes. ML classifiers including random forest (RF), support vector machine (SVM), and linear discriminant analysis (LDA) were applied to construct PURS, IURS, and the combined P-IURS radiomics models. SVM of the TN subtype obtained the most favorable performance with an AUC of 0.917 (95%CI: 0.859, 0.960) in PURS models, RF of the HER2+ subtype yielded the highest efficacy in IURS models [AUC = 0.935 (95% CI: 0.843, 0.976)]. The RF-based combined P-IURS model of the HER2+ subtype improved the efficacy to a maximum AUC of 0.952 (95%CI: 0.868, 0.994). ML-based US radiomics can be a promising biomarker to predict axillary response.

INTRODUCTION

For patients with breast cancer (BC) with node-positive, neoadjuvant chemotherapy (NAC) is the standard therapeutic option to reduce tumor burden, eliminate lymph node metastasis, and improve the probability of breast and axillary conservation surgery.[1] Axillary nodal pathological complete response (pCR) is associated with excellent local regional and survival outcomes.[1,2] Previous studies had indicated that axillary pCR was a more important indicator as compared with primary tumor response for predicting prognosis.[1–3] Ideally, for patients who are expected to achieve nodal pCR, axillary lymph node dissection (ALND) can be spared, and minimally invasive approaches are recommended.[4,5] However, in clinical practice, ALND was still regarded as the standard procedure for patients with node-positive BC after NAC due to the high false-negative results of sentinel lymph node biopsy (SLNB) reported in previous studies.[6,7] ALND is notable for its morbidity and complications such as pain, lymphedema, shoulder dysfunction, and hypesthesia.[8] Accurate and non-invasive methods to predict axillary pCR could assist clinicians in stratifying patients from the avoidance of over-treatment axillary surgery.

Previous studies have used clinical and pathological factors,[2] breast MRI,[9,10] or US[10] to predict axillary response after NAC. Recent researchers applied deep learning US radiomics for the prediction of treatment response.[11] However, most of the previous studies mainly focused on the tumor heterogeneity of intratumoral structure, the considerations about the peritumoral region which contained peripheral tumor lymphatic, micro-vascular infiltration and marginal inflammatory factors were limited. Moreover, to our knowledge, no prior study had applied peritumoral US radiomics signatures (PURS) of breast tumors for the prediction of axillary pCR. So, can PURS be utilized to predict axillary response after NAC? In addition, as a high heterogeneity disease, BC is well known for its diverse response to NAC according to different molecular subtypes. For example, the introduction of HER-targeted drugs (i.e., trastuzumab and pertuzumab) improved the rate of tumor pCR of patients with HER2+ than patients with HR+/HER2-and TNBC.[12,13] However, what are the rates of axillary pCR for patients with BC with node-positive in different biological subtypes?

Machine learning (ML) algorithms such as random forest (RF), support vector machine (SVM), and linear discriminant analysis (LDA) have been introduced for the analysis of medical images, and can effectively generate quantitative biomarkers.[14,15] A prior study of our team has applied four ML classifiers-based US radiomics for the preoperative prediction of axillary sentinel lymph node metastasis burden in patients with early-stage BC.[16] However, what are the efficacy of various ML-based PURS and IURS for the prediction of axillary nodal response? Moreover, it is curious about the predictive performance of the combination of PURS and IURS (P-IURS) models.

Table 1. Clinicopathologic data in relation to axillary pCR in molecular subtypes

Characteristics	HR+/HER2- (n = 174) No. (%)			HER2+ (n = 165) No. (%)			TN (n = 96) No. (%)		
	(n = 28) pCR	(n = 146) Non-pCR	p value	(n = 86) pCR	(n = 79) Non-pCR	p value	(n = 40) pCR	(n = 56) Non-pCR	p value
Age, years	48.0 ± 11.7	49.8 ± 13.8	0.516	47.5 ± 12.5	50.9 ± 14.2	0.497	48.2 ± 12.9	51.1 ± 14.7	0.588
Tumor size (mm)	37.2 ± 11.3	39.5 ± 12.7	0.715	36.8 ± 11.5	38.7 ± 13.3	0.681	35.9 ± 12.2	39.0 ± 14.1	0.694
Family history of breast cancer			0.330			0.493			0.663
No	20 (71.4)	116 (79.5)		60 (69.8)	59 (74.7)		26 (65.0)	39 (69.6)	
Yes	8 (28.6)	30 (20.5)		26 (30.2)	20 (25.3)		14 (35.0)	17 (30.4)	
Symptom			0.510			0.263			0.531
Palpable mass	17 (60.7)	100 (68.5)		50 (58.1)	53 (67.1)		21 (52.5)	34 (60.7)	
others	11 (39.3)	46 (31.5)		36 (41.9)	26 (32.9)		19 (47.5)	22 (39.3)	
Tumor location			0.652			0.424			0.675
Outer upper quadrant	13 (46.4)	58 (39.7)		35 (40.7)	27 (34.2)		17 (42.5)	21 (37.5)	
others	15 (53.6)	88 (60.3)		51 (59.3)	52 (65.8)		23 (57.5)	35 (62.5)	
Clinical T stage			0.473			0.08			0.229
I	2 (7.1)	5 (3.4)		7 (8.1)	2 (2.5)		3 (7.5)	2 (3.6)	
II	15 (53.6)	69 (47.3)		48 (55.8)	37 (46.9)		23 (57.5)	25 (44.6)	
III	11 (39.3)	72 (49.3)		31 (36.1)	40 (50.6)		14 (35.0)	29 (51.8)	
Histologic type			0.776			0.679			0.575
Invasive ductal carcinoma	23 (82.1)	124 (84.9)		73 (84.9)	65 (82.3)		35 (87.5)	46 (82.1)	
Others	5 (17.9)	22 (15.1)		13 (15.1)	14 (17.7)		5 (12.5)	10 (17.9)	
Tumor grade			0.252			0.121			0.437
Low/medium	5 (17.9)	44 (30.1)		13 (15.1)	20 (25.3)		6 (15.0)	13 (23.2)	
High	23 (82.1)	102 (69.9)		73 (84.9)	59 (74.7)		34 (85.0)	43 (76.8)	
Ki67 levels			0.644			0.566			0.458
≤20%	6 (21.4)	39 (26.7)		16 (18.6)	18 (22.8)		7 (17.5)	14 (25.0)	
>20%	22 (78.6)	107 (73.3)		70 (81.4)	61 (77.2)		33 (82.5)	42 (75.0)	

The analyses of clinicopathologic data found no significant difference between axillary pCR and non-pCR groups in HR+/HER2-, HER2+, and TN subtypes, p value >0.05 for all.

Based on the above questions, this study aimed to construct various ML classifiers-based PURS, IURS, and the combined P-IURS models in different molecular subtypes, and explore the value of these novel approaches, for the early prediction of axillary pCR after NAC in patients with node-positive.

RESULTS

Clinicopathologic characteristics

The axillary pCR rate was 35.4% (154/435) in all patients. Among the subtypes, the axillary pCR rate was higher in patients with HER2+ [52.2% (60/115) and 52.0% (26/50)] than in patients with TN [40.3% (27/67) and 44.8% (13/29)], and patients with HR+/HER2-[16.5% (20/121) and 15.1% (8/53)] in the training and test set, respectively. However, no significant difference was found in terms of age, tumor maximum size, family history of BC, symptom, tumor location, clinical T stage, histologic type, tumor grade, and Ki-67 levels between the axillary pCR and non-pCR groups in three molecular subtypes (p > 0.05 for all) (Table 1).

Machine learning-based peritumoral US radiomics signatures, intratumoral ultrasound radiomics signatures, and peritumoral US radiomics signatures-intratumoral ultrasound radiomics signatures models according to molecular subtypes

Predictive performance in the hormone receptor+/human epidermal growth factor receptor 2-subtype

Figures 1–3 show the study design, the examples of peritumoral and intratumoral ROIs, and the procedure of the workflow, respectively. For the PURS models, the RF classifier achieved a better AUC of 0.838 (95%CI: 0.761, 0.897), than SVM [AUC = 0.781 (95%CI: 0.711, 0.850], and LDA [AUC = 0.748 (95%CI: 0.669, 0.825)] in the training set. In the test set, the AUCs were 0.823 (95%CI: 0.708, 0.916) for RF, 0.757 (95%CI: 0.596, 0.870) for SVM, and 0.681 (95%CI: 0.500, 0.805) for LDA. The RF classifier yielded a SEN of 76.6%, SPE of 82.6%, ACC of 80.9%, PPV of 63.7%,
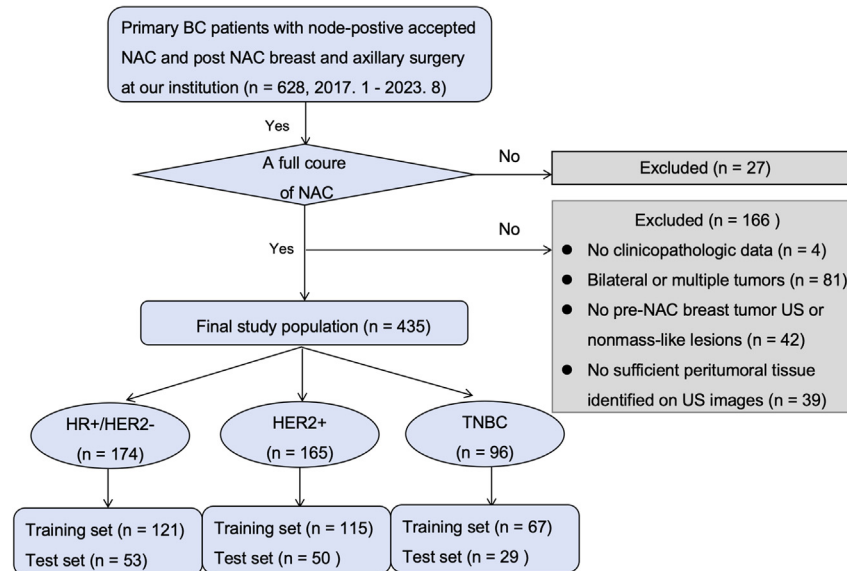
**Figure 1. Flowchart shows patient recruitment and study design**

and NPV of 91.7% in the training set, and a SEN of 85.0%, SPE of 84.3%, ACC of 84.2%, PPV of 62.2%, and NPV of 94.0% in the test set, respectively (Table 2).

For the IURS models, the RF classifier obtained a higher AUC of 0.847 (95%CI: 0.789, 0.906), compared with SVM [AUC = 0.823 (95%CI: 0.759, 0.882)], and LDA [AUC = 0.782 (95%CI: 0.701, 0.856)] in the training set. In the test set, the AUCs were 0.851 (95%CI: 0.732, 0.918)
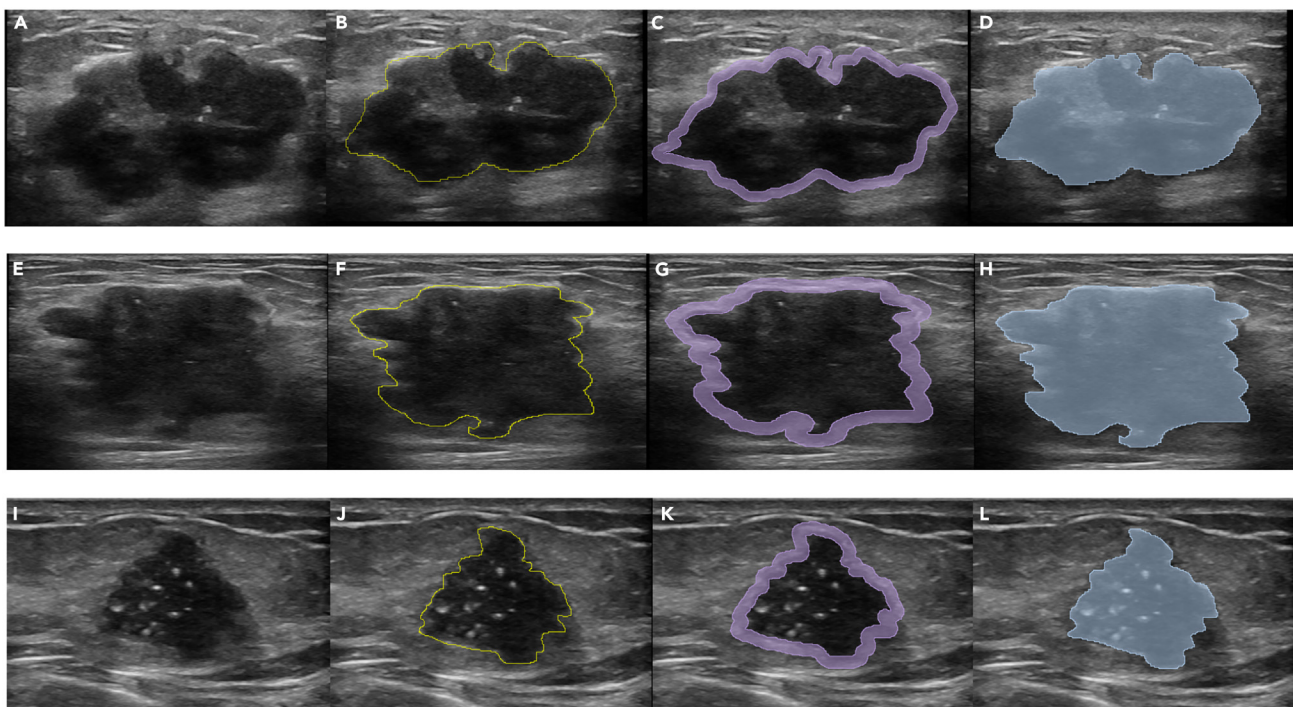
Figure 2. Examples of the regions of interest (ROIs) segmentation in different molecular subtypes
(A) The baseline grayscale US image of a 61-year-old woman with the right axillary node-positive and HR+/HER2-subtype invasive ductal carcinoma in the right breast, attained axillary non-pCR after NAC, (E) the baseline grayscale US image of a 49-year-old woman with the right axillary node-positive and HER2+ subtype invasive ductal carcinoma in the right breast yielded axillary pCR after NAC, (I) The baseline grayscale US image of a 44-year-old woman with the left axillary node-positive and TN subtype invasive ductal carcinoma in the left breast achieved axillary pCR after NAC, (B, F, and J) the corresponding ROIs were manually delineated along the contour of the tumor, (C, G, and K) the ROIs of peritumoral regions (purple), and (D, H, and L) the ROIs of intratumoral regions (blue).
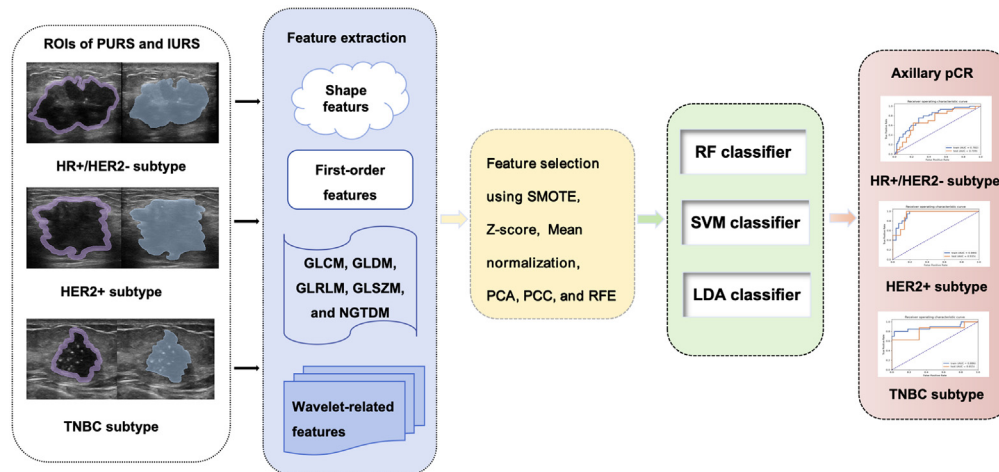
Figure 3. The overview of the workflow

(1) The peritumoral and intratumoral ROIs of 435 breast tumors according to the molecular subtype.

(2) Features extraction including shape features, first-order features, texture features (i.e., GLCM, GLDM, GLRLM, GLSZM, and NGTDM), and wavelet-related features.

(3) Features selection by using SMOTE, *Z* score, Mean normalization, PCA, PCC, and RFE methods.

(4) RF, SVM, and LDA classifiers were applied to construct PURS, IURS, and the combined P-IURS radiomics models for the prediction of axillary pCR after NAC.

(5) The performance of predictive models in three subtypes.

for RF, 0.802 (95%CI: 0.698, 0.866) for SVM, and 0.709 (95%CI: 0.591, 0.818) for LDA. The RF classifier also obtained a SEN of 77.7%, SPE of 83.1%, ACC of 81.8%, PPV of 61.2%, and NPV of 89.9% in the training set, and a SEN of 83.3%, SPE of 86.2%, ACC of 84.8%, PPV of 69.2%, and NPV of 93.8% in the test set, respectively (Table 2).

For the combined P-IURS models, in the training set, the AUCs were 0.852 (95%CI: 0.776, 0.890) for RF, 0.864 (95%CI: 0.790, 0.908) for SVM, and 0.859 (95%CI: 0.784, 0.901) for LDA. In the test set, the RF classifier yielded a better AUC of 0.857 (95%CI: 0.726–0.903) than SVM [AUC = 0.819 (95%CI: 0.700, 0.875)], and LDA [AUC = 0.765 (95%CI: 0.601, 0.859)] (Figures 4A–4I). The RF classifier achieved a SEN of 76.9%, SPE of 82.9%, ACC of 80.6%, PPV of 62.3%, and NPV of 89.4% in the training set, and a SEN of 83.7%, SPE of 85.4%, ACC of 85.0%, PPV of 67.4%, and NPV of 92.9% in the test set, respectively (Table 2). Two, 3, and 4 optimal radiomics features were selected for the RF classifier-based PURS, IURS, and P-IURS models, respectively. The details and coefficients of the selected features are shown in Table 3.

### Predictive performance in the human epidermal growth factor receptor 2+ subtype

For the PURS models, the SVM classifier yielded a favorable AUC of 0.888 (95%CI: 0.834, 0.936), compared with RF [AUC = 0.873 (95%CI: 0.819, 0.923)], and LDA [AUC = 0.804 (95%CI: 0.739, 0.871)]) in the training set. However, in the test set, the AUCs were 0.867 (95%CI: 0.772, 0.952) for RF, 0.816 (95%CI: 0.697, 0.933) for SVM, and 0.781 (95%CI: 0.644, 0.901) for LDA (Figures 5A–5C). The SVM classifier yielded a SEN of 78.9%, SPE of 86.0%, ACC of 86.7%, PPV of 66.3%, and NPV of 93.1% in the training set, and the RF classifier obtained a SEN of 86.5%, SPE of 88.9%, ACC of 86.5%, PPV of 70.1%, and NPV of 95.5% in the test set, respectively (Table 4).

For the IURS models, the RF classifier achieved a better AUC of 0.944 (95%CI: 0.868, 0.990), than SVM [AUC = 0.882 (95%CI: 0.826, 0.937)], and LDA [AUC = 0.819 (95%CI: 0.744, 0.873)] in the training set. In the test set, the AUCs were 0.935 (95%CI: 0.843, 0.976) for RF, 0.883 (95%CI: 0.779, 0.944) for SVM, and 0.824 (95%CI: 0.712, 0.890) for LDA (Figures 5D–5F). The RF classifier also obtained a favorable SEN of 95.0%, SPE of 90.1%, ACC of 92.6%, PPV of 73.3%, and NPV of 98.9% in the training set, and a SEN of 87.5%, SPE of 93.1%, ACC of 92.3%, PPV of 70.0%, and NPV of 96.7% in the test set, respectively (Table 4).

For the combined P-IURS models, in the training set, the AUCs were 0.954 (95%CI: 0.870, 0.992) for RF, 0.933 (95%CI: 0.857, 0.974) for SVM, and 0.852 (95%CI: 0.761, 0.880) for LDA classifier. In the test set, the RF classifier also obtained a higher AUC of 0.952 (95%CI: 0.868–0.994) compared with SVM [AUC = 0.906 (95%CI: 0.801, 0.957)], and LDA [AUC = 0.857 (95%CI: 0.771, 0.892)] (Figures 5G–5I). The RF classifier achieved a substantial SEN of 95.5%, SPE of 91.2%, ACC of 93.0%, PPV of 74.8%, and NPV of 98.7% in the training set, and a SEN of 88.8%, SPE of 95.0%, ACC of 93.4%, PPV of 73.6%, and NPV of 98.5% in the test set, respectively (Table 4). Five, 9, and 10 optimal features were selected for the RF classifier-based PURS, IURS, and P-IURS models, respectively. The details and coefficients of the selected features are shown in Table 5.

Predictive performance in the triple-negative subtype

For the PURS models, the SVM classifier yielded a better predictive efficacy with an AUC of 0.928 (95%CI: 0.880, 0.964), compared with RF [AUC = 0.886 (95%CI: 0.831, 0.939)], and LDA [AUC = 0.876 (95%CI: 0.828, 0.926)] in the training set. In the test set, the AUCs were 0.917
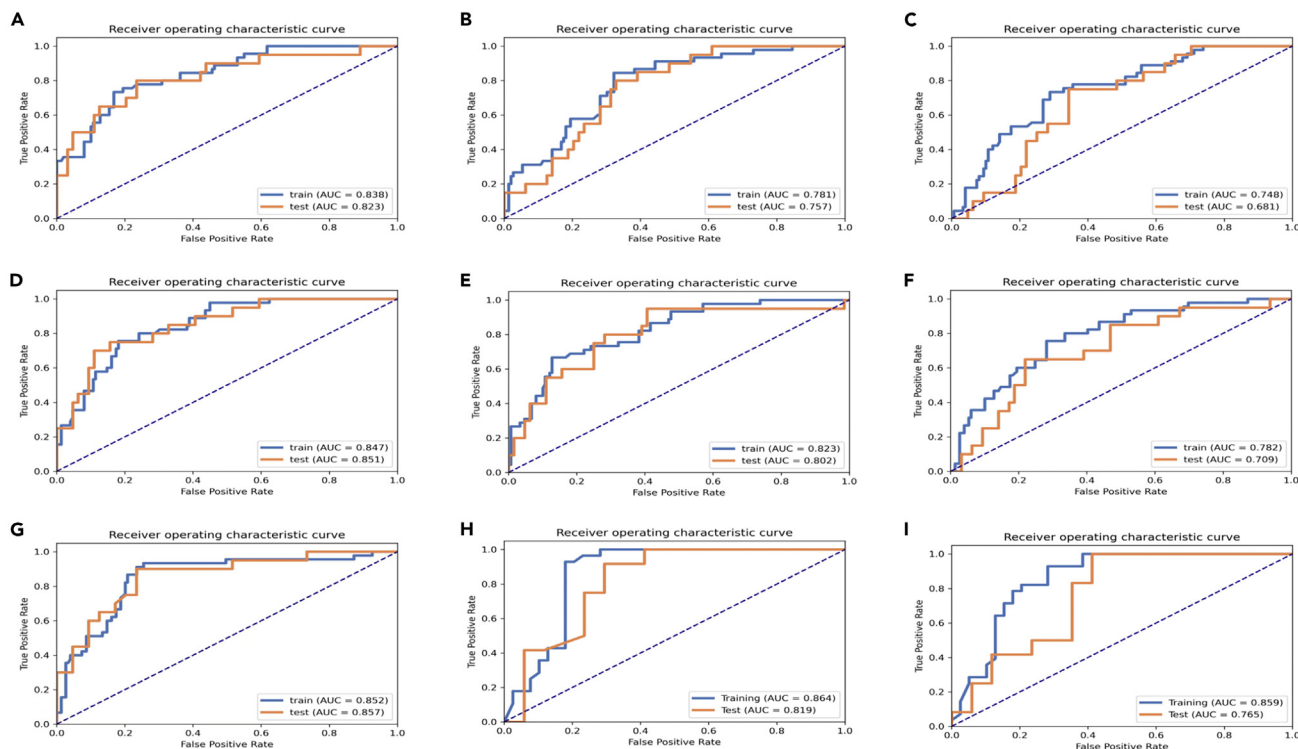
Figure 4. The ROC curves in the HR+/HER2-subtype

The PURS model with RF (A), SVM (B), and LDA (C) classifiers. The AUCs for the three classifiers were 0.838, 0.781, and 0.748 in the training set, and 0.823, 0.757, and 0.681 in the test set, respectively. The IURS model with RF (D), SVM (E), and LDA (F) classifiers. The AUCs for the three classifiers were 0.847, 0.823, and 0.782 in the training set, and 0.851, 0.802, and 0.709 in the test set, respectively. The combined P-IURS model with RF (G), SVM (H), and LDA (I) classifiers. The AUCs for the three classifiers were 0.852, 0.864, and 0.859 in the training set, and 0.857, 0.819, and 0.765 in the test set, respectively.

(95%CI: 0.859, 0.960) for SVM, 0.815 (95%CI: 0.694, 0.920) for RF, and 0.773 (95%CI: 0.664, 0.901) for LDA (Figures 6A–6C). The SVM classifier obtained a satisfactory SEN of 93.5%, SPE of 81.2%, ACC of 85.3%, PPV of 70.8%, and NPV of 96.1% in the training set, and SEN of 90.5%, SPE of 85.5%, ACC of 88.2%, PPV of 73.1%, and NPV of 95.9% in the test set, respectively (Table 6).

For the IURS models, the SVM classifier yielded a better AUC of 0.903 (95%CI: 0.855, 0.962), than RF [AUC = 0.875 (95%CI: 0.820, 0.923)], and LDA [AUC = 0.825 [95%CI: 0.760, 0.891]] in the training set. In the test set, the AUCs were 0.866 (95%CI: 0.739, 0.943) for SVM, 0.789 (95% CI: 0.701, 0.887) for RF, and 0.750 (95%CI: 0.637, 0.842) for LDA (Figures 6D–6F). The SVM classifier achieved an acceptable SEN of 85.7%, SPE of 89.7%, ACC of 88.8%, PPV of 85.7%, and NPV of 90.1% in the training set, and a SEN of 81.6%, SPE of 84.3%, ACC of 82.2%, PPV of 68.6%, and NPV of 92.3% in the test set, respectively (Table 6).

For the combined P-IURS models, in the training set, the AUCs were 0.873 (95%CI: 0.818, 0.925) for RF, 0.923 (95%CI: 0.866, 0.971) for SVM, and 0.886 (95%CI: 0.803, 0.939) for LDA. In the test set, the SVM classifier achieved a more satisfactory AUC of 0.934 (95%CI: 0.877, 0.983) than RF [AUC = 0.918 (95%CI: 0.862, 0.970)], and LDA [AUC = 0.849 (95%CI: 0.745, 0.900)] (Figures 6G–6I). The SVM classifier achieved a substantial SEN of 88.9%, SPE of 91.2%, ACC of 90.4%, PPV of 79.5%, and NPV of 93.1% in the training set, and a SEN of 91.4%, SPE of 85.7%, ACC of 90.2%, PPV of 75.6%, and NPV of 95.9% in the test set, respectively (Table 6). Six, 4, and 6 optimal radiomics features were selected for the SVM classifier-based PURS, IURS, and P-IURS models, respectively (Table 7).

Comparison of predictive models

Delong test showed that among the PURS models, SVM of the TN subtype obtained the highest predictive performance than RF of the HER2+ subtype (AUC of 0.917 vs. 0.867, z = 2.581, p < 0.05), and RF of the HR+/HER2-subtype (AUC of 0.917 vs. 0.823, z = 5.243, p < 0.001). Among the IURS models, RF of the HER2+ subtype achieved the best ability than the SVM of the TN subtype (AUC of 0.935 vs. 0.866, z = 4.447, p < 0.001), and RF of the HR+/HER2-subtype (AUC of 0.935 vs. 0.851, z = 4.890, p < 0.001). Among the combined P-IURS models, RF of the HER2+ subtype achieved the most favorable efficacy than SVM of the TN subtype (AUC of 0.952 vs. 0.934, z = 2.073, p < 0.05), and RF of the HR+/HER2-subtype (AUC of 0.952 vs. 0.857, z = 5.103, p < 0.001) for the prediction of axillary pCR in the test sets. Encouragingly, the RF-based combined P-IURS model in the HER2+ subtype improved the performance of RF-based IURS and yielded the highest predictive efficacy (AUC of 0.952 vs. 0.935, z = 2.003, p < 0.05). Meanwhile, the LDA-based PURS in the HR+/HER2-subtype obtained the lowest ability (AUC of 0.681, 95%CI: 0.500, 0.805) to predict axillary response in the test sets among all radiomics models.

Table 2. The predictive performance of ML classifiers-based PURS, IURS and the combined P-IURS models in the HR+/HER2-subtype in the training and test sets

Training set	SEN (%)	SPE (%)	ACC (%)	PPV (%)	NPV (%)	AUC (95% CI)
PURS model						
RF	76.6	82.6	80.9	63.7	91.7	0.838 (0.761–0.897)
SVM	74.5	71.8	72.7	47.5	90.6	0.781 (0.711–0.850)
LDA	73.3	71.2	71.5	43.4	89.3	0.748 (0.669–0.825)
IURS model						
RF	77.7	83.1	81.8	61.2	89.9	0.847 (0.789–0.906)
SVM	73.8	82.0	81.0	56.7	90.7	0.823 (0.759–0.882)
LDA	72.3	77.1	76.2	50.2	90.3	0.782 (0.701–0.856)
P-IURS model						
RF	76.9	82.9	80.6	62.3	89.4	0.852 (0.776–0.890)
SVM	77.6	83.5	78.2	60.1	90.1	0.864 (0.790–0.908)
LDA	77.5	79.8	81.0	59.5	88.5	0.859 (0.784–0.901)
Test set						
PURS model						
RF	85.0	84.3	84.2	62.1	94.0	0.823 (0.708–0.916)
SVM	80.0	67.9	70.2	50.0	91.8	0.757 (0.596–0.870)
LDA	62.5	72.7	71.1	49.4	91.4	0.681 (0.500–0.805)
IURS model						
RF	83.3	86.2	84.8	69.2	93.8	0.851 (0.732–0.918)
SVM	72.0	80.1	76.8	53.6	90.1	0.802 (0.698–0.866)
LDA	65.0	78.1	74.5	48.8	87.7	0.709 (0.591–0.818)
P-IURS model						
RF	83.7	85.4	85.0	67.4	92.9	0.857 (0.726–0.903)
SVM	72.4	79.8	78.5	54.8	88.9	0.819 (0.700–0.875)
LDA	70.3	80.0	75.8	51.3	89.0	0.765 (0.601–0.859)

Table 2 showed that in the HR+/HER2-subtype, the RF-based PURS, IURS, and the combined P-IURS models achieved better performance with AUCs of 0.823, 0.851, and 0.857, while the LDA-based PURS obtained low predictive ability with AUC of 0.681 in the test set.

DISCUSSION

Axillary pCR has been identified as a more vital prognostic indicator associated with improved overall survival in comparison with breast tumor pCR.[1–3,17] Accurate and non-invasive prediction of nodal pCR can stratify patients from less aggressive axillary procedures.[4,5] Previous studies used clinicopathologic risk factors, traditional medical images, and radiomics features to predict axillary pCR.[9–11,18,19] Recently, researchers began to extract both intratumoral and peritumoral radiomic features from DCE-MRI, or contrast-enhanced spectral mammography to predict breast pCR.[20,21] However, whether peritumoral US radiomics features can be applied to predict axillary response exists doubts. Moreover, whether the rates of axillary pCR depend on different molecular subtypes remains unknown. The present study is the first attempt to assess the efficacy of various ML classifiers-based PURS and IURS to predict nodal response after NAC. Additionally, we also investigate the performance of the combined P-IURS models according to molecular subtypes. Our results showed that the RF-based P-IURS of the HER2+ subtype achieved the highest predictive ability (AUC of 0.952, 95%CI: 0.868, 0.994), while the LDA-based PURS of the HR+/HER2-subtype obtained the lowest predictive ability (AUC of 0.681, 95%CI: 0.500, 0.805) in the test set.

With the breakthrough improvement of HER2-targeted drugs such as trastuzumab and pertuzumab, the HER2+ subtype has been regarded as a good prognostic factor of NAC.[12,13] In contrast, the absence of HER2+ and the presence of HR + are poor predictors for treatment response.[9,10] A meta-analysis including 33 studies with 57531 patients from Samiei et al.[1] indicated that the HR-/ERBB2+ subtype was associated with the highest axillary pCR rate (60%), followed by 48% for TN and 18% for HR+/ERBB2-. In agree with previous studies, our results displayed that axillary pCR occurred most commonly in the HER2+ subtype and achieved the lowest rate in the HR+/HER2-subtype. The results may be explained by the homology and the same nature between the axillary metastatic nodes and the primary breast tumor. Regarding other clinicopathologic factors, Kantor et al.[18] reported that younger age, high grade, ductal histology, and the extent of breast response were significant independent predictors for nodal pCR. Vila et al.[2] revealed that high nuclear grade and higher Ki-67 levels were more likely to achieve nodal pCR. Regrettably, our results indicated that patient's age, tumor maximum size, family history of BC, symptom,

Table 3. Coefficients of selected features in the HR+/HER2-subtype

Features	Coef in model
RF-based PURS model	
Wavelet-LHH_GLDM_LargeDependenceLowGrayLevelEmphasis	−2.181
Wavelet-LHL_GLRLM_ShortRunEntropy	−1.467
RF-based IURS model	
Wavelet-LHL_GLCM_Imc2	−0.811
Wavelet-LHL_GLRLM_RunLengthNonUniformityNormalized	0.932
Wavelet-HLH_GLSZM_HighGrayLevelZoneEmphasis	−2.167
RF-based combined P-IURS model	
Wavelet-LHL_GLCM_ClusterEndency	−0.418
Wavelet-HLH_GLSZM_HighGrayLevelZoneEmphasis	1.097
Wavelet-LHH_GLDM_LargeDependenceLowGrayLevelEmphasis	0.536
Wavelet-HHL_GLRLM_ShortRunEmphasis	−1.170

There were 2, 3, and 4 optimal features selected in RF-based PURS, IURS, and combined P-IURS models in the HR+/HER2-subtype.

tumor location, clinical T stage, histologic type, tumor grade and Ki-67 levels all had no predictive value for axillary pCR in three molecular subtypes. The discrepancies may be due to the difference in database and sampling selection, a large series of multicenter external patients are needed.

Radiomics provides potential biomarkers for the prediction of clinical outcomes through thousands of high-dimensional data extracted from traditional medical images.[11,22] Previous studies have demonstrated that radiomics features could comprehensively reflect tumor
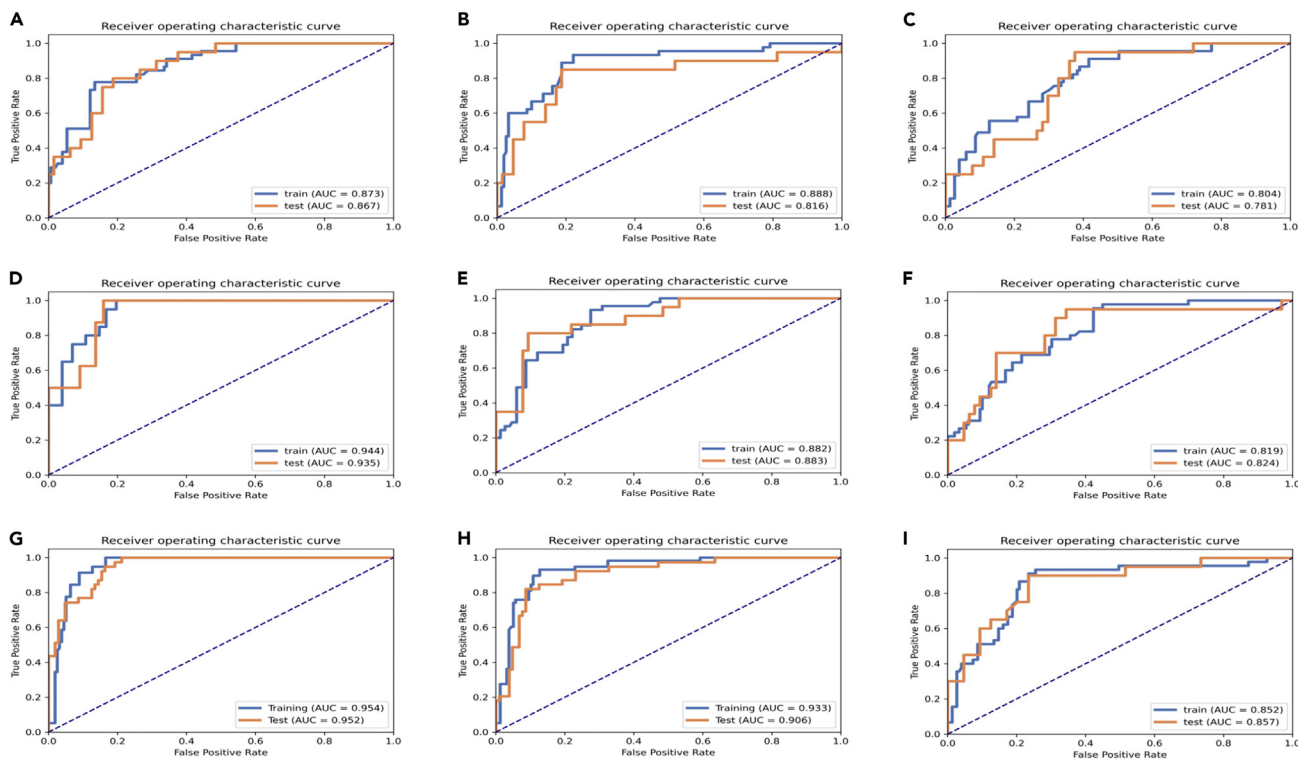
Figure 5. The ROC curves in the HER2+ subtype

The PURS model with RF (A), SVM (B), and LDA (C) classifiers. The AUCs for the three classifiers were 0.873, 0.888, and 0.804 in the training set, and 0.867, 0.816, and 0.781 in the test set, respectively. The IURS model with RF (D), SVM (E), and LDA (F) classifiers. The AUCs for the three classifiers were 0.944, 0.882, and 0.819 in the training set, and 0.935, 0.883, and 0.824 in the test set, respectively. The combined P-IURS model with RF (G), SVM (H), and LDA (I) classifiers. The AUCs for the three classifiers were 0.954, 0.933, and 0.852 in the training set, and 0.952, 0.906, and 0.857 in the test set, respectively.

Table 4. The predictive performance of ML classifiers-based PURS, IURS, and the combined P-IURS models in the HER2+ subtype in the training and test sets

Training set	SEN (%)	SPE (%)	ACC (%)	PPV (%)	NPV (%)	AUC (95% CI)
PURS model						
RF	77.8	86.5	84.5	65.2	92.8	0.873 (0.819–0.923)
SVM	78.9	86.0	86.7	66.3	93.1	0.888 (0.834–0.936)
LDA	73.3	75.6	74.2	48.8	90.2	0.804 (0.739–0.871)
IURS model						
RF	95.0	90.1	92.6	73.3	98.9	0.944 (0.868–0.990)
SVM	83.3	89.1	86.8	68.0	92.4	0.882 (0.826–0.937)
LDA	73.2	80.0	77.1	55.8	90.5	0.819 (0.744–0.873)
P-IURS model						
RF	95.5	91.2	93.0	74.8	98.7	0.954 (0.870–0.992)
SVM	90.3	88.5	89.9	74.0	96.4	0.933 (0.857–0.974)
LDA	82.6	84.1	83.7	69.3	90.1	0.852 (0.761–0.880)
Test set						
PURS model						
RF	86.5	88.9	86.5	70.1	95.5	0.867 (0.772–0.952)
SVM	84.2	81.3	82.1	68.4	94.0	0.816 (0.697–0.933)
LDA	81.6	70.5	79.3	58.7	92.3	0.781(0.644–0.901)
IURS model						
RF	87.5	93.1	92.3	70.0	96.7	0.935 (0.843–0.976)
SVM	82.4	90.1	86.8	67.2	93.0	0.883 (0.779–0.944)
LDA	80.0	81.3	80.9	59.7	89.6	0.824 (0.712–0.890)
P-IURS model						
RF	88.8	95.0	93.4	73.6	98.5	0.952 (0.868–0.994)
SVM	85.1	91.3	87.5	69.1	93.7	0.906 (0.801–0.957)
LDA	82.3	84.0	83.2	68.5	90.4	0.857 (0.771–0.892)

Table 4 showed that the RF-based PURS, IURS, and the combined P-IURS models in the HER2+ subtype yielded robust efficacy with AUCs of 0.867, 0.935, and 0.952 in the test set.

heterogeneity, which was related to tumor progression and metastatic behavior.[19,22] In the present study, the dominant features selected in the most robust RF-based P-IURS predictive model in the HER2+ subtype were wavelet-related features. After the wavelet transform, the texture features including GLDM, GLSZM, and GLRLM are the main features. GLDM is a gray level dependence matrix and has been used to calculate the complexity of image textures.[23] GLSZM and GLRLM reflect the roughness and heterogeneity of texture features by calculating the size, length, or number of connected gray level zones of the image.[22,23] With the alliance of transform, wavelet features can detect more micro-environmental information about tumors, and provide more valuable predictors for biological behavior.[24] Our results showed that wavelet-related features are substantial signatures associated with axillary pCR after NAC.

The peritumoral region containing lymphatic vessel infiltration, micro-vascular proliferation, or stroma response surrounding the tumor has served as an additional prognostic factor. Previous studies had reported that the presence of peritumoral edema of BC was associated with tumor aggressiveness and ALNM.[25,26] MacColl et al.[27] showed that the residual carcinoma restricted to lymphatic spaces after NAC, which was called predominantly pure intralymphatic carcinoma, was related to residual positive lymph nodes. Previous studies also indicated that HR + tumors would benefit less from NAC due to the intrinsic or resistance to therapy, resulting in disease progression and poor outcomes.[28] In the present study, the LDA classifier-based PURS model in the HR+/HER2-subtype yielded the lowest ability for the prediction of axillary response. The result may be due to the HR + tumor with peritumoral lymphocytic infiltration being apt to no reaction to NAC, causing the residual nodal metastasis, resulting in axillary non-pCR.[27,28]

In addition to the predictive performance, Vila et al.[2] constructed clinicopathologic nomograms and obtained an AUC of 0.787. Kim et al.[10] reported that models based on breast MRI and US could predict axillary pCR with AUCs from 0.78 to 0.84. Gan et al.[19] developed a clinical-radiomics yielded an AUC of 0.878 for the prediction of axillary response in the test set. The present study applied RF, SVM, and LDA classifiers to construct PURS, IURS, and the combined P-IURS models in different molecular subtypes. RF is a regression tree technique that can use bootstrap aggregation and randomization of predictors to achieve a high accuracy.[29] SVM is a generalized

**Table 5. Coefficients of selected features in the HER2+ subtype**

Features	Coef in model
RF-based PURS model	
Wavelet-LHL_GLSZM_SmallAreaEmphasis	0.914
Original_GLSZM_LowGrayLevelZoneEmphasis	−1.579
Wavelet-LHL_GLCM_MCC	0.708
Wavelet-HHL_GLDM_SmallDependenceEmphasis	−3.309
Wavelet-HHL_GLSZM_GrayLevelNonUniformityNormalized	2.130
RF-based IURS model	
Original_Firstorder_Skewness	−1.267
Original_GLRLM_LowGrayLevelRunEmphasis	0.776
Wavelet-LHL_Firstorder_Median	−1.084
Wavelet-LHH_GLDM_DependenceNonUniformityNormalized	1.685
Wavelet-LHH_GLDM_LargeDependenceLowGrayLevelEmphasis	−0.420
Wavelet-LHH_GLRLM_RunLengthNonUniformityNormalized	3.560
Wavelet-HLH_GLSZM_HighGrayLevelZoneEmphasis	4.980
Wavelet-HHL_GLRLM_ShortRunEmphasis	−2.368
Wavelet-LLL_GLRLM_LowGrayLevelRunEmphasis	−0.596
RF-based combined P-IURS model	
Original_GLRLM_LowGrayLevelRunEmphasis	−0.928
Wavelet-LHL_GLCM_Imc2	−0.184
Wavelet-LHL_GLCM_ClusterTendency	−2.785
Wavelet-HHL_GLDM_SmallDependenceEmphasis	5.120
Wavelet-LHH_GLDM_LargeDependenceLowGrayLevelEmphasis	−1.530
Wavelet-LHH_GLRLM_RunLengthNonUniformityNormalized	−0.780
Wavelet-HLH_GLSZM_HighGrayLevelZoneEmphasis	2.479
Wavelet-HHL_GLRLM_ShortRunEmphasis	−1.391
Wavelet-LHL_GLRLM_RunEntropy	1.720
Wavelet-HHL_GLSZM_GrayLevelNonUniformityNormalized	3.852

There were 5, 9, and 10 optimal features selected in RF-based PURS, IURS, and combined P-IURS models in the HER2+ subtype.

linear classifier that can better solve small sample problems than other ML algorithms.[30] LDA is a classical ML method, which aims to find a linear data transformation increasing class discrimination in an optimal discriminant subspace.[31] Previous studies had reported that different classifiers had different predictive performances.[16,32] Beig et al.[33] reported that the SVM classifier with intranodular lung CT radiomic features achieved an AUC of 0.75 in the test set, and the combining radiomics of intranodular with perinodular regions improved the AUC to 0.80 to distinguish adenocarcinomas from granulomas. Braman et al.[20] indicated that a combined intratumoral and peritumoral DCE-MRI radiomic feature set using a diagonal linear discriminant analysis (DLDA) yielded a maximum AUC of 0.74 to predict breast pCR after NAC. Our study revealed that RF classifier-based IURS in the HER2+ subtype achieved an AUC of 0.935 (95%CI: 0.843, 0.976), and the combined P-IURS improved the AUC to 0.952 (95%CI: 0.868, 0.994) ($p < 0.05$) in the test set. The results showed that the combination of peritumoral and intratumoral US radiomics could improve the performance for predicting axillary pCR after NAC, which was consistent with previous studies.[20,33] Furthermore, our results also showed that the RF classifier obtained more robust predictive efficacy as compared with the SVM and LDA classifiers. Nevertheless, the LDA classifier was less efficient than RF and SVM, with the lowest ability (AUC of 0.681, 95%CI: 0.500, 0.805) in PURS of the HR+/HER2-subtype to predict axillary response after NAC. The inconsistencies may be due to the inclusion criteria or the different images such as US, MRI, or CT settings, further studies with larger data and multimodal images are required in the future.

This study contributes to the field of US-based radiomics analysis in the following ways:

First, with the increasing considerations about peritumoral regions, it is the first attempt to predict axillary response using peritumoral and intratumoral US radiomics signatures in patients with BC with node-positive. Second, compared with prior radiomics studies, most of which are limited to a single statistical method for analysis.[34,35] The present study applied various ML classifiers to construct different predictive models which may provide more robust results. Moreover, as compared with MRI or other images, the US is a more convenient diagnostic tool in breast examination, thus making the US-based radiomics analysis more wider application in clinical practice. Finally, we additionally
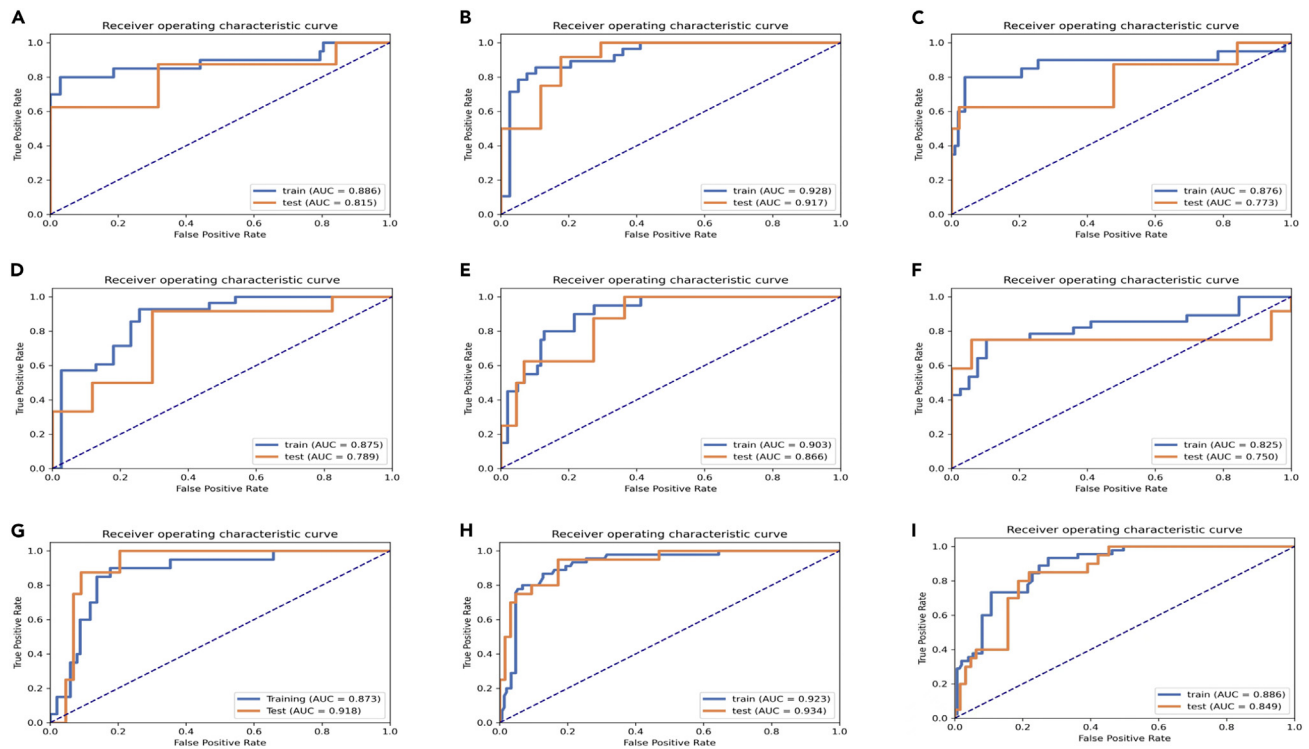
Figure 6. ROC curves in the TNBC subtype

The PURS model with RF (A), SVM (B), and LDA (C) classifiers. The AUCs for the three classifiers were 0.886, 0.928, and 0.876 in the training set, and 0.815, 0.917, and 0.773 in the test set, respectively. The IURS model with RF (D), SVM (E), and LDA (F) classifiers. The AUCs for the three classifiers were 0.875, 0.903, and 0.825 in the training set, and 0.789, 0.866, and 0.750 in the test set, respectively. The combined P-IURS model with RF (G), SVM (H), and LDA (I) classifiers. The AUCs for the three classifiers were 0.873, 0.923, and 0.886 in the training set, and 0.918, 0.934, and 0.849 in the test set, respectively.

assessed the predictive value of axillary pCR after NAC according to different molecular subtypes, which may assist in the selection of a clinical therapeutic regime.

In conclusion, ML-based PURS, IURS, and the combined P-IURS models in different molecular subtypes can assist in the estimation of axillary response after NAC. The RF classifier-based combined P-IURS in the HER2+ subtype achieved a favorable predictive accuracy and may be a promising clinical approach to help the selection of appropriate axillary surgical interventions in patients with BC with node-positive.

Limitations of the study

There are several limitations in our study: First, it was a retrospective study, and data were collected in a single institution, which may result in biased selection and a lack of external validation. Second, to better analyze the correlation between tumor PURS and IURS features and axillary pCR, we excluded patients with bilateral, multifocal tumors, nonmass-like tumors, and tumors with no clear peritumoral region on US images, which may also cause selected bias. Third, the present study did not include genomics data. Although genomics identification had been considered a promising predictive tool, they were not routinely performed in clinical practice. Further studies will be expected to address this issue in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Table 6. Predictive performance of ML classifiers-based PURS, IURS, and the combined P-IURS models in the TN subtype in the training and test sets

Training set	SEN (%)	SPE (%)	ACC (%)	PPV (%)	NPV (%)	AUC (95% CI)
PURS model						
RF	80.0	83.2	85.7	66.0	93.0	0.886 (0.831–0.939)
SVM	93.5	81.2	85.3	70.8	96.1	0.928 (0.880–0.964)
LDA	78.8	82.1	81.9	63.0	91.5	0.876 (0.828–0.926)
IURS model						
RF	79.1	81.8	81.4	62.1	90.9	0.875 (0.820–0.923)
SVM	85.7	89.7	88.8	85.7	90.1	0.903 (0.855–0.962)
LDA	74.0	82.1	80.5	60.3	89.9	0.825 (0.760–0.891)
P-IURS model						
RF	80.0	80.7	81.6	63.0	90.4	0.873 (0.818–0.925)
SVM	88.9	91.2	90.4	79.5	93.1	0.923 (0.866–0.971)
LDA	84.6	85.5	87.6	70.2	92.0	0.886 (0.803–0.939)
Test set						
PURS model						
RF	84.0	80.6	82.3	66.1	92.7	0.815 (0.694–0.920)
SVM	90.5	85.5	88.2	73.1	95.9	0.917 (0.859–0.960)
LDA	80.4	69.2	77.4	58.7	89.8	0.773 (0.664–0.901)
IURS model						
RF	70.6	81.4	75.5	52.0	89.0	0.789 (0.701–0.887)
SVM	81.6	84.3	82.2	68.6	92.3	0.866 (0.739–0.943)
LDA	68.8	77.6	74.9	50.8	88.5	0.750 (0.637–0.842)
P-IURS model						
RF	90.2	84.9	89.5	74.0	95.6	0.918 (0.862–0.970)
SVM	91.4	85.7	90.2	75.6	95.9	0.934 (0.877–0.983)
LDA	81.2	84.4	82.0	61.5	90.1	0.849 (0.745–0.900)

Table 6 showed that SVM-based PURS, IURS, and the combined P-IURS models in the TN subtype yielded better performance, with AUCs of 0.917, 0.866, and 0.934 in the test set.

- ○ Study design and patients
- ○ Clinical and pathological data
- ● METHOD DETAILS
  - ○ US image acquisition and regions of interest (ROIs) segmentation
  - ○ Feature extraction, selection and classifiers implementation
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

AUTHOR CONTRIBUTIONS

J.J.Y., X.H.J., W.Z., and J.Q.Z. contributed to the design of this study and the writing of the draft. J.J.Y. and X.H.J. analyzed the data. Y.Z. and X.S.C. assisted with the data collection and verification. W.W.Z. and J.Q.Z. supervised data collection, and reviewed the article for important intellectual content. W.Z. and J.Q.Z. supervised the study and revised the article. All authors read and approved the final version of the article.

DECLARATION OF INTERESTS

The authors declare no conflicts of interest.

Table 7. Coefficients of selected features in the TN subtype

Features	Coef in model
SVM-based PURS model	
Original_GLRLM_LowGrayLevelRunEmphasis	−0.925
Wavelet-LHL_Firstorder_Median	1.655
Wavelet-HHL_GLDM_SmallDependenceEmphasis	−1.913
Wavelet-HHL_GLRLM_ShortRunEmphasis	0.861
Wavelet-HHL_GLSZM_GrayLevelNonUniformityNormalized	−1.853
Wavelet-HHH_GLDM_LargeDependenceLowGrayLevelEmphasis	1.084
SVM-based IURS model	
Wavelet-LHL_GLCM_Imc2	−2.479
Wavelet-LHH_GLDM_LargeDependenceLowGrayLevelEmphasis	−1.913
Wavelet-LHH_GLRLM_RunLengthNonUniformityNormalized	0.861
Wavelet-HHL_GLRLM_ShortRunEmphasis	1.921
SVM-based combined P-IURS model	
Wavelet-HHL_GLSZM_GrayLevelNonUniformityNormalized	−0.962
Wavelet-LHH_GLDM_LargeDependenceLowGrayLevelEmphasis	2.031
Wavelet-LHH_GLRLM_RunLengthNonUniformityNormalized	−0.592
Wavelet-HHL_GLRLM_ShortRunEmphasis	1.921
Wavelet-LHL_GLCM_Imc2	−1.280
Wavelet-HHL_GLDM_SmallDependenceEmphasis	−0.053

There were 6, 4, and 6 optimal selected features in SVM-based PURS, IURS, and combined P-IURS models in the TN subtype.

REFERENCES

1. Samiei, S., Simons, J.M., Engelen, S.M.E., Beets-Tan, R.G.H., Classe, J.M., and Smidt, M.L.; EUBREAST Group (2021). Axillary pathologic complete response after neoadjuvant systemic therapy by breast cancer subtype in patients with initially clinically node-positive disease: a systematic review and meta-analysis. JAMA Surg. 156, e210891. https://doi.org/10.1001/jamasurg.2021.0891.

2. Vila, J., Mittendorf, E.A., Farante, G., Bassett, R.L., Veronesi, P., Galimberti, V., Peradze, N., Stauder, M.C., Chavez-MacGregor, M., Litton, J.F., et al. (2016). Nomograms for predicting axillary response to neoadjuvant chemotherapy in clinically node-positive patients with breast cancer. Ann. Surg Oncol. 23, 3501–3509. https://doi.org/10.1245/s10434-016-5277-1.

3. Mougalian, S.S., Hernandez, M., Lei, X., Lynch, S., Kuerer, H.M., Symmans, W.F., Theriault, R.L., Fornage, B.D., Hsu, L., Buchholz, T.A., et al. (2016). Ten-Year outcomes of patients with breast cancer with cytologically confirmed axillary lymph node metastases and pathologic complete response after primary systemic chemotherapy. JAMA Oncol. 2, 508–516. https://doi.org/10.1001/jamaoncol.2015.4935.

4. Osorio-Silla, I., Gómez Valdazo, A., Sánchez Méndez, J.I., York, E., Díaz-Almirón, M., Gómez Ramírez, J., Rivas Fidalgo, S., Oliver, J.M., Álvarez, C.M., Hardisson, D., et al. (2019). Is it always necessary to perform an axillary lymph node dissection after neoadjuvant chemotherapy for breast cancer? Ann. R. Coll. Surg. Engl. 101, 186–192. https://doi.org/10.1308/rcsann.2018.0196.

5. Mamtani, A., Barrio, A.V., King, T.A., Van Zee, K.J., Plitas, G., Pilewskie, M., El-Tamer, M., Gemignani, M.L., Heerdt, A.S., Sclafani, L.M., et al. (2016). How often does neoadjuvant chemotherapy avoid axillary dissection in patients with histologically confirmed nodal metastases? results of a prospective study. Ann. Surg Oncol. 23, 3467–3474. https://doi.org/10.1245/s10434-016-5246-8.

6. Kuehn, T., Bauerfeind, I., Fehm, T., Fleige, B., Hausschild, M., Helms, G., Lebeau, A., Liedtke, C., von Minckwitz, G., Nekljudova, V., et al. (2013). Sentinel-lymph-node biopsy in patients with breast cancer before and after neoadjuvant chemotherapy (SENTINA): a prospective, multicentre cohort study. Lancet Oncol. 14, 609–618. https://doi.org/10.1016/S1470-2045(13)70166-9.

7. Boughey, J.C., Suman, V.J., Mittendorf, E.A., Ahrendt, G.M., Wilke, L.G., Taback, B., Leitch, A.M., Kuerer, H.M., Bowling, M., Flippo-Morton, T.S., et al. (2013). Sentinel lymph node surgery after neoadjuvant chemotherapy in patients with node-positive breast cancer: the ACOSOG Z1071 (Alliance) clinical trial. JAMA 310, 1455–1461. https://doi.org/10.1001/jama.2013.278932.

8. Schulze, T., Mucke, J., Markwardt, J., Schlag, P.M., and Bembenek, A. (2006). Long-term morbidity of patients with early breast cancer after sentinel lymph node biopsy compared to axillary lymph node dissection. J. Surg. Oncol. 93, 109–119. https://doi.org/10.1002/jso.20406.

9. Al-Hattali, S., Vinnicombe, S.J., Gowdh, N.M., Evans, A., Armstrong, S., Adamson, D., Purdie, C.A., and Macaskill, E.J. (2019). Breast MRI and tumour biology predict axillary lymph node response to neoadjuvant chemotherapy for breast cancer. Cancer Imag. 19, 91. https://doi.org/10.1186/s40644-019-0279-4.

10. Kim, R., Chang, J.M., Lee, H.B., Lee, S.H., Kim, S.Y., Kim, E.S., Cho, N., and Moon, W.K. (2019). Predicting axillary response to neoadjuvant chemotherapy: breast MRI and US in patients with node-positive breast cancer. Radiology 293, 49–57. https://doi.org/10.1148/radiol.2019190014.

11. Gu, J., Tong, T., Xu, D., Cheng, F., Fang, C., He, C., Wang, J., Wang, B., Yang, X., Wang, K., et al. (2023). Deep learning radiomics of

ultrasonography for comprehensively predicting tumor and axillary lymph node status after neoadjuvant chemotherapy in breast cancer patients: A multicenter study. Cancer *129*, 356–366. https://doi.org/10.1002/cncr.34540.

12. Takada, M., and Toi, M. (2020). Neoadjuvant treatment for HER2-positive breast cancer. Chin. Clin. Oncol. *9*, 32. https://doi.org/10.21037/cco-20-123.

13. Loibl, S., and Gianni, L. (2017). HER2-positive breast cancer. Lancet *389*, 2415–2429. https://doi.org/10.1016/S0140-6736(16)32417-5.

14. Haug, C.J., and Drazen, J.M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023. N. Engl. J. Med. *388*, 1201–1208. https://doi.org/10.1056/NEJMra2302038.

15. Zhang, T., Tan, T., Samperna, R., Li, Z., Gao, Y., Wang, X., Han, L., Yu, Q., Beets-Tan, R.G.H., and Mann, R.M. (2023). Radiomics and artificial intelligence in breast imaging: a survey. Artif. Intell. Rev. *56*, 857–892. https://doi.org/10.1007/s10462-023-10543-y.

16. Yao, J., Zhou, W., Xu, S., Jia, X., Zhou, J., Chen, X., and Zhan, W. (2024). Machine learning-based breast tumor ultrasound radiomics for pre-operative prediction of axillary sentinel lymph node metastasis burden in early-stage invasive breast cancer. Ultrasound Med. Biol. *50*, 229–236. https://doi.org/10.1016/j.ultrasmedbio.2023.10.004.

17. Flores, R., Roldan, E., Pardo, J.A., Beight, L., Ubellacker, J., Fan, B., Davis, R.B., and James, T.A. (2023). Discordant breast and axillary pathologic response to neoadjuvant chemotherapy. Ann. Surg Oncol. *30*, 8302–8307. https://doi.org/10.1245/s10434-023-14082-2.

18. Kantor, O., Sipsy, L.M., Yao, K., and James, T.A. (2018). A predictive model for axillary node pathologic complete response after neoadjuvant chemotherapy for breast cancer. Ann. Surg Oncol. *25*, 1304–1311. https://doi.org/10.1245/s10434-018-6345-5.

19. Gan, L., Ma, M., Liu, Y., Liu, Q., Xin, L., Cheng, Y., Xu, L., Qin, N., Jiang, Y., Zhang, X., et al. (2021). A clinical-radiomics model for predicting axillary pathologic complete response in breast cancer with axillary lymph node metastases. Front. Oncol. *11*, 786346. https://doi.org/10.3389/fonc.2021.786346.

20. Braman, N.M., Etesami, M., Prasanna, P., Dubchuk, C., Gilmore, H., Tiwari, P., Plecha, D., and Madabhushi, A. (2017). Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res. *19*, 57. https://doi.org/10.1186/s13058-017-0846-1.

21. Mao, N., Shi, Y., Lian, C., Wang, Z., Zhang, K., Xie, H., Zhang, H., Chen, Q., Cheng, G., Xu, C., and Dai, Y. (2022). Intratumoral and peritumoral radiomics for preoperative prediction of neoadjuvant chemotherapy effect in breast cancer based on contrast-enhanced spectral mammography. Eur. Radiol. *32*, 3207–3219. https://doi.org/10.1007/s00330-021-08414-7.

22. Wang, X., Xie, T., Luo, J., Zhou, Z., Yu, X., and Guo, X. (2022). Radiomics predicts the prognosis of patients with locally advanced breast cancer by reflecting the heterogeneity of tumor cells and the tumor microenvironment. Breast Cancer Res. *24*, 20. https://doi.org/10.1186/s13058-022-01516-0.

23. Abbasian Ardakani, A., Bureau, N.J., Ciaccio, E.J., and Acharya, U.R. (2022). Interpretation of radiomics features - a pictorial review. Comput. Methods Programs Biomed. *215*, 106609. https://doi.org/10.1016/j.cmpb.2021.106609.

24. Sudarshan, V.K., Mookiah, M.R.K., Acharya, U.R., Chandran, V., Molinari, F., Fujita, H., and Ng, K.H. (2016). Application of wavelet techniques for cancer diagnosis using ultrasound images: a review. Comput. Biol. Med. *69*, 97–111. https://doi.org/10.1016/j.compbiomed.2015.12.006.

25. Cheon, H., Kim, H.J., Kim, T.H., Ryeom, H.K., Lee, J., Kim, G.C., Yuk, J.S., and Kim, W.H. (2018). Invasive breast cancer: prognostic value of peritumoral edema identified at preoperative MR imaging. Radiology *287*, 68–75. https://doi.org/10.1148/radiol.2017171157.

26. Kettunen, T., Okuma, H., Auvinen, P., Sudah, M., Tiainen, S., Sutela, A., Masarwah, A., Tammi, M., Tammi, R., Oikari, S., and Vanninen, R. (2020). Peritumoral ADC values in breast cancer: region of interest selection, associations with hyaluronan intensity, and prognostic significance. Eur. Radiol. *30*, 38–46. https://doi.org/10.1007/s00330-019-06361-y.

27. MacColl, C.E., Paré, G., Salehi, A., Hodgson, N., and Williams, P. (2021). Postneoadjuvant pure and predominantly pure intralymphatic breast carcinoma: case series and literature review. Am. J. Surg. Pathol. *45*, 537–542. https://doi.org/10.1097/PAS.0000000000001610.

28. Miranda, F., Prazeres, H., Mendes, F., Martins, D., and Schmitt, F. (2022). Resistance to endocrine therapy in HR + and/or HER2 + breast cancer: the most promising predictive biomarkers. Mol. Biol. Rep. *49*, 717–733. https://doi.org/10.1007/s11033-021-06863-3.

29. Paul, A., Mukherjee, D.P., Das, P., Gangopadhyay, A., Chintha, A.R., and Kundu, S. (2018). Improved Random Forest for Classification. IEEE Trans. Image Process. *27*, 4012–4024. https://doi.org/10.1109/TIP.2018.2834830.

30. Chao, C.F., and Horng, M.H. (2015). The construction of support vector machine classifier using the firefly algorithm. Comput. Intell. Neurosci. *2015*, 212719. https://doi.org/10.1155/2015/212719.

31. Xu, L., Raitoharju, J., Iosifidis, A., and Gabbouj, M. (2022). Saliency-Based Multilabel Linear Discriminant Analysis. IEEE Trans. Cybern. *52*, 10200–10213. https://doi.org/10.1109/TCYB.2021.3069338.

32. Uddin, S., Khan, A., Hossain, M.E., and Moni, M.A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC Med. Inform. Decis. Mak. *19*, 281. https://doi.org/10.1186/s12911-019-1004-8.

33. Beig, N., Khorrami, M., Alilou, M., Prasanna, P., Braman, N., Orooji, M., Rakshit, S., Bera, K., Rajiah, P., Ginsberg, J., et al. (2019). Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. Radiology *290*, 783–792. https://doi.org/10.1148/radiol.2018180910.

34. Bian, T., Wu, Z., Lin, Q., Wang, H., Ge, Y., Duan, S., Fu, G., Cui, C., and Su, X. (2020). Radiomic signatures derived from multiparametric MRI for the pretreatment prediction of response to neoadjuvant chemotherapy in breast cancer. Br. J. Radiol. *93*, 20200287. https://doi.org/10.1259/bjr.20200287.

35. Bitencourt, A.G.V., Gibbs, P., Rossi Saccarelli, C., Daimiel, I., Lo Gullo, R., Fox, M.J., Thakur, S., Pinker, K., Morris, E.A., Morrow, M., and Jochelson, M.S. (2020). MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer. EBioMedicine *61*, 103042. https://doi.org/10.1016/j.ebiom.2020.103042.

36. Penault-Llorca, F., André, F., Sagan, C., Lacroix-Triki, M., Denoux, Y., Verriele, V., Jacquemier, J., Baranzelli, M.C., Bibeau, F., Antoine, M., et al. (2009). Ki67 expression and docetaxel efficacy in patients with estrogen receptor–positive breast cancer. J. Clin. Oncol. *27*, 2809–2815. https://doi.org/10.1200/JCO.2008.18.2808.

37. Gradishar, W.J., Moran, M.S., Abraham, J., Aft, R., Agnese, D., Allison, K.H., Anderson, B., Burstein, H.J., Chew, H., Dang, C., et al. (2022). Breast Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. J. Natl. Compr. Canc. Netw. *20*, 691–722. https://doi.org/10.6004/jnccn.2022.0030.

38. Mercado, C.L. (2014). BI-RADS update. Radiol. Clin. North Am. *52*, 481–487. https://doi.org/10.1016/j.rcl.2014.02.008.

39. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al. (2012). 3D Slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imaging *30*, 1323–1341. https://doi.org/10.1016/j.mri.2012.05.001.

40. Xu, Z., Shen, D., Kou, Y., and Nie, T. (2024). A synthetic minority oversampling technique based on gaussian mixture model filtering for imbalanced data classification. IEEE Trans. Neural Netw. Learn. Syst. *35*, 3740–3753. https://doi.org/10.1109/TNNLS.2022.3197156.

41. Borstelmann, S.M. (2020). Machine learning principles for radiology investigators. Acad. Radiol. *27*, 13–25. https://doi.org/10.1016/j.acra.2019.07.030.

42. Song, Y., Zhang, J., Zhang, Y.D., Hou, Y., Yan, X., Wang, Y., Zhou, M., Yao, Y.F., and Yang, G. (2020). FeAture explorer (FAE): A tool for developing and comparing radiomics models. PLoS One *15*, e0237587. https://doi.org/10.1371/journal.pone.0237587.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw data	This paper	https://pan.baidu.com/s/1ugLf5dkN2ZZilytmachiGA?pwd=yjj6
Supplementary data	This paper	Tables S1, S2, and Data S1–S9
Analyzed data	This paper	Tables 1, 2, 3, 4, 5, 6, and 7
Software and algorithms		
3D slicer software	Fedorov et al.[39]	https://www.slicer.org
Pyradiomics software	This paper	https://pyradiomics.readthedocs.io/en/latest/
Python software	This paper	https://www.python.org
FeAture Explorer Pro (FAEPro, V0.5.3)	Song et al.[42]	https://github.com/salan668/FAE
Random forest (RF) algorithm	This paper	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#
Support vector machine (SVM) algorithm	This paper	https://scikit-learn.org/stable/modules/svm.html#svm
Linear discriminant analysis (LDA) algorithm	This paper	https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html#
SPSS software (version 23.0)	This paper	https://www.ibm.com/products/spss-statistics
MedCalc software (version 22.013)	This paper	http://www.medcalc.com.cn

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and data should be directed to and will be fulfilled by the lead contact, Jian-qiao Zhou (e-mail:zhousu30@126.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data supporting the findings of this study can be downloaded. The DOI is listed in the key resources table.
- All original code can be seen in the official website, DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study design and patients

This study was a retrospective analysis, and was conducted in accordance with the Declaration of Helsinki, approved by the Ethics Committee of our institution (approval number 2023221). Informed consent was waived due to its retrospective nature. Datasets were obtained from the Shanghai Jiaotong University Breast Cancer Database (SJTU-BCDB). From January 2017 to August 2023, we firstly collected 628 BC patients with node-positive accepted NAC and post NAC surgery at our institution. The included criteria were as follows: (i) patients were confirmed to be primary BC with node-positive, and without distant metastasis; (ii) patients underwent a full course of NAC; (iii) patients accepted post NAC breast and axillary surgery, and the axillary nodal pCR was proved by surgical specimen histopathological examination; (iv) patients with high quality pre-NAC baseline breast tumor US images. In the present study, node-positive was defined as axillary lymph node metastasis (ALNM) confirmed by either fine needle aspiration (FNA) or core needle biopsy (CNB) before the initiation of NAC. Axillary pCR was defined as the complete absence of micro- and macro-metastases in ALNs. The exclusion criteria were as follows: (i) patients who did not complete NAC regimen; (ii) patients with no baseline breast tumor US images and insufficient clinicopathologic data; (iii) patients with bilateral or multiple tumors; (iv) patients with nonmass-like lesions or no sufficient peritumoral tissue identified on US images. Finally, a total of 435 patients were included in the study population. All patients were individuals of East Asian descent, Chinese women, and Han nationality, with mean age 46.3 ± 11.1 years, median age 47.9 years, and age range from 36 to 78 years.

Clinical and pathological data

The clinical data including patients' age, clinical T stage, and NAC regimens were also retrieved from the Database (SJTU-BCDB). The pathological data such as tumor histological type, estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) status, and tumor proliferation rate (Ki-67 levels) were determined from the results of CNBs performed before NAC. HER2 status was confirmed with fluorescence *in situ* hybridization. A cut-off value for Ki67 positive was established at 20%.[36] Tumors were classified into three molecular subtypes based on the expression of ER, PR, and HER2 status: HR+/HER2- (HR+, HER2-); HER2+ (HER2+, ER + or ER-, PR + or PR-); and TN (ER-, PR-, HER2-). HR+ was defined as ER+ and/or PR+. The NAC regimens such as adriamycin with cyclophosphamide (AC), adriamycin with docetaxel (AT), adriamycin with cyclophosphamide plus docetaxel (AC-T) were treated for non-HER2+ patients. HER2+ patients also received trastuzumab, or trastuzumab with pertuzumab. All patients accepted six or eight cycles of NAC before surgery according to the National Comprehensive Cancer Network (NCCN) guideline.[37] Among the 435 BC patients, 174 were HR+/HER2-subtype, 165 were HER2+ subtype, and 96 were TN subtype patients. Patients was randomly (7:3) divided into a training set including 121 of HR+/HER2-, 115 of HER2+, and 67 of TN subtype. An independent test set included 53 patients of HR+/HER2-, 50 of HER2+, and 29 of TN subtype. Figure 1 shows the patient recruitment and study design.

METHOD DETAILS

US image acquisition and regions of interest (ROIs) segmentation

Breast US examinations were performed 1 week before biopsy by Resona 7 and Resona 8 (Mindray Medical International, Shenzhen, China) with a linear probe at 3–11 MHz, and Esaote MyLab 60 (Esaote, Genoa, Italy) with a linear probe at 4–13 MHz. Tumors were assessed according to the Breast Imaging Reporting and Data System (BI-RADS).[38] The maximum size of the breast tumors measured by US were also recorded.

The tumor regions of interest (ROIs) segmentation and extraction were performed by using 3D-slicer (3D Slicer version 5.0.3) and PyRadiomics software.[39] The intratumoral region was segmented by dilating the delineated tumor contour manually in the largest cross-sectional area. The peritumoral ROI was obtained with a 3mm-thick surrounding zone outside the intratumoral region automatically using "Hollow" and "Margin" segment editors. The ROIs of peritumoral and intratumoral area were extracted separately according to molecular subtype. Figures 2A–2L shows the examples of the tumor US images and its corresponding ROIs. The P-IURS was calculated by combining peritumoral and intratumoral radiomics signatures in each subtype.

For the reproducibility of feature extraction, two experienced radiologists (author 1 and author 2, with over ten years of experience in breast US, and three years of experience in the software) initially segmented the peritumoral and intratumoral regions of 60 randomly selected breast tumors, and extracted the radiomics signatures separately. Both radiologists were blinded to the treatment outcomes. One weeks later, author 1 repeated the same procedure and analyzed the remaining images. An intra- and interclass correlation coefficient (ICC) equal to or higher than 0.75 was regarded as good intra- and interobserver agreement, and was included in the further feature selection process.

Feature extraction, selection and classifiers implementation

The extracted radiomics features included 14 shape features, 18 first-order features, 38 s-order texture features [24 Gy level co-occurrence matrix (GLCM) and 14 Gy level dependence matrix (GLDM)], 37 higher-order texture features [16 Gy level run length matrix (GLRLM), 16 Gy level size zone matrix (GLSZM) and 5 neighboring gray tone difference matrix (NGTDM)], and 744 wavelet-related features (details are shown in Tables S1 and S2). The final dataset comprised each of 148,074 PURS, IURS and the combined P-IURS in HR+/HER-subtype (Data S1–S3), 140,415 in HER2+ subtype (Data S4–S6), and 81,696 in TN subtype (Data S7–S9).

In the features processing and selection procedure, synthetic minority oversampling technique (SMOTE) was firstly used to remove the unbalance samples in each of dataset.[40] Z score and Mean normalization were applied to standardize the corresponding features. Then, principal component analysis (PCA) and pearson correlation coefficient (PCC) were employed to reduce the features dimension and meanwhile prevent over-fitting. After that, recursive features elimination (RFE) was utilized to detect the most relevant predictive features.[41] Finally, the most robust radiomics signatures selected by the above procedures were input to RF, SVM, and LDA classifiers with a 5-fold cross validation to construct and validate PURS, IURS and the combined P-IURS models for the prediction of axillary pCR in different molecular subtypes. Figure 3 shows the overview of the workflow.

QUANTIFICATION AND STATISTICAL ANALYSIS

All numerical data were presented as mean ± standard deviation. Continuous and categorical variables were compared using the two-sided independent *t* test, and the Chi-square test or Fisher's exact test, respectively. The training dataset was used to construct various ML classifiers-based PURS, IURS and the combined P-IURS models for predicting axillary pCR in three subtypes, the independent test dataset was used to validate the models. The predictive efficacy were assessed with respect to sensitivity (SEN), specificity (SPE), accuracy (ACC), positive predictive value (PPV), negative predictive value (NPV), and the area under the receiver operating characteristic (ROC) curve (AUC). Comparisons of AUCs between ML classifiers-based predictive models were made by using the DeLong test. All of the processes were implemented with FeAture Explorer Pro (FAEPro, V0.5.3) in Python (3.7.6) (https://github.com/salan668/FAE),[42] SPSS software (version 23.0), and MedCalc software (version 22.013). A *p* value less than 0.05 was regarded as significant difference.