Am J Cancer Res 2022;12(1):152-164
www.ajcr.us /ISSN:2156-6976/ajcr0136954

Original Article

Association of machine learning ultrasound radiomics and disease outcome in triple negative breast cancer

Haoyu Wang[1*], Xiaokang Li[2*], Ying Yuan[3*], Yiwei Tong[1], Siyi Zhu[1], Renhong Huang[1], Kunwei Shen[1], Yi Guo[2#], Yuanyuan Wang[2#], Xiaosong Chen[1#]

[1]Department of General Surgery, Comprehensive Breast Health Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; [2]Department of Electronic Engineering, Fudan University, Shanghai 200433, China; [3]Department of Radiology, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200010, China. [*]Equal contributors. [#]Equal contributors.

Received July 2, 2021; Accepted December 13, 2021; Epub January 15, 2022; Published January 30, 2022

Abstract: Triple negative breast cancer (TNBC) is a breast cancer subtype with unfavorable prognosis. We aimed to establish a machine learning-based ultrasound radiomics model to predict disease-free survival (DFS) in TNBC. Invasive TNBC>T1b between January 2009 and June 2018 with preoperative ultrasound were enrolled and assigned to training and independent test cohort. Radiomics and clinicopathological features related with DFS were selected by univariate and multivariate regression analysis. Training cohort of combined features was resampled with SMOTEENN to balance distribution and put into classifiers. Areas Under Curves (AUCs) of models were compared by DeLong's test. 562 women were included with 68 DFS events observed. Twenty prognostic radiomics features were extracted. Machine learning model by Naïve Bayes combining radiomics, clinicopathological features, and SMOTEENN had an AUC of 0.86 (95% CI 0.84-0.88), with sensitivity of 74.7% and specificity of 80.1% in training cohort. In independent test cohort, this three-combination model delivered an AUC of 0.90 (95% CI 0.83-0.95), higher than models based on radiomics (AUC=0.69, P=0.016) or radiomics + SMOTEENN (AUC=0.73, P=0.019). Integrating machine learning radiomics model based on ultrasound and clinicopathological features can predict DFS events for TNBC patients.

Keywords: Triple negative breast cancer, ultrasonography, radiomics, machine learning, prognosis

Introduction

Breast cancer is one of the most commonly diagnosed malignancy among females and has become the second leading cause of tumor-related death for women worldwide [1]. In the era of precise diagnoses and individualized treatment, classifications of molecular subtypes have become the backbone of management strategy for breast cancer [2, 3]. Triple negative breast cancer (TNBC), which accounts for approximately 10-15% of newly diagnosed breast cancer, harbors more malignant biological behaviors compared with other molecular subtypes [4]. With higher nuclear grade, larger tumor size, and more aggressive proliferative documents, people with TNBC had a higher risk of recurrence and worse overall survival [5]. Thus, the spotlight of clinical and translational research in TNBC field always includes identifying risk factors of developing relapse, thus to find out high risk populations to guide individualized therapy [6]. Among traditional clinicopathological factors, younger age at diagnosis, axillary lymph node (ALN) involvement, and lymphatic vessel invasion (LVI) have been reported to associate with the higher relapse rate of TNBC in long-term follow-up studies [7, 8]. However, to better understand the recurrence pattern of TNBC, more novel biomarkers need to be studied.

Ultrasound (US) has been widely used in screening and diagnosis of breast cancer with its advantages of no radiation and good accessibility in clinical practice [9, 10]. Several studies have been exploring its predictive and prognostic values for breast cancer. It was reported that that breast cancers classified as Breast Imaging Reporting and Data System (BI-RADS) 4A cate-

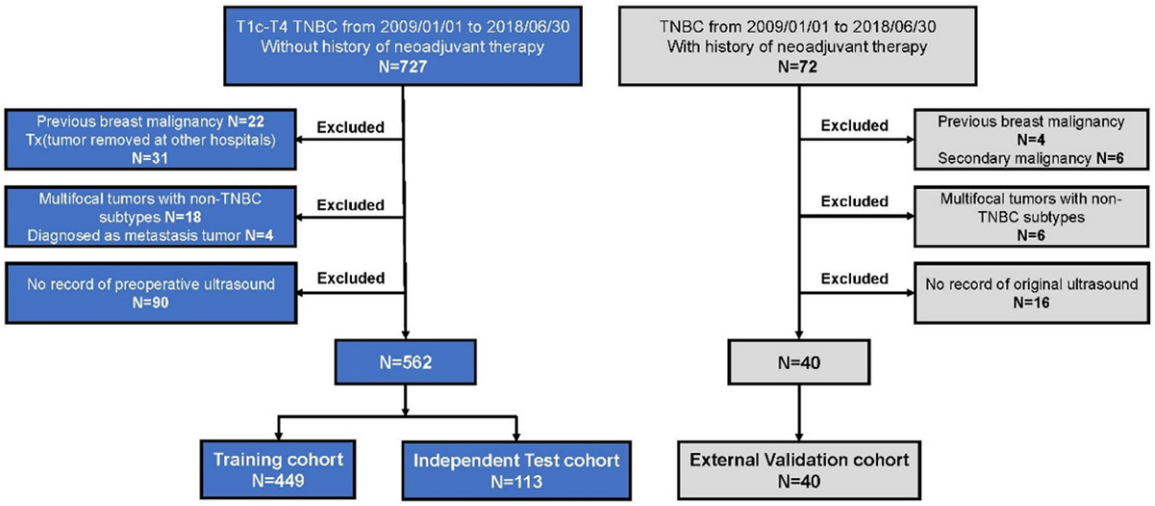

Figure 1. Flow chart of enrollment. Eligibility and exclusive criteria were shown in the flow chart. Finally, 562 patients were retrospectively included, among which 449 patients were randomized into training cohort while 113 patients into Independent Test cohort. On the other hand, 40 TNBC patients were included as the External Validation cohort.

gory in screening US had a higher risk of recurrence compared with tumors with 4B-5 categories [11]. Notably, our previous studies reviewed the preoperative sonographic features of TNBC and found out that TNBC tumors with vertical orientation had worse RFS and more ALN metastases, indicating that ultrasound characteristics could provide prognostic information for TNBC patients [12, 13]. However, the accuracy of this feature recognition was limited by subjective evaluation of US operators.

Radiomics was able to automatically extract quantitative image features with large scales and high accuracy [14]. Our previous studies have shown that radiomics analysis of breast cancer ultrasound had a high reproducibility and was able to predict molecular classifications and biological behaviors for breast tumor [15, 16]. Furthermore, artificial intelligence (AI), especially the machine learning algorithm, has gained extensive attention in the field of breast cancer research, especially in screening and diagnostic settings [17]. Machine learning-based radiomics model with convolutional network method on screening mammography have been reported to reach an Area Under Curve (AUC) of 0.98 in breast cancer detection [18]. As for sonographic radiomics, Arturo Brunetti et al. managed to distinguish malignancy breast tumors from benign lesions through an ultrasound radiomic analysis combined with machine learning [19]. Meanwhile, Zheng et al. have established a predictive model for lymph node metastasis by machine learning radiomics of preoperative ultrasound with an AUC value of 0.90 [20].

As shown above, previous literatures regarding machine learning radiomics based on ultrasound have mostly focused on optimizing diagnostic efficacy including recognition of malignancies or axillary lymph nodes. However, whether machine learning-based radiomics models with sonography could predict patients' long-term outcomes, especially in TNBC, has barely been explored. Hence, the purpose of our study was to evaluate the prognostic predictive value of machine learning radiomics based on ultrasound for disease outcomes in TNBC patients, thus to establish a machine learning-based model to further classify TNBC patients with various disease outcome.

Methods and materials

Patients

For model establishment, patients diagnosed with TNBC at the Breast Health Center of our hospital between January 1st 2009 to June 30th 2018 and underwent surgical treatment were screened. Patients with invasive TNBC larger than 1.0 cm with record of preoperative ultrasound were included for analysis. Patients with history of neoadjuvant treatment, previous breast malignancy, multifocal tumors with other molecular subtypes or history of other malignancy were excluded (**Figure 1**).

On the other hand, another panel of TNBC patients who received neoadjuvant therapy between January 1st 2009 to June 30th 2018 were included as the external validation cohort. Patients with records of original ultrasound images before treatments without history of previous malignancy were retrospectively enrolled (**Figure 1**). The study was performed under the Declaration of Helsinki and has been approved by the Institutional review board of our hospital.

Pathological evaluation

Pathological evaluation was conducted by the Department of Pathology in our hospital. Breast tumors were fixed in formalin, embedded in paraffin, stained with hematoxylin-eosin, and then evaluated for pathology types. ER, PR, HER2, and Ki67 expression was examined by immunohistochemistry (IHC). Nuclear staining in at least 1% tumor specimen was defined as ER or PR positivity [21]. HER2 negativity was determined as IHC 0-1+ or negative on fluorescence in situ hybridization (FISH), while positivity as IHC 3+ or positive on FISH [22]. TNBC was defines as breast cancer with no expression of ER, PR, and HER2.

Data collection and follow-up

Clinicopathological profiles and follow-up data of patients were recorded and retrieved from the Shanghai Jiaotong University Breast Cancer Database (SJTU-BCDB). Clinicopathological features including patients' age, menstrual status, breast and axillary surgery types, pathology types, tumor size, ALN metastases, nuclear grade, Ki-67, LVI and adjuvant treatments were taken into analysis.

Information of follow-up was collected by specialized nurses. Disease-free survival (DFS) events were recorded and analyzed, which was defined as the interval between the date of surgery and the date of breast cancer recurrence, secondary primary cancer, or death of any reason.

Ultrasound examination and image segmentation

Preoperative ultrasounds were performed and reviewed by two proficient radiologists with more than 10-years' experience in breast imaging. Sonograms were all conducted by the machines of MyLab60 (Esaote, Genoa, Italy) or Philip HD15 (Philips, Rochester, NY, USA) equipped with 5-12 MHz linear probes. Static images and video profiles were then stored in the system of Digital Imaging and Communications in Medicine (DICOM). The ultrasound imaging was then assessed with the criteria of ACR BI-RADS® Atlas.

1-3 representative ultrasound images of targeted lesions were selected. Contours of tumors were manually extracted by Polygen mode in ITK-SNAP (Windows 3.4.0 version) and independently reviewed by two sonographic specialists (**Figure 2**).

Feature extraction and selection

A total of 460 radiomics characteristics were extracted and quantified from each ultrasound images in MATLAB (Windows 2020a version). The features include morphological (15 features), histogram-based (16 features), texture features (73 features) and wavelet features (356 features). For tumors with more than one representative sonographic images, mean index of each feature was measured and taken into analysis. Radiomics features and clinicopathological features associated with occurrence of DFS events were then explored by Logistic Regression. Characteristics with *P* value <0.05 were considered as significantly associated with DFS events and then included for further model construction.

Data resampling and machine learning models

SMOTEENN [23] was used to conduct data balancing in order to improve predictive performance in our imbalanced dataset. It is a hybrid sampling method combined oversampling technique SMOTE (Synthetic Minority Oversampling Technique) and under-sampling technique ENN (Edited Nearest Neighbor). In the procedure of the method, firstly SMOTE generates synthetic samples by randomly interpolating between existing samples in the minority class [24]. Then ENN cleans the newly generated dataset to prevent overlap of samples between the minority class and the majority class. Specifically, a sample from one class will be eliminated if more than half of its K nearest neighbors do not belong to the same class. As a result, SMOTEENN makes the sample numbers of the two classes closer and the boundaries between them clearer. Therefore, the classi-

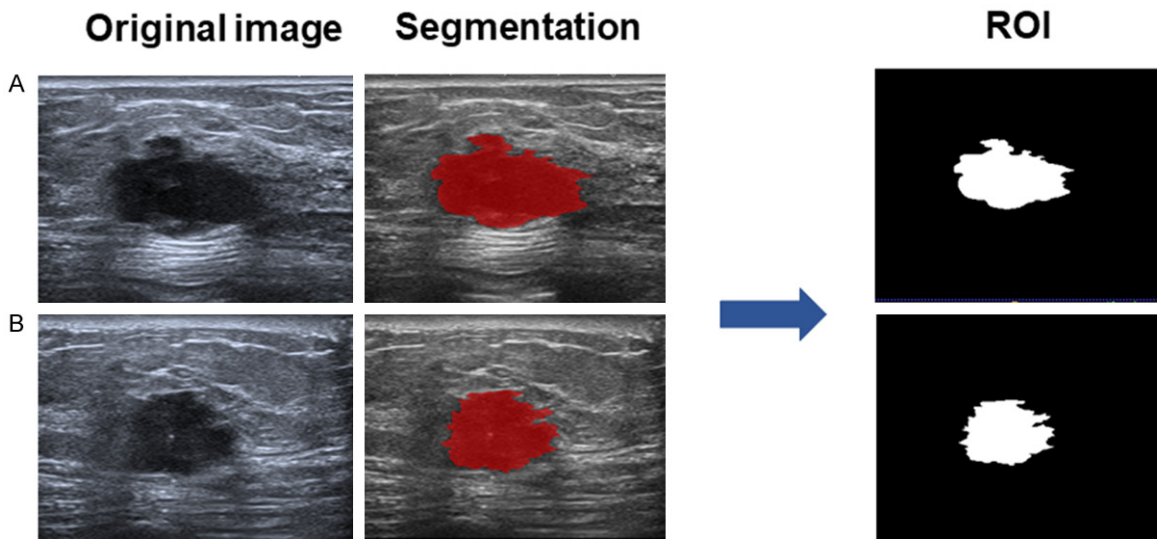

Figure 2. Examples of ultrasound segmentations. Typical examples of ultrasound segmentations were shown. Representative ultrasound images of breast tumor were selected. The contours of the lesions were manually drawn in ITK-SNAP and the ROIs were then extracted.

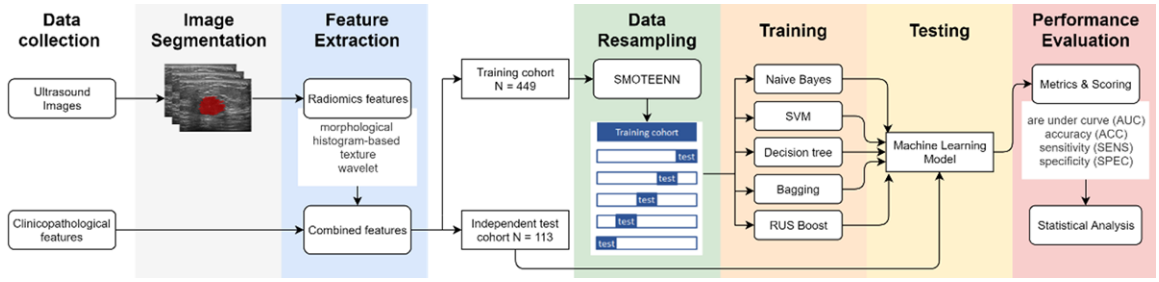

Figure 3. Workflow of the machine learning algorithm. High-throughout radiomic features were extracted from segmentations of breast ultrasound. Taken clinicopathological features together, enrolled samples were randomly assigned into the Training cohort and Independent test cohort with the ratio of 4:1. Data resampling was then performed through the methods of SMOTE-ENN to balance the events. 5 classifiers were conducted with 5-fold cross validation and compared in the training cohort and then tested in independent test cohort to find out the best Machine Learning Model. Performance of each model was evaluated by AUC, ACC, SENS and SPEC.

fier can easily learn the differences between the two classes, thus improving the prediction performance.

In our study, five machine learning classifiers were used to predict DFS, including Naive Bayes, SVM, Decision Tree, Bagging, and RUS Boost. The first three of these classifiers are traditional machine learning classifiers that assume roughly equal numbers of samples and the same cost of misclassification in each class. However, if these traditional classifiers were used on our imbalanced dataset, they would be prone to misclassify the minority class. Therefore, two ensemble classifiers Bagging and RUS Boost, were also employed.

The whole cohort was randomly assigned into the training cohort and the independent test cohort with a ratio of 4:1. Among the training cohort, predictive performance of 5 classifiers was compared by 5-fold cross-validation test. Performance of different models was then further validated in the external validation cohort. The workflow of the machine learning algorithm was shown in **Figure 3**.

Statistical analysis

Analysis was conducted by IBM SPSS Statistics (Windows 25.0 version), R (Windows 3.6.3 version), Python (Windows 3.8.5 version) and MATLAB (Windows 2020a version). All tests were two-sided and *P* value <0.05 was consid-

Table 1. Clinicopathological features of the enrolled population

Variables		Total N=562	DFS event N=68	No DFS event N=494	*P* value
Race	Asian	562 (100.0)	68 (100.0)	494 (100.0)	NA
Age (yrs)	≤55	289 (51.4)	29 (41.6)	260 (52.6)	0.122
	>55	273 (48.6)	39 (57.4)	234 (47.4)	
Menstruation status	Pre-/peri-	218 (38.8)	20 (29.4)	198 (40.1)	0.090
	Post-	344 (61.2)	48 (70.6)	296 (59.9)	
Breast surgery	BCS	215 (38.3)	19 (27.9)	196 (39.7)	0.104
	Mastectomy	347 (61.7)	49 (72.1)	298 (60.3)	
Axillary surgery	SLNB	285 (50.7)	18 (26.9)	261 (53.4)	<0.001
	ALND	277 (49.3)	49 (73.1)	228 (46.6)	
Pathology type	IDC	483 (85.9)	63 (92.6)	420 (85.0)	0.090
	Non-IDC	79 (14.1)	5 (7.4)	74 (15.0)	
Tumor size	Mean ± SE	2.5±0.1	2.9±0.1	2.5±0.1	0.006
	≤2 cm	246 (43.8)	18 (26.5)	228 (46,2)	0.002
	>2 cm	316 (56.2)	50 (73.5)	266 (53.8)	
ALN metastases	No	401 (71.7)	35 (51.5)	370 (74.9)	<0.001
	Yes	161 (28.3)	33 (48.5)	124 (25.1)	
Nuclear grade	I-II	91 (16.2)	10 (14.7)	81 (16.4)	0.546
	III	395 (70.3)	52 (76.5)	343 (69.4)	
	NA	76 (13.5)	6 (8.8)	70 (14.2)	
Ki-67 (%)	Mean ± SE	54.6±1.1	54.8±1.1	53.6±3.3	0.720
	≤30	148 (26.3)	16 (23.5)	132 (26.7)	0.575
	>30	414 (73.7)	52 (76.5)	362 (73.3)	
LVI	No	515 (91.6)	56 (82.4)	459 (92.9)	0.003
	Yes	47 (8.4)	12 (17.6)	35 (7.1)	
TNM stage	I	196 (35.1)	12 (17.9)	184 (37.4)	<0.001
	II	308 (55.1)	36 (53.7)	272 (55.3)	
	III	55 (9.8)	19 (28.4)	36 (7.3)	
Chemotherapy	No	52 (9.3)	11 (16.2)	41 (8.3)	0.036
	Yes	509 (90.7)	57 (83.8)	452 (91.7)	
Radiotherapy	No	244 (43.5)	27 (39.7)	217 (44.0)	0.502
	Yes	317 (56.5)	41 (60.3)	276 (56.0)	

Abbreviations: DFS, disease-free survival; BCS, breast conserving surgery; SLNB, sentinel lymph node biopsy; ALND, axillary lymph node dissection; IDC, invasive ductal carcinoma; SE, standard error; ALN, axillary lymph node; NA, not available; LVI, lymphatic vascular invasion; TNM, tumor node metastasis.

ered as significantly important. Regarding baseline characteristics, categorical variables were shown as numbers and percentages and analyzed by Pearson's Chi-square test (or Fisher's exact test); while continuous variables were shown as means and standard errors (SEs) and was analyzed by independent sample t test. Performances of classifiers and prediction models were evaluated in four merits, including the Area Under Curve (AUC), the accuracy (ACC), the Specificity (SPEC) and the Sensitivity (SENS). AUC of different prognostic models was compared with the method of DeLong's test [25].

Results

Basic characteristics

From January 2009 to June 2018, 562 patients diagnosed with TNBC were included for model establishment (**Figure 1**). As shown in **Table 1**, mean age of enrolled patients was 55.5 (27-87) years old and 266 (61.4%) patients were post-menopausal at diagnosis. There were 215 (38.3%) patients underwent breast conserving surgery (BCS) while 347 (61.7%) received mastectomy. Sentinel lymph node biopsy (SLNB) was performed among 285

Table 2. Disease outcomes of patients

Events	N	Percentage
Recurrences	57	10.1%
LRR	7	1.2%
Distance	50	8.9%
Secondary tumors	4	0.7%
Deaths	34	6.0%
Breast cancer-specific	27	4.8%
Non-breast cancer-specific	7	1.2%

Abbreviations: LRR, locoregional recurrence.

(50.7%) patients. There were 316 (56.2%) patients with tumor size >2.0 cm and 395 (70.3%) with grade III disease. Mean Ki-67 value for enrolled patients was 54.6% (0-95%). ALN metastases were detected in 161 (28.3%) patients.

Disease outcomes

Disease outcomes of the enrolled population were listed in **Table 2**. With a median follow-up of 76.0 months, 68 (12.1%) DFS events were observed. Fifty-seven (10.1%) patients had distance recurrence, among which 27 (4.8%) patients have died with breast cancer events. Four (0.7%) patients developed secondary tumors. A total of 7 (1.2%) patients died without breast recurrence: 3 patients for myocardial infarction, 2 for cerebrovascular accident, 1 for respiratory function failure, and 1 for renal function failure.

Training and independent test cohort

The whole cohort was randomly assigned to training cohort (N=499) or independent test cohort (N=113). Clinicopathological characteristics including tumor size, lymph node metastasis, nuclear grade, and Ki-67 index were well-balanced between two cohorts (all P>0.05, **Table 3**). Fifty-seven (12.7%) patients in the training cohort while 11 (9.7%) in the independent test cohort had DFS events respectively, which also showed no significant difference. The training cohort was taken into 5-fold cross-validation test and then validated in the independent test cohort.

Classifier selection and clinical information integration

With Logistic Regression test, 20 radiomic characteristics related with DFS events were selected and taken into model construction (Table S1), including one morphological feature, one histogram-based feature, 3 texture features and 15 wavelet features. The boxplots of four representative radiomics features are shown in **Figure 4**, where the significant differences in feature means reflect the strong correlation with DFS events. In order to explore the most suitable algorithm for prediction model, classifiers including Naive Bayes, SVM, Decision Tree, Bagging, and RUS Boost in predicting DFS events were compared. As illustrated in **Table 4**, the classifier Naive Bayes had the best performance in predicting DFS events when only radiomic features were taken into consideration with AUC 0.69 in the independent test cohort, which was then adopted for further model construction.

Regarding clinicopathological features, both clinicopathological characteristics and treatment choices were taken into consideration. As shown in **Table 1**, larger tumor size (P=0.006), more lymph node metastases (P<0.001), presence of LVI (P=0.003), and higher TNM stage (P<0.001) was significantly related with elevated risk of DFS events and was further selected into modeling. The AUC value of model based on clinicopathological factors was 0.79 but the sensitivity was only 54.5%. Moreover, for combination of US radiomics and clinicopathological features, the AUC value can reach to 0.86 in the independent test cohort, but which was only 0.65 in the training cohort (**Table 5**). In addition, the sensitivity of combination model was only 25.6% and 63.6% in the training and independent-test cohorts, respectively.

Prediction of DFS events with machine learning radiomics

Due to the relatively small number of DFS events, SMOTEENN algorithm was further used to build prediction model. Radiomics + SMOTEENN, clinicopathological + SMOTEENN, and Radiomics + clinicopathological + SMOTEENN models were built and compared, which had the AUC values 0.84 (95% CI 0.82-0.86), 0.81 (95% CI 0.76-0.85), and 0.86 (95% CI, 0.84-0.88) in the training cohort and 0.73 (95% CI 0.64-0.81), 0.80 (95% CI 0.72-0.87), and 0.90 (95% CI 0.83-0.95) in the independent test cohort, respectively (**Figure 5**; **Table 5**). The radiomics + clinicopathological + SMOTEENN model had a higher AUC than models

Table 3. Features of training and independent test cohorts

Variables		Training N=449	Independent test N=113	*P* value
Age (yr)	≤55	225 (50.1)	64 (56.6)	0.215
	>55	224 (49.9)	49 (43.4)	
Menstruation status	Pre-/peri-	168 (37.4)	50 (44.2)	0.183
	Post-	281 (62.6)	63 (55.8)	
Breast surgery	BCS	177 (39.4)	38 (33.6)	0.257
	Mastectomy	272 (60.6)	75 (66.4)	
Axillary surgery	SLNB	228 (51.1)	51 (46.4)	0.395
	ALND	218 (48.9)	59 (53.6)	
Pathology type	IDC	390 (86.9)	93 (82.3)	0.226
	Non-IDC	59 (13.1)	20 (17.7)	
Tumor size	Mean ± SE	2.6±0.1	2.4±0.1	0.232
	≤2 cm	197 (43.9)	49 (43.4)	0.922
	>2 cm	252 (56.1)	64 (56.6)	
ALN metastases	No	330 (73.5)	75 (66.4)	0.131
	Yes	119 (26.5)	38 (33.6)	
Nuclear grade	I-II	73 (16.3)	18 (15.9)	0.991
	III	315 (70.2)	80 (70.8)	
	NA	61 (13.6)	15 (13.3)	
Ki-67 (%)	Mean ± SE	54.2±1.2	56.2±2.4	0.719
	≤30	122 (27.2)	26 (23.0)	0.369
	>30	327 (72.8)	87 (77.0)	
LVI	No	417 (92.9)	98 (86.7)	0.055
	Yes	32 (7.1)	15 (13.3)	
TNM stage	I	163 (36.4)	33 (29.7)	0.383
	II	243 (54.2)	65 (58.6)	
	III	42 (9.4)	13 (11.7)	
Chemotherapy	No	44 (9.8)	8 (7.1)	0.469
	Yes	404 (90.2)	105 (92.9)	
Radiotherapy	No	194 (43.3)	50 (44.2)	0.916
	Yes	254 (56.7)	63 (55.8)	
DFS events	No	392 (87.3)	102 (90.3)	0.518
	Yes	57 (12.7)	11 (9.7)	

Abbreviations: DFS, disease-free survival; BCS, breast conserving surgery; SLNB, sentinel lymph node biopsy; ALND, axillary lymph node dissection; IDC, invasive ductal carcinoma; SE, standard error; ALN, axillary lymph node; NA, not available; LVI, lymphatic vascular invasion; TNM, tumor node metastasis.

based only on radiomic features (AUC 0.69, P=0.016) or radiomics + SMOTEENN (AUC 0.73, P=0.019) (Figure S1). Furthermore, the radiomics + clinicopathological + SMOTEENN model exhibited a high sensitivity (SENS=81.8%) and specificity (SPEC=82.3%) in predicting DFS events in TNBC patients (**Table 5**).

Performance of machine learning radiomics in external validation cohort

To further test the reproductivity of the machine learning radiomics model, a cohort with 40 TNBC patients who underwent neoadjuvant therapy were introduced as the external validation cohort. As shown in Table S2, 21 DFS events were observed. The Radiomics + clinicopathological + SMOTEENN model showed an AUC of 0.84 (95% CI 0.69-0.94), with a high sensitivity of 81.0% (**Table 5** and **Figure 5**) in the external validation cohort.

Discussion

In this current study, we established and validated a novel machine learning-based model by combining ultrasound radiomics, clinicopathological features and data sampling method SMOTEENN for DFS prediction in TNBC patients, which had a high AUC value of 0.90, exhibiting a significant better performance than models based only on radiomics or resampled radiomics features.

Ultrasound is one of the most prevailing imaging techniques in breast cancer screening and diagnosis [9]. Traditionally, breast ultrasound was recorded according to BI-RADS system and conventional sonographic features including orientation, shape, margin, posterior acoustic patterns were evaluated, which could reveal certain biological features of breast cancer. Thus, researcher has explored the role of breast ultrasound in predictive and prognostic value in breast cancer patients. Vandana D et al. reported that pathological features combined with sonographic features including well-circumscribed oval mass, vascularity and posterior enhancement were able to predict Oncotype Dx Recurrence Score (r=0.79) with a sensitivity of 89% and specificity of 83% [26]. Our previous study found that the feature of vertical ori-

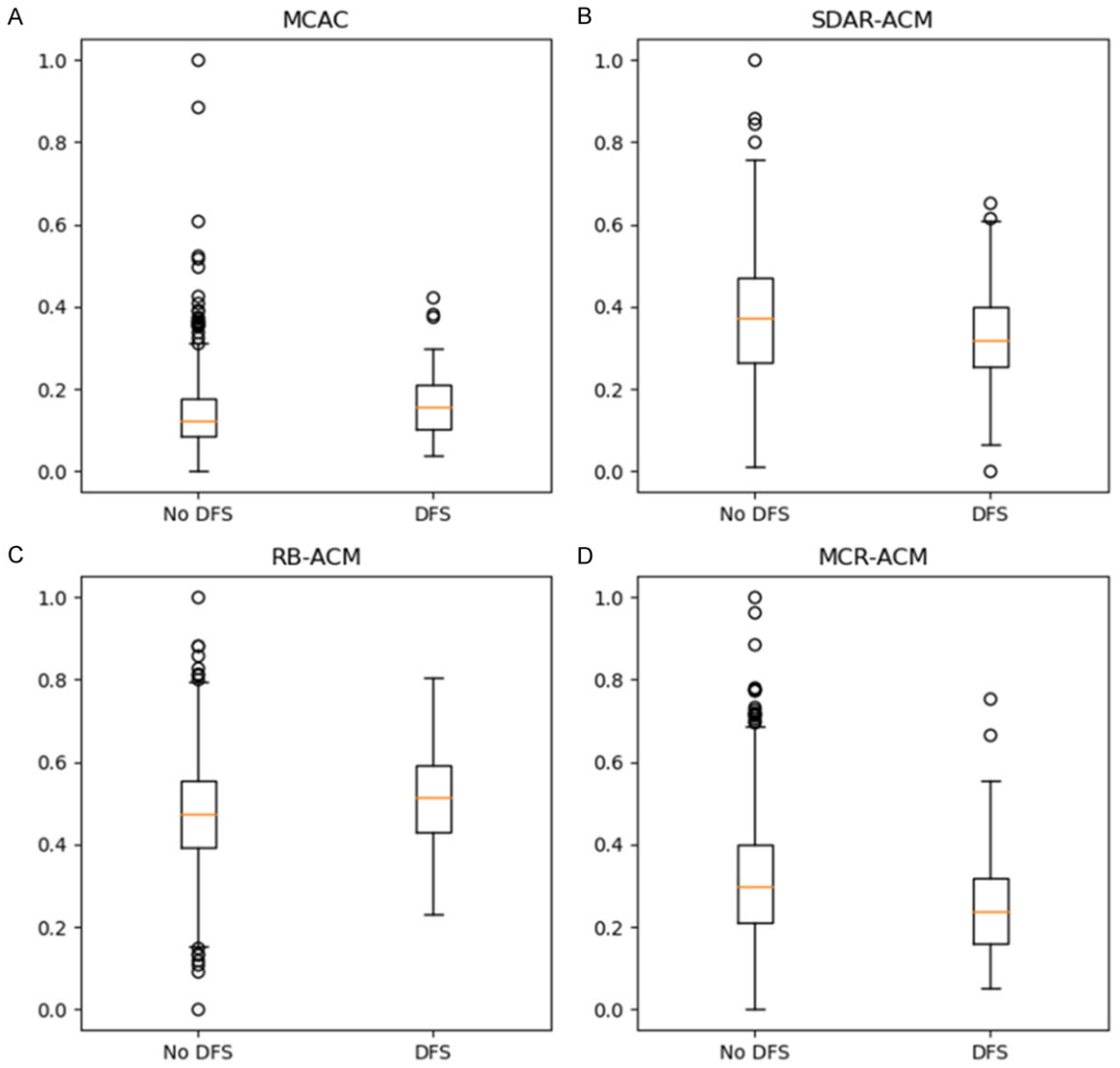

Figure 4. The boxplots of four representative radiomics features. A. MCAC: Mean of the contrast of the internal and external region autocorrelation coefficients. B. SDAR-ACM: Standard deviation of annular region based on approximation coefficients matrix. C. RB-ACM: Relative brightness between inner region and Annular region based on approximation coefficients matrix. D. MCR-ACM: Mean of covariance in ROI based on approximation coefficients matrix.

entation in preoperative ultrasound was an independently risk factor for inferior disease outcome in TNBC patients [12, 13], indicating a promising value for conventional sonographic features in predicting long-term prognosis. However, traditional ultrasound images were assessed by radiologists, which may lead to relatively large inter-observer variability and bad reproducibility [27].

Being able to extract large scales of quantitative features from medical images, radiomics has shown great advantages and optimistic prospective in translational studies of breast cancer [28]. Radiomics studies have focused on the roles of radiological features in aiding breast cancer diagnosis, characterization, and prediction [14, 29]. Our team has established a novel automatic radiomics approach which provided 463 features from conventional breast ultrasound images, which demonstrated a strong correlation between receptor status and molecular subtypes with an AUC of 0.760 ($P< 0.05$) [15, 16]. Efforts also have been made in predicting disease outcomes with radiomics. Hyunjin Park et al. constructed a nomogram

Table 4. Performance comparisons among different classifiers of Radiomics

Classifiers	Dataset	AUC	ACC (%)	SENS (%)	SPEC (%)
Naive Bayes	Training	0.61	84.4	13.9	94.6
	Independent-test	0.69	81.5	18.2	67.6
SVM	Training	0.58	87.3	0.0	100.0
	Independent-test	0.63	89.4	0.0	99.0
Decision Tree	Training	0.48	74.2	10.5	83.4
	Independent-test	0.49	81.4	9.1	89.2
Bagging	Training	0.61	83.5	8.8	94.4
	Independent-test	0.66	88.5	18.2	96.1
RUS Boost	Training	0.58	69.3	36.7	74.0
	Independent-test	0.56	76.1	18.2	82.4

Abbreviations: AUC, areas under curve; ACC, accuracy; SENS, sensitivity; SPEC, specificity; SVM, support vector machines; RUS, random under-sampling.

Table 5. Performance of different models in predicting DFS events for TNBC

Models	Dataset	AUC	ACC (%)	SENS (%)	SPEC (%)
US only	T	0.61 [0.55, 0.67]	84.4 [81.4, 87.4]	13.9 [9.9, 17.9]	94.6 [91.5, 97.8]
	I-T	0.69 [0.60, 0.78]	81.5 [73.1, 88.2]	18.2 [39.0, 94.0]	67.6 [57.7, 76.6]
	E-V	0.51 [0.35, 0.67]	45.0 [29.3, 61.5]	9.5 [1.2, 30.4]	84.2 [60.4, 96.6]
CP only	T	0.67 [0.54, 0.80]	84.0 [80.5, 87.5]	20.8 [7.1, 34.4]	93.1 [89.3, 96.9]
	I-T	0.79 [0.70, 0.86]	88.5 [81.1, 93.7]	54.5 [23.4, 83.3]	93.1 [86.4, 97.2]
	E-V	0.70 [0.54, 0.84]	57.5 [40.9, 73.0]	42.9 [21.8, 66.0]	73.7 [48.8, 90.9]
US + CP	T	0.65 [0.54, 0.75]	84.6 [79.4, 89.9]	25.6 [4.2, 47.0]	93.1 [89.0, 97.3]
	I-T	0.86 [0.78, 0.92]	91.3 [84.5, 95.8]	63.6 [30.8, 89.1]	95.1 [88.9, 98.4]
	E-V	0.77 [0.61, 0.89]	65.0 [48.3, 79.4]	81.0 [58.1, 94.6]	47.4 [54.5, 71.1]
US + SMOTEENN	T	0.84 [0.82, 0.86]	73.1 [69.8, 76.4]	83.5 [80.4, 86.6]	70.1 [65.9, 74.3]
	I-T	0.73 [0.64, 0.81]	59.3 [49.6, 68.4]	90.9 [58.7, 99.8]	54.90 [44.7, 64.8]
	E-V	0.70 [0.54, 0.84]	55.0 [38.5, 70.7]	33.3 [14.6, 57.0]	79.0 [54.4, 94.0]
CP + SMOTEENN	T	0.81 [0.76, 0.85]	73.1 [69.8, 76.4]	47.8 [41.5, 54.2]	90.6 [86.6, 94.7]
	I-T	0.80 [0.72, 0.87]	88.5 [81.1, 93.7]	54.5 [23.4, 83.3]	93.1 [86.4, 97.2]
	E-V	0.79 [0.64, 0.91]	65.0 [48.3, 79.4]	81.0 [58.1, 94.6]	47.4 [24.5, 71.1]
US + CP + SMOTEENN	T	0.86 [0.84, 0.88]	76.5 [72.2, 80.9]	74.7 [68.4, 81.0]	80.1 [78.0, 82.2]
	I-T	0.90 [0.83, 0.95]	82.3 [74.0, 88.8]	81.8 [48.2, 97.7]	82.3 [73.6, 89.2]
	E-V	0.84 [0.69, 0.94]	77.5 [61.6, 89.2]	81.0 [58.1, 94.6]	73.7 [48.8, 90.9]

Abbreviations: US, ultrasound; CP, clinicopathological; AUC, areas under curve; ACC, accuracy; SENS, sensitivity; SPEC, specificity; T, Training; I-T, Independent-test; E-V, external-validation; DFS, disease-free survival.

combining MRI radiomics and clinicopathological features to successfully estimate DFS for breast cancer patients with a C index of 0.76 [30]. Similarly, a radiomic signature based on MRI developed by Yunfang Yu et al. managed to predict 3-year DFS with an AUC of 0.73 in validation cohort [31]. However, most of the current studies focused on MRI, whether ultrasound radiomics could predict prognosis for breast cancer patients still lack convincing evidences. In current study, we used 20 radiomics features to build a model in TNBC patients, which found with a moderate accuracy with AUC value 0.61-0.69 in DFS events prediction which mainly due to relatively low incidence of DFS events, indicating that a new algorithm needs to be investigated to overcome class imbalance.

In this current study, SMOTEENN, a hybrid sampling method to optimize the imbalanced positive classification, was applied to predict disease outcome. In addition, as studies have revealed that molecular subtypes of breast

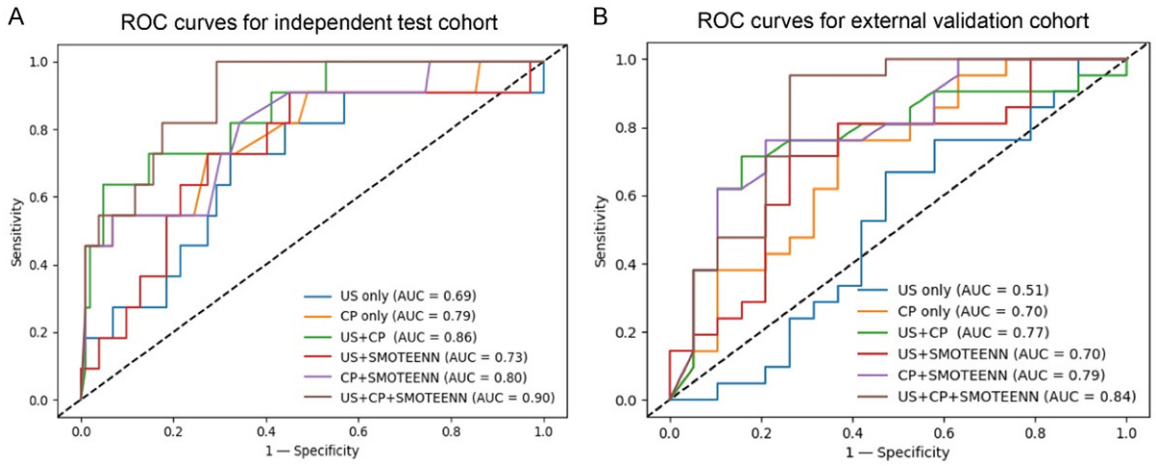

Figure 5. ROC curves of different prognostic models based on Naïve Bayes Classifier in independent test and external validation cohort. ROC curves of 6 different Machine Learning Models in (A) the independent test cohort and (B) the external validation cohort were demonstrated and compared.

cancer may have an impact on signatures of ultrasound radiomics [32, 33], we focused on the TNBC in current study to avoid the interference of radiological heterogeneity. We found that the model combining ultrasound radiomics, clinicopathological features, and data sampling method SMOTEENN had a significantly higher AUC (0.90) value compared with models based on radiomics only (AUC=0.69) or radiomics + SMOTEENN (AUC=0.73) in the independent test cohort, indicating data sampling method SMOTEENN could significantly improve the predictive performance of ultrasound radiomics in predicting DFS events. To our knowledge, this is the first study that established a machine learning-based radiomics model based on preoperative ultrasound and data sampling method to in predict DFS in TNBC patients.

In our study, a total of 20 radiomics features were selected. Briefly, the spiculation of the tumor was selected from the morphological features, which quantifies the degree of irregularity and roughness of the tumor boundary. Roughness of the tumor boundary could imply that the tumor has invaded surrounding tissue [34] and thus could be associated with poor survival. The wavelet features were calculated from the histogram-based and texture features of the single-level discrete 2-D wavelet transform. The low-frequency information features of the ultrasound image were extracted, as well as the high-frequency information features in the horizontal, vertical, and diagonal directions. 15 of the 20 selected features were obtained from the wavelet transformed images, which indicate the importance of radiomics that they redisplay the texture characters and show discriminative ability [16, 35].

Our study has several strengths. First of all, this is the first and largest study to predict long-term prognosis based on machine learning ultrasound radiomics in TNBC patients. Notably, our integrated model showed a stable performance and promising potency with a highest AUC of 0.90. Secondly, to overcome the possible imbalance caused by relatively few events, the hybrid sampling method 'SMOTEENN' was innovatively introduced to our machine learning radiomics model. Last but not least, compared with other examination methods as MRI and tomography, ultrasound was more approachable and affordable in clinical practice. Thus, our machine learning model based on sonographic radiomics tended to have a broader application scenario, which showed potential in risk stratification and precision medicine for TNBC patients.

Several limitations have to be mentioned in our study. Firstly, the study was based on a retrospectively enrolled cohort within a single center, which may unavoidably bring selection bias to the analysis. Secondly, the number of DFS events was relatively small due to the early diagnosis and standardized treatment of TNBC, indicating that a prospectively designed study in larger cohorts with longer follow-up time is warranted to further validate the performance of our integrated model. Thirdly, due to

the lack of available data, novel prognostic biomarkers as tumor infiltrating lymphocytes (TILs), which may further increase efficacy of the model, were unable to be taken into modeling. Last but not least, TNBC can be further divided into several classifications based on genomic expression level [6, 36], including 2 basal-like, immunomodulatory, mesenchymal, mesenchymal stem-like and luminal androgen receptor subtypes. The accuracy of machine learning radiomics in these certain subtypes was not known, which needs further evaluation.

Conclusion

Novel machine learning-based radiomics of preoperative ultrasound combined with clinicopathological features can predict DFS in TNBC patients, warranting further studies validation.

Acknowledgements

The authors thank all participating patients and clinicians for contributing data to this study. This work was funded by financial support from the National Natural Science Foundation of China (81772797, 82072937, 82072897, 81627804 and 81830058); Shanghai Municipal Education Commission-Gaofeng Clinical Medicine Grant Support (20172007); the Science and Technology Commission of Shanghai Municipality (Grant 20DZ1100104); Shanghai Sailing Program 21YF1427400.

Disclosure of conflict of interest

None.

Address correspondence to: Xiaosong Chen, Comprehensive Breast Health Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, No. 197 Ruijin Second Road, Shanghai 200025, China. E-mail: chenxiaosong0156@hotmail.com; Yi Guo and Yuanyuan Wang, Department of Electronic Engineering, Fudan University, 220 Handan Road, Shanghai 200433, China. E-mail: guoyi@fudan.edu.cn (YG); yywang@fudan.edu.cn (YYW)

References

[1] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020; 70: 7-30.

[2] Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thürlimann B and Senn HJ; Panel members. Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. Ann Oncol 2009; 20: 1319-1329.

[3] Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B and Senn HJ; Panel members. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2013. Ann Oncol 2013; 24: 2206-2223.

[4] Foulkes WD, Smith IE and Reis-Filho JS. Triple-negative breast cancer. N Engl J Med 2010; 363: 1938-1948.

[5] Bianchini G, Balko JM, Mayer IA, Sanders ME and Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol 2016; 13: 674-690.

[6] Jiang YZ, Ma D, Suo C, Shi J, Xue M, Hu X, Xiao Y, Yu KD, Liu YR, Yu Y, Zheng Y, Li X, Zhang C, Hu P, Zhang J, Hua Q, Zhang J, Hou W, Ren L, Bao D, Li B, Yang J, Yao L, Zuo WJ, Zhao S, Gong Y, Ren YX, Zhao YX, Yang YS, Niu Z, Cao ZG, Stover DG, Verschraegen C, Kaklamani V, Daemen A, Benson JR, Takabe K, Bai F, Li DQ, Wang P, Shi L, Huang W and Shao ZM. Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. Cancer Cell 2019; 35: 428-440, e425.

[7] Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P and Narod SA. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 2007; 13: 4429-4434.

[8] Penault-Llorca F and Viale G. Pathological and molecular diagnosis of triple-negative breast cancer: a clinical perspective. Ann Oncol 2012; 23 Suppl 6: vi19-22.

[9] Bevers TB, Helvie M, Bonaccio E, Calhoun KE, Daly MB, Farrar WB, Garber JE, Gray R, Greenberg CC, Greenup R, Hansen NM, Harris RE, Heerdt AS, Helsten T, Hodgkiss L, Hoyt TL, Huff JG, Jacobs L, Lehman CD, Monsees B, Niell BL, Parker CC, Pearlman M, Philpotts L, Shepardson LB, Smith ML, Stein M, Tumyan L, Williams C, Bergman MA and Kumar R. Breast cancer screening and diagnosis, version 3.2018, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2018; 16: 1362-1389.

[10] Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH and Shah K. Supplemental screening for breast cancer in women with dense breasts: a systematic review for the U.S. preventive services task force. Ann Intern Med 2016; 164: 268-278.

[11] Kim SY, Han BK, Kim EK, Choi WJ, Choi Y, Kim HH and Moon WK. Breast cancer detected at

screening US: survival rates and clinical-pathologic and imaging factors associated with recurrence. Radiology 2017; 284: 354-364.
[12] Wang H, Zhan W, Chen W, Li Y, Chen X and Shen K. Sonography with vertical orientation feature predicts worse disease outcome in triple negative breast cancer. Breast 2020; 49: 33-40.
[13] Wang H, Yao J, Zhu Y, Zhan W, Chen X and Shen K. Association of sonographic features and molecular subtypes in predicting breast cancer disease outcomes. Cancer Med 2020; 9: 6173-6185.
[14] Valdora F, Houssami N, Rossi F, Calabrese M and Tagliafico AS. Rapid review: radiomics and breast cancer. Breast Cancer Res Treat 2018; 169: 217-229.
[15] Hu Y, Qiao M, Guo Y, Wang Y, Yu J, Li J and Chang C. Reproducibility of quantitative high-throughput BI-RADS features extracted from ultrasound images of breast cancer. Med Phys 2017; 44: 3676-3685.
[16] Guo Y, Hu Y, Qiao M, Wang Y, Yu J, Li J and Chang C. Radiomics analysis on ultrasound for prediction of biologic behavior in breast invasive ductal carcinoma. Clin Breast Cancer 2018; 18: e335-e344.
[17] Le EPV, Wang Y, Huang Y, Hickman S and Gilbert FJ. Artificial intelligence in breast imaging. Clin Radiol 2019; 74: 357-366.
[18] Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R and Sieh W. Deep learning to improve breast cancer detection on screening mammography. Sci Rep 2019; 9: 12495.
[19] Romeo V, Cuocolo R, Apolito R, Stanzione A, Ventimiglia A, Vitale A, Verde F, Accurso A, Amitrano M, Insabato L, Gencarelli A, Buonocore R, Argenzio MR, Cascone AM, Imbriaco M, Maurea S and Brunetti A. Clinical value of radiomics and machine learning in breast ultrasound: a multicenter study for differential diagnosis of benign and malignant lesions. Eur Radiol 2021; 31: 9511-9519.
[20] Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, Mao R, Li F, Xiao Y, Wang Y, Hu Y, Yu J and Zhou J. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun 2020; 11: 1236.
[21] Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M, Hicks DG, Lester S, Love R, Mangu PB, McShane L, Miller K, Osborne CK, Paik S, Perlmutter J, Rhodes A, Sasano H, Schwartz JN, Sweep FC, Taube S, Torlakovic EE, Valenstein P, Viale G, Visscher D, Wheeler T, Williams RB, Wittliff JL and Wolff AC. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol 2010; 28: 2784-2795.
[22] Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, Bilous M, Ellis IO, Fitzgibbons P, Hanna W, Jenkins RB, Press MF, Spears PA, Vance GH, Viale G, McShane LM and Dowsett M. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. J Clin Oncol 2018; 36: 2105-2122.
[23] Batista GEAPA, Prati RC and Monard MC. A study of the behavior of several methods for balancing machine learning training data. 2004; 6: 20-29.
[24] Chawla NV, Bowyer KW, Hall LO and Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 2002; 16: 321-357.
[25] DeLong ER, DeLong DM and Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44: 837-845.
[26] Dialani V, Gaur S, Mehta TS, Venkataraman S, Fein-Zachary V, Phillips J, Brook A and Slanetz PJ. Prediction of low versus high recurrence scores in estrogen receptor-positive, lymph node-negative invasive breast cancer on the basis of radiologic-pathologic features: comparison with oncotype dx test recurrence scores. Radiology 2016; 280: 370-378.
[27] Hooley RJ, Scoutt LM and Philpotts LE. Breast ultrasonography: state of the art. Radiology 2013; 268: 642-659.
[28] Gillies RJ, Kinahan PE and Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278: 563-577.
[29] Conti A, Duggento A, Indovina I, Guerrisi M and Toschi N. Radiomics in breast cancer classification and prediction. Semin Cancer Biol 2020; 72: 238-250.
[30] Park H, Lim Y, Ko ES, Cho HH, Lee JE, Han BK, Ko EY, Choi JS and Park KW. Radiomics signature on magnetic resonance imaging: association with disease-free survival in patients with invasive breast cancer. Clin Cancer Res 2018; 24: 4705-4714.
[31] Yu Y, Tan Y, Xie C, Hu Q, Ouyang J, Chen Y, Gu Y, Li A, Lu N, He Z, Yang Y, Chen K, Ma J, Li C, Ma M, Li X, Zhang R, Zhong H, Ou Q, Zhang Y, He Y, Li G, Wu Z, Su F, Song E and Yao H. Development and validation of a preoperative magnetic resonance imaging radiomics-based signature to predict axillary lymph node metastasis and disease-free survival in patients with ear-

ly-stage breast cancer. JAMA Netw Open 2020; 3: e2028086.
[32] Jaber MI, Song B, Taylor C, Vaske CJ, Benz SC, Rabizadeh S, Soon-Shiong P and Szeto CW. A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. Breast Cancer Res 2020; 22: 12.
[33] Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, Perou CM, Troester MA and Niethammer M. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. NPJ Breast Cancer 2018; 4: 30.
[34] Huang SF, Chang RF, Chen DR and Moon WK. Characterization of spiculation on ultrasound lesions. IEEE Trans Med Imaging 2004; 23: 111-121.
[35] Sudarshan VK, Mookiah MR, Acharya UR, Chandran V, Molinari F, Fujita H and Ng KH. Application of wavelet techniques for cancer diagnosis using ultrasound images: a review. Comput Biol Med 2016; 69: 97-111.
[36] Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y and Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011; 121: 2750-2767.

Table S1. Radiomics features involved in machine learning radiomics

No.	Feature categories	Feature names
1	Morphological	Spiculation
2	Histogram-based	Median absolute deviation
3	Texture	Mean of the contrast of the internal and external region autocorrelation coefficients
4		Mean of contrast in ROI
5		Mean of covariance in ROI
6	Wavelet	Standard deviation of annular region based on approximation coefficients matrix
7		Relative brightness between Inner region and annular region based on approximation coefficients matrix
8		Standard deviation of contrast in ROI based on approximation coefficients matrix
9		Mean of covariance in ROI based on approximation coefficients matrix
10		Standard deviation of covariance in ROI based on approximation coefficients matrix
11		Energy base on horizontal detail coefficients
12		Skewness based on horizontal detail coefficients
13		Histogram skewness based on horizontal detail coefficients
14		Variance contrast of Inside and outside based on horizontal detail coefficients
15		Energy base on vertical detail coefficients
16		Relative brightness between inner region and annular region based on diagonal detail coefficients
17		Sum variance based on diagonal detail coefficients
18		Information measure of correlation 1 based on diagonal detail coefficients
19		High gray based on diagonal detail coefficients
20		High gray-Level Zone Emphasis based on diagonal detail coefficients

Abbreviation: ROI, region of interest.

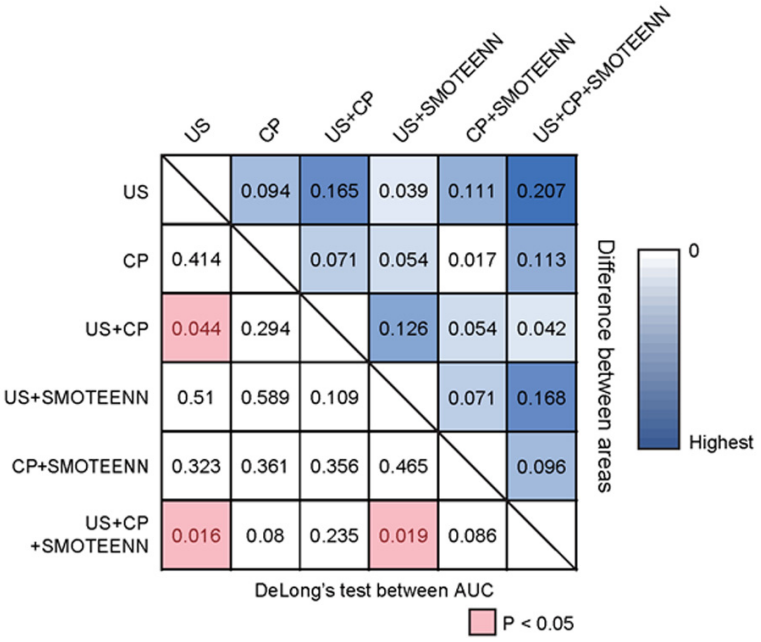

Figure S1. Comparison between different predictive models in independent test cohort. The difference between areas and *p* value of Delong's test of each two machine learning models was demonstrated. For difference between areas, the largest difference was painted for deep blue while minimum for white. *P* values with significant importance were painted as pink. Abbreviations: US, ultrasound; CP, clinicopathological; AUC, areas under curve.

Table S2. Clinicopathological features of external validation cohorts

Variables		Number (%)
Age (yr)	≤55	24 (60.0)
	>55	16 (40.0)
Menstruation status	Pre-/peri-	18 (45.0)
	Post-	22 (55.0)
Breast surgery	BCS	5 (12.5)
	Mastectomy	35 (87.5)
Axillary surgery	SLNB	5 (12.5)
	ALND	35 (87.5)
Tumor size	Mean ± SE	2.9±0.3
	≤2 cm	18 (45.0)
	>2 cm	22 (55.0)
ALN metastases	No	12 (30.0)
	Yes	28 (70.0)
Nuclear grade	I-II	10 (25.0)
	III	25 (62.5)
	NA	5 (12.5)
Ki-67 (%)	Mean ± SE	54.2±1.2
	≤30	16 (40.0)
	>30	24 (60.0)
LVI	No	32 (80.0)
	Yes	8 (20.0)
DFS events	No	19 (47.5)
	Yes	21 (52.5)

Abbreviations: DFS, disease-free survival; BCS, breast conserving surgery; SLNB, sentinel lymph node biopsy; ALND, axillary lymph node dissection; IDC, invasive ductal carcinoma; SE, standard error; ALN, axillary lymph node; NA, not available; LVI, lymphatic vascular invasion; TNM, tumor node metastasis.