European Radiology (2022) 32:1590–1600
https://doi.org/10.1007/s00330-021-08224-x

BREAST

Prediction for pathological and immunohistochemical characteristics of triple-negative invasive breast carcinomas: the performance comparison between quantitative and qualitative sonographic feature analysis

Jia-wei Li[1,2] · Yu-cheng Cao[3] · Zhi-jin Zhao[1,2] · Zhao-ting Shi[1,2] · Xiao-qian Duan[3] · Cai Chang[1,2] · Jian-gang Chen[3]

Received: 6 March 2021 / Revised: 28 June 2021 / Accepted: 15 July 2021 / Published online: 14 September 2021

Abstract

Objective Sonographic features are associated with pathological and immunohistochemical characteristics of triple-negative breast cancer (TNBC). To predict the biological property of TNBC, the performance using quantitative high-throughput sonographic feature analysis was compared with that using qualitative feature assessment.

Methods We retrospectively reviewed ultrasound images, clinical, pathological, and immunohistochemical (IHC) data of 252 female TNBC patients. All patients were subgrouped according to the histological grade, Ki67 expression level, and human epidermal growth factor receptor 2 (HER2) score. Qualitative sonographic feature assessment included shape, margin, posterior acoustic pattern, and calcification referring to the Breast Imaging Reporting and Data System (BI-RADS). Quantitative sonographic features were acquired based on the computer-aided radiomics analysis. Breast cancer masses were manually segmented from the surrounding breast tissues. For each ultrasound image, 1688 radiomics features of 7 feature classes were extracted. The principal component analysis (PCA), least absolute shrinkage and selection operator (LASSO), and support vector machine (SVM) were used to determine the high-throughput radiomics features that were highly correlated to biological properties. The performance using both quantitative and qualitative sonographic features to predict biological properties of TNBC was represented by the area under the receiver operating characteristic curve (AUC).

Results In the qualitative assessment, regular tumor shape, no angular or spiculated margin, posterior acoustic enhancement, and no calcification were used as the independent sonographic features for TNBC. Using the combination of these four features to predict the histological grade, Ki67, HER2, axillary lymph node metastasis (ALNM), and lymphovascular invasion (LVI), the AUC was 0.673, 0.680, 0.651, 0.587, and 0.566, respectively. The number of high-throughput features that closely correlated with biological properties was 34 for histological grade (AUC 0.942), 27 for Ki67 (AUC 0.732), 25 for HER2 (AUC 0.730), 34 for ALNM (AUC 0.804), and 34 for LVI (AUC 0.795).

Conclusion High-throughput quantitative sonographic features are superior to traditional qualitative ultrasound features in predicting the biological behavior of TNBC.

Key Points • *Sonographic appearances of TNBCs showed a great variety in accordance with its biological and clinical characteristics.*

• *Both qualitative and quantitative sonographic features of TNBCs are associated with tumor biological characteristics.*

• *The quantitative high-throughput feature analysis is superior to two-dimensional sonographic feature assessment in predicting tumor biological property.*

Keywords Breast · Triple-negative breast neoplasms · Ultrasonography · Radiomics

Abbreviations

ABUS	Automated breast ultrasonography
ALNM	Axillary lymph node metastasis
AUC	Area under the curve
BI-RADS	Breast imaging – report and data system
BLIS	Basal-like and immune-suppressed
CI	Confidence interval

Jia-wei Li, Yu-cheng Cao, and Zhi-jin Zhao contributed equally to this work. Jian-gang Chen is the first corresponding author and Cai Chang is the second corresponding author.

Extended author information available on the last page of the article

DICOM	Digital imaging and communication in medicine
ER	Estrogen receptor
FISH	Fluorescent in situ hybridization
GLCM	Gray level co-occurrence matrix
GLDM	Gray level dependence matrix
GLRLM	Gray level run length matrix
GLSZM	Gray level size zone matrix
HE	Hematoxylin-eosin
HER2	Human epidermal growth factor receptor-2
IHC	Immunohistochemical
IM	Immunomodulatory
IQR	Interquartile range
LAR	Luminal androgen receptor
LASSO	Least absolute shrinkage and selection operator
LBP2D	Local binary pattern 2-dimensional
LoG	Laplacian of Gaussian
LVI	Lymphovascular invasion
MES	Mesenchymal-like
MRI	Magnetic resonance imaging
NGTDM	Neighboring gray tone difference matrix
OR	Odds ratio
PCA	Principal component analysis
PR	Progesterone receptor
ROI	Region of interest
SD	Standard deviation
SVM	Support vector machine
TNBC	Triple-negative breast carcinomas

Introduction

Triple-negative breast cancer (TNBC) is one of the molecular subtypes with the worst prognosis among all breast cancers [1, 2]. It is characteristic of no expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). TNBC is associated with high heterogeneity, aggressive proliferation, and low differentiation [2–4]. As a result of these biomolecular characteristics, the treatment of TNBCs has been a challenge without effective drugs for well-defined molecular targets. The high chance of recurrence and distant metastasis results in a poor prognosis among TNBC patients [1–6].

Ultrasonography is an important imaging tool for breast diseases. In our previous study, we addressed the heterogeneity of sonographic features of TNBCs [7]. The most common sonographic features for TNBCs included regular shape, no angular/spiculated margin, posterior acoustic enhancement, and no calcifications which are characteristic of benign breast masses [7]. These benign-like TNBCs are associated with a high cellular proliferation rate and poor differentiation [7] which indicated a high risk of recurrence and distant metastasis.

Unfortunately, these benign-like TNBC masses especially in young patients are easy to be missed [8]. An early and accurate recognition of this kind of breast tumor with aggressive biological property will, therefore, be beneficial for improving clinical outcomes. Thus, benign-like breast masses require more attention during ultrasound examinations, particularly by less experienced doctors. The challenge for sonographic physicians calls for more advanced methods to improve diagnostic performance.

In the last decade, the concept of radiomics emerges and flourishes. Computer-aided analysis converts imaging information into quantitative numerical data using a series of computational algorithms [9–11]. Numerous studies have confirmed that information obtained from medical images such as ultrasound and magnetic resonance imaging (MRI) is closely related to the characteristics of genes, proteins, and tumor phenotypes [9, 10, 12–15]. However, the assessment of ultrasound images was operator-dependent with large intra- and inter-observer variability [16]. Computer-aided feature analysis is expected to reduce the variability among observers. Our preliminary results showed that the high throughput quantitative analysis for ultrasound images of breast cancers was reliable [17] which could be used to predict the biological behavior of breast cancers [18] and TNBCs [19].

In the present study, the performance using quantitative high-throughput sonographic feature analysis to predict pathological and immunohistochemical (IHC) characteristics of TNBCs was compared with that using qualitative feature assessment.

Materials and methods

We reviewed the clinical data including preoperative ultrasound reports and images, surgical records, postoperative pathological, and IHC results from 6758 patients who accepted breast cancer surgeries at our center from June 2014 to June 2019. Patients presented as a solitary mass on preoperative ultrasound images and confirmed as TNBC by postoperative pathology were eligible for the study. The exclusion criteria included the following: bilateral or multiple masses, recurrences, previous breast cancer surgeries, neoadjuvant chemotherapy, or mass diameter larger than 5 cm and ultrasound images with poor qualities. Finally, 252 eligible TNBC patients were included. Our study acquired ethical approval from the institutional review board at Fudan University Shanghai Cancer Center. Informed consent was waived as the data were retrospectively collected.

Ultrasound equipment and preoperative image acquisition

Ultrasound equipment used in the study included Aixplorer (SuperSonic Imagine), Logic E9 (GE Healthcare), xMATRIX and IU22 (Philips Medical Systems), Aplio 500 (Cannon Medical System), and Mylab90 and MylabTwice (Esaote). The high frequency (5–14 MHz) linear array transducer was used for the scanning of breast masses. Ultrasound images were recorded as the format of digital imaging and communication in medicine (DICOM).

Qualitative sonographic feature assessment

All ultrasound images were reviewed by two experienced US physicians who were blinded to the patients' clinical characteristics and histological results. The sonographic features of TNBC masses were assessed referring to the Breast Imaging Reporting and Data System (BI-RADS) [20]. In the present study, TNBCs were assessed in terms of tumor shape (regular and irregular), angular or spiculated margin (yes and no), posterior acoustic pattern (shadow, enhancement, no change, and mixed change), and calcification (yes and no). A consensus was reached after discussion when there was a disagreement between the two ultrasound physicians.

Image selection for radiomics analysis

One typical ultrasound image which mostly matches with the morphologic description of the breast mass was selected for each patient to perform computer-aided radiomics analysis. Firstly, the region of interest (ROI) delineating margins of the mass was outlined on the selected ultrasound images by the ultrasound physician ZJ Zhao for radiomics feature analysis. The data set was randomly divided into the training set and testing set at a ratio of 7:3. Figure 1 shows the flowchart of the computer-aided radiomics analysis.

Table 1 Radiomics features of each feature class

Feature class	Number of features
1. First order	19
2. Shape	26
3. GLCM	24
4. GLSZM	16
5. GLRLM	16
6. NGTDM	5
7. GLDM	14

Feature extraction from the US image

A total of 1688 high-throughput radiomics features based on the ROI were extracted to analyze the internal heterogeneity from 10 types of images for each ultrasound image. These image types include original, wavelet, Laplacian of Gaussian (LoG), square, square root, exponential, logarithm, gradient, and local binary pattern 2-dimensional (LBP2D). All features are grouped into the following seven feature classes: first order, shape, gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), gray level run length matrix (GLRLM), neighboring gray tone difference matrix (NGTDM), and gray level dependence matrix (GLDM) (Table 1).

Feature selection and classification

To remove the redundant features for the purpose of reducing overfitting, machine learning was used for data analysis to select the most robust radiomics features correlating with the biological features. In this study, the combination of principal component analysis (PCA) [21] and least absolute shrinkage and selection operator (LASSO) [22] algorithms, named PCA + LASSO method, was used to perform feature selection.

The PCA algorithm comprehensively considers all high-throughput features and reduces the feature dimensions according to the selected coefficients. And then the LASSO

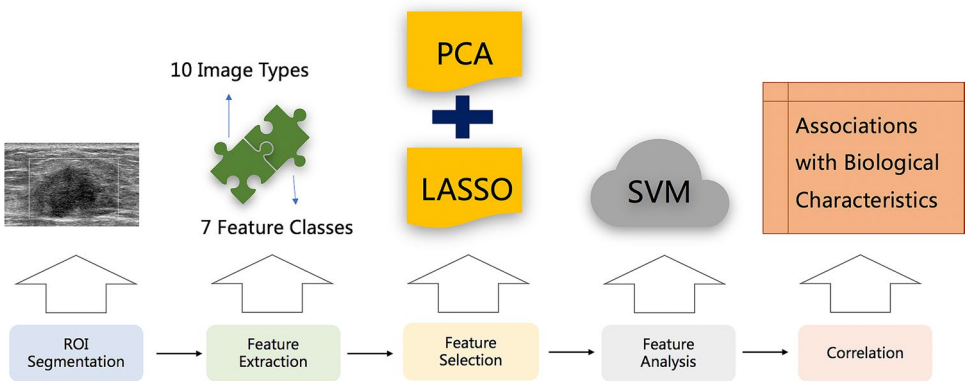

Fig. 1 Flowchart of the computer-aided radiomics analysis

algorithm extracts feature with high correlation with biological characteristics from the reduced-dimensional features. This algorithm is suitable not only for linear cases but also for nonlinear cases. Table 2 shows the number of selected radiomics features after using PCA + LASSO method for each biological property.

After selection, the selected features were input to the support vector machine (SVM) classifier for further classification. Three kinds of classification models were proposed and compared: (I) all features, (II) features selected with the PCA method, and (III) features selected the combined method of PCA + LASSO.

Postoperative pathology and IHC

The postoperative specimens were fixed in formalin, embedded in paraffin, and stained with hematoxylin–eosin (HE) to prepare for histological and IHC assessments. Before the preparation of histological specimen, the tumor size was firstly determined based on the gross sample. The pathological characteristics evaluated by HE staining included pathological type, nuclear grade, status of lymphovascular invasion, papilla invasion, and axillary lymph node metastasis. Based on the pathological type, all patients were divided into infiltrating ductal carcinoma, infiltrating lobular carcinoma, and other types of invasive breast carcinomas. Based on nuclear grades, all patients were divided into three groups: grade I (highly differentiated), grade II (moderately differentiated), and grade III (poorly differentiated).

IHC analysis was performed to determine the expression of ER, PR, HER2, and Ki67. The negative expression of ER and PR was defined as less than 1% nuclei staining. HER2 status was considered as negative when IHC was 0 or 1 + , or HER2 amplification was absent (ratio < 2.2) in the fluorescent in situ hybridization (FISH) test. TNBCs were defined as the negativity of ER, PR, and HER2 in accordance with the St. Gallen International Expert Consensus [23]. The expression level of Ki67 was based on the ratio of the nucleus with positive staining.

Patients were divided into two groups according to the pathological grade: low grade (I and II) and high grade (III); two groups according to Ki67 level: < 40% and ≥ 40%; and two groups according to HER2 score: low score (0 or 1) and high score (2); axillary lymph node metastasis (yes and no); and lymphovascular invasion (yes and no). The subgroup according to HER2 score was based on our previous finding that the higher HER2 score (2 + with FISH negative) was associated with the higher chance of calcifications in TNBCs [7]. The cutoff of 40% was used for Ki67 level of TNBC cohort as the median Ki67 expression level was 60–70% and mean Ki67 level was about 60% according to our previous experience. A cutoff of 20% or 14% in the guideline would cause a bias that most TNBCs cases are of high Ki67 expression (> 20% or 14%). Therefore, in our TNBC-related articles, the cutoff of 40% was used as the criterion for defining Ki67 expression level [7, 24, 25].

Table 2 Number of radiomics features after dimension reduction for each biological property

	Histological grade	Ki67	HER2	ALNM	LVI
All features	1688	1688	1688	1688	1688
After PCA	200	200	200	200	200
After PCA + LASSO	34	27	25	34	34

Statistical analysis

The statistical analyses for qualitative sonographic features were performed using SPSS for Windows version 22.0 (IBM Corp.). Continuous numerical data were presented as mean ± standard deviation SD (range) or median (interquartile range, IQR) after testing the data normality with the Kolmogorov–Smirnov test. Categorical data were presented as frequency (percentage %, 95% confidence interval CI). Multivariate logistic regression analysis was used to identify independent qualitative sonographic features that were correlated with the pathological characteristics of TNBCs. Odds ratio (OR) was calculated for the qualitative sonographic feature with the best predicative value.

Machine learning was used for analyzing the high-throughput features. The PCA algorithm was used for dimension reduction, which projected the feature space owning all 1688 features into a smaller subspace and ensured the loss of feature information when the overall influence of the original subject information is not large. The LASSO algorithm was used to select contributory variables from the afore-obtained features of reduced dimension by adding penalty terms. In addition, the small coefficients were compressed to be 0 with insignificant variables discarded. Figure 2 shows the coefficients of selected features in the LASSO model.

The predictive efficiency for pathological and IHC data using quantitative and qualitative sonographic features was evaluated by sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) [26]. The two-tailed p value less than 0.05 was considered statistically significant.

Results

Table 3 shows the demographics, surgical information, postoperative pathology, and immunohistochemistry of the 252 patients. Nine patients were excluded for histological grade–related analysis due to missing data. All

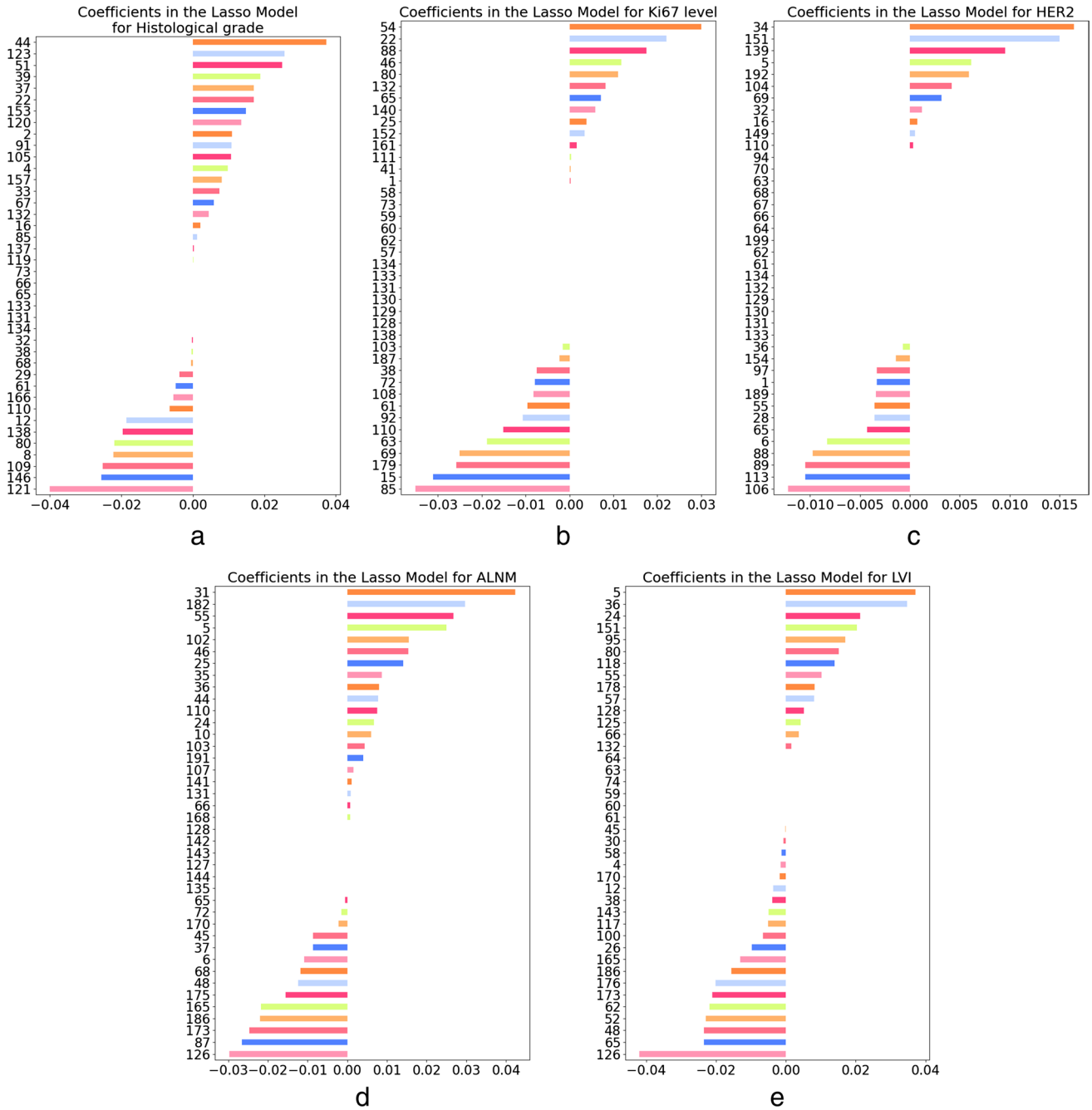

Fig. 2 The confusion matrix of high-throughput radiomics features analysis. Column: the predicted category; Row: the true category of the data. **a** Histological grade; **b** Ki67; **c** HER2; **d** ALNM; **e** LVI

252 TNBC patients had an average age of 50.9 years old (SD 11.7, range 22–82) and an average tumor size of 2.3 cm (SD 0.8, range 0.8–4.9). Most patients were infiltrative ductal carcinoma (96.0%) with high histological grade (84.0%), high Ki67 level (median 70%), but without axillary lymph node metastasis (68.7%) or lymphovascular invasion (67.5%).

The combination of four sonographic features including regular tumor shape, no angular or spiculated margin, posterior acoustic enhancement, and no calcification was used to predict the clinical, histological, and IHC characteristics of TNBCs as shown in Table 4. The AUC was 0.673 ($p=0.001$) for histological grade, 0.680 ($p<0.0005$) for Ki67 level, 0.651 ($p=0.01$) for HER2, 0.587 for ALNM ($p=0.026$), and 0.566 ($p=0.088$) for LVI. Posterior acoustic enhancement

Table 3 Demographics, surgical, and pathological features of 252 patients. Data are presented as mean ± SD (range) for age and tumor size, median (IQR) for Ki67 expression level, and frequency (%, 95% confidence interval, CI) for categorial data

Age (yrs)	50.9 ± 11.7 [22–82]
Tumor size (cm)	2.3 ± 0.8 [0.8–4.9]
Surgical type	
MRM	91 (36.1%, 30.2–42.1%)
M&SLNB	72 (28.6%, 23.0–34.1%)
M&ALND	7 (2.8%, 1.2–5.2%)
BCS&SLNB	61 (24.2%, 19.0–29.4%)
BCS&ALND	21 (8.3%, 4.8–11.9%)
Pathological type	
Infiltrative ductal carcinoma	242 (96.0%, 93.7–98.4%)
Infiltrative lobular carcinoma	2 (0.8%, 0–2.0%)
Other types of infiltrative carcinoma	8 (3.2%, 1.2–5.6%)
Histological grade	
I and II (low grade)	39 (16.0%, 11.5–21.0%)
III (high grade)	204 (84.0%, 79.0–88.5%)
Axillary lymph node metastasis	
Yes	79 (31.3%, 25.8–37.3%)
No	173 (68.7%, 62.7–74.2%)
Lymphovascular invasion	
Yes	82 (32.5%, 26.6–38.5%)
No	170 (67.5%, 61.5–73.4%)
Papilla invasion	
Yes	3 (1.2%, 0–2.8%)
No	249 (98.8%, 97.2–100%)
Ki67 expression value (%)	70 [40–80]
Ki67 expression level	
≤ 40%	57 (22.6%, 17.9–27.8%)
> 40%	195 (77.4%, 72.2–82.1%)
HER2 score	
0	146 (57.9%, 51.6–63.9%)
1	79 (31.3%, 25.8–36.9%)
2	27 (10.7%, 7.1–14.7%)

yrs years old, *MRM* modified radical mastectomy, *M* mastectomy, *SLNB* sentinel lymph node biopsy, *ALND* axillary lymph node dissection, *BCS* breast conservative surgery

was the predominant sonographic feature responsible for high histological grade (OR = 3.81). No angular/spiculated margin was the predominant sonographic feature for high Ki67 expression (OR = 2.31). Calcification was predominantly associated with high HER2 score (OR = 2.48). There was no independent sonographic feature associated with axillary lymph node metastasis.

The performance of radiomics analysis to predict the biological property was displayed in Table 5. Compared with the method of PCA algorithm, the combination of PCA and LASSO algorithms increased the AUC value of each biological property by 46.0–88.4%. Meanwhile, compared with directly using all 1688 features for classification, after using the method of PCA and LASSO, the AUC value for predicting biological features increased 30.0–116.7%. With a certain number of selected features, the AUC was 0.942 for histological grade (34 features), 0.732 for Ki67 (27 features), 0.730 for HER2 (25 features), 0.804 for ALNM (34 features), and 0.795 for LVI (34 features). Figure 3 shows the confusion matrix reflecting accuracy of classification for each biological property. In the confusion matrix, the higher value on the diagonal indicates the greater possibility that the predicted category matched with the actual category. In Fig. 3, the number on the non-main diagonal is very close to 0, which means that the probability of misjudgment in our model's prediction process is very small. For example, for histological grade, the values on the non-main diagonal are 1 and 2, which are much smaller than 11 and 59 on the main diagonal, which means that the model has good performance.

Discussion

While being recognized as an aggressive disease, TNBC-related research has been immensely studied in terms of imaging features, clinical outcome assessment, and therapeutic target exploration [1, 7, 27–30]. As a result of the high heterogeneity of biological property at cellular and genetic levels [28, 29, 31–34], clinical outcome of TNBC is highly diverse among patients. Similarly, the imaging appearances of TNBC showed a great variety in accordance

Table 4 The efficacy of using two-dimensional ultrasound features (regular shape, no angular/spiculated margin, posterior acoustic enhancement, and no calcification) to correlate with pathological and IHC characteristics of TNBCs

	Sensitivity (%)	Specificity (%)	AUC (95% CI)	*p* value	The predominant US feature with OR (95% CI)
Histological grade	51.5	76.9	0.673 (0.585–0.760)	0.001	Posterior acoustic enhancement 3.81 (1.34–10.80)
Ki67 level	74.4	54.4	0.680 (0.600–0.760)	< 0.0005	No angular/spiculated margin 2.31 (1.20–4.48)
HER2	66.7	63.6	0.651 (0.543–0.760)	0.01	Calcification 2.48 (1.03–5.99)
ALNM	68.8	51.9	0.587 (0.514–0.660)	0.026	None
LVI	76.8	36.5	0.566 (0.493–0.640)	0.088	None

AUC area under the curve; *OR* odds ratio; *CI* confidence interval

Table 5 The efficacy of using high-throughput sonographic features to correlate with pathological and IHC characteristics of TNBCs

	Indices	Sensitivity (%)	Specificity (%)	AUC
Histological grade	All features ($n=1688$)	25.0	83.6	0.543
	PCA ($n=200$)	100.0	0	0.50
	PCA + LASSO ($n=34$)	91.6	96.7	0.942
Ki67	All features ($n=1688$)	40.0	72.7	0.563
	PCA ($n=200$)	100.0	0	0.5
	PCA + LASSO ($n=27$)	60.0	86.4	0.732
HER2	All features ($n=1688$)	67.5	0	0.337
	PCA ($n=200$)	100	0	0.5
	PCA + LASSO ($n=25$)	95.9	50.0	0.730
ALNM	All features ($n=1688$)	48.2	50.0	0.491
	PCA ($n=200$)	71.4	20.0	0.457
	PCA + LASSO ($n=34$)	85.7	75.0	0.804
LVI	All features ($n=1688$)	100	0	0.50
	PCA ($n=200$)	71.4	35.0	0.532
	PCA + LASSO ($n=34$)	83.9	75.0	0.795

AUC area under the ROC curve

with its biological and clinical characteristics [7, 27, 35, 36] as illustrated in Figs. 4 and 5. The heterogeneity of sonographic features hindered the early and accurate diagnosis for TNBCs, especially for those TNBCs with benign-like sonographic appearance. In the literature, it was reported that some TNBCs are prone to be confused with fibroadenomas [37]. These TNBCs with benign sonographic features tend to show more proliferative and aggressive biological properties such as high histological grade and high Ki67 expression level [7]. In the present study, we used the quantitative high-throughput feature analysis to further validate the association between sonographic features and biological property.

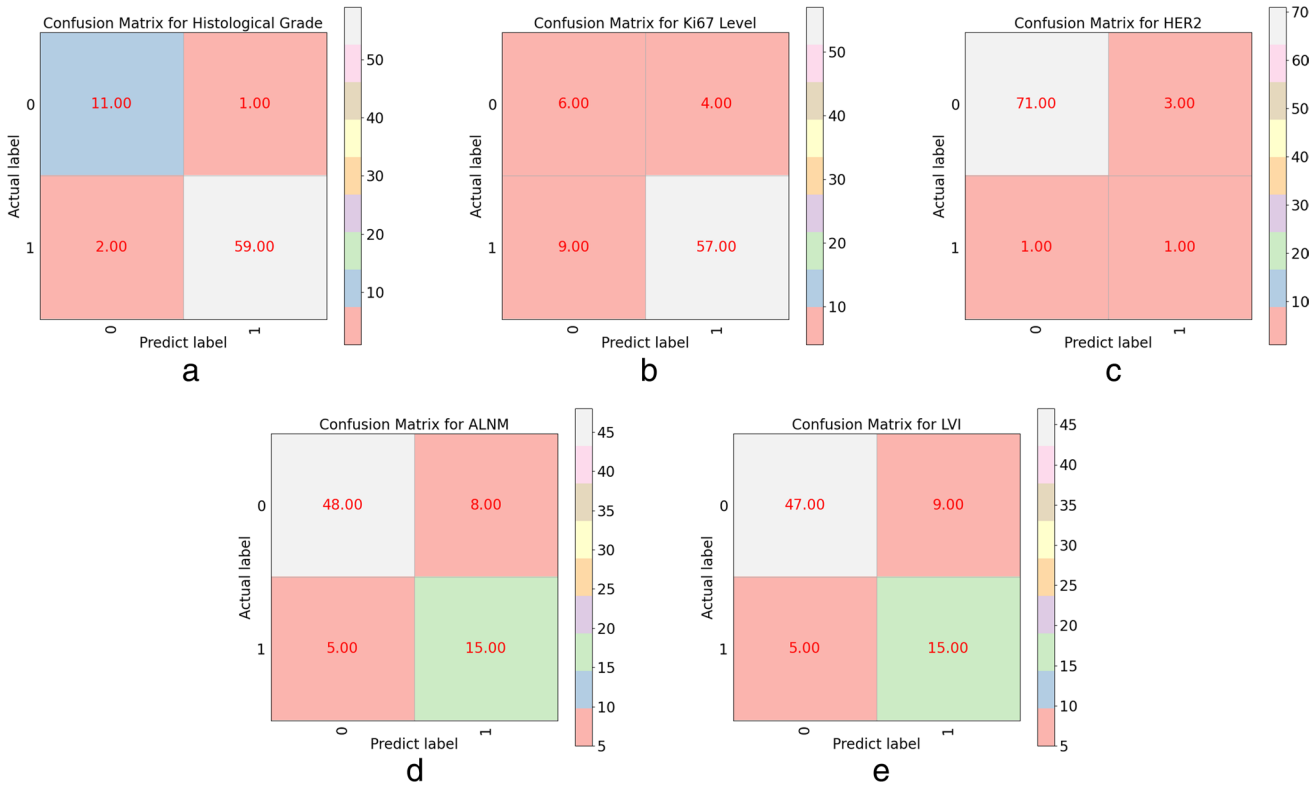

Fig. 3 The coefficients of selected features in the LASSO model. The ordinate lists the serial numbers of certain features, and the corresponding color bars representing the coefficients of selected features show the significance of the corresponding features. **a** Histological grade; **b** Ki67; **c** HER2; **d** ALNM; **e** LVI

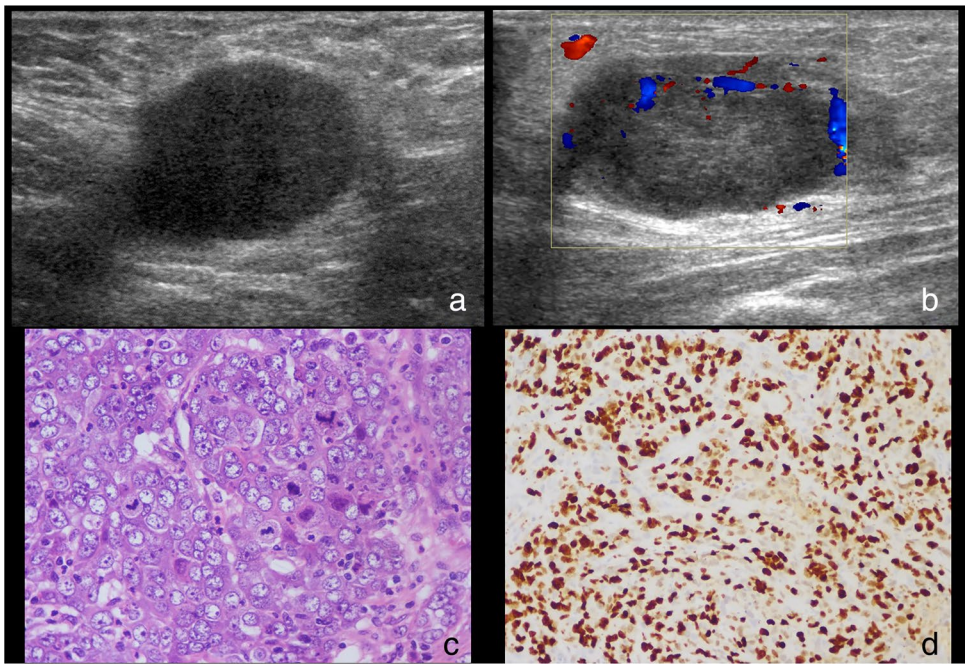

Fig. 4 TNBC with regular tumor shape, circumscribed margin, and posterior acoustic enhancement at sonography in a 27-year-old female patient (BI-RADS: 4A). **a** Gray-scale US image; **b** Color Doppler US image; **c** HE staining showing the nuclear pleomorphism and nuclear mitosis (original magnification × 400, histological grade III); **d** IHC staining for Ki67 quantification (original magnification × 200, Ki67 80%)

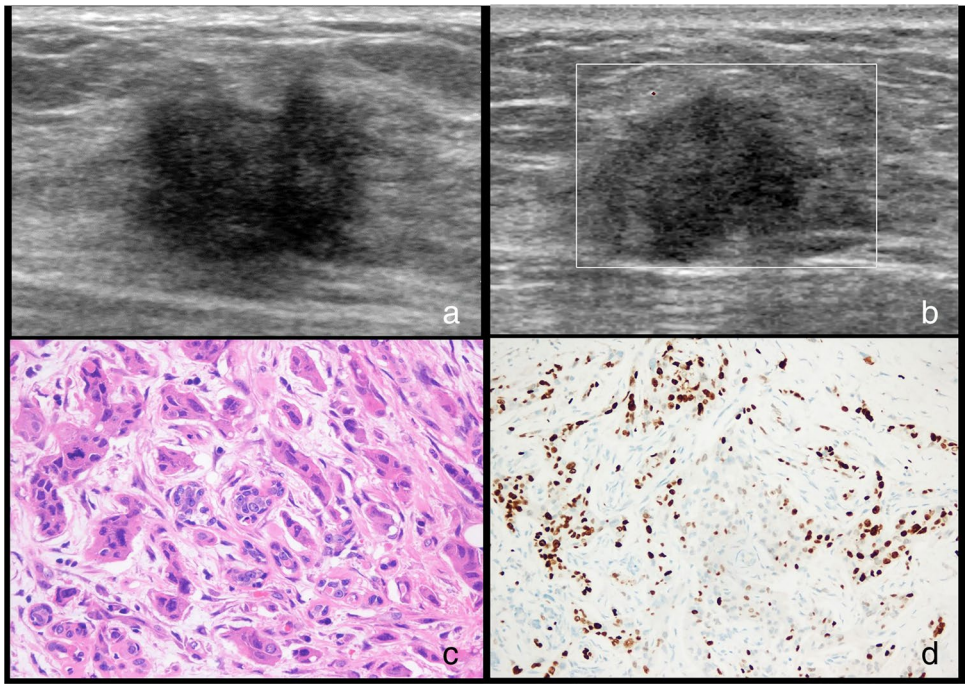

Fig. 5 TNBC with irregular tumor shape, angular, and spiculated margin at sonography in a 48-year-old female patient (BI-RADS: 4C). **a** Gray-scale US image; **b** Color Doppler US image; **c** HE staining showing the sporadic nuclear mitosis (original magnification × 400, histological grade II); **d** IHC staining for Ki67 quantification (original magnification × 200, Ki67 30%)

Our results showed that both qualitative and quantitative sonographic features of TNBCs are associated with tumor biological characteristics. The quantitative high-throughput feature analysis is superior to two-dimensional sonographic feature assessment in predicting tumor biological property.

Most TNBCs have a higher histological grade and higher cellular proliferation rate than other molecular subtypes of breast cancers. The active and rapid growth of TNBCs results in less matrix reaction which leads to clear boundaries between tumors and surrounding normal tissues [38, 39]. Some TNBCs presented inactive growth pattern which allowed enough time to have interaction with host cells, leading to fibrosis, inflammation, and neovascularization [38, 39]. These interactions result in angular burrs on the margins of tumors, which are the areas where normal breast tissue and tumor tissue cross-grow. These differences in growth characteristics lead to variability in sonographic features of TNBC [7].

This is the first study to compare the performance of quantitative and quantitative sonographic feature assessment for TNBCs. The AUC value for quantitative features was higher than that for the qualitative sonographic features. Traditional medical imaging techniques including ultrasonography were based on the general anatomical and morphological images of tissues, organs, or lesions. The images were subjectively interpreted, while the tumor biology-related information hidden in the images was not well-considered [9]. Medical images also contain digitalized information in addition to displaying conventional descriptive imaging signs visualized by naked eyes [9, 10, 40]. High-throughput image analysis is the kernel part in radiomics and artificial intelligence. By digitizing the information hidden in the image, the high-throughput image analysis is capable of associating the imaging phenotype with the tumor biological characteristics [9, 10]. Relevant studies have also confirmed that the information obtained from medical imaging is closely related to the characteristics of genes, proteins, and tumor phenotypes [9, 10]. In the past decade, numerous studies on radiomics and radiogenomics have emerged and have proved that radiomics can assist clinical decision-making in many ways. It has been shown that radiomics analysis of MRI can predict the histological outcome of breast cancers [41] and also predict the therapeutic effect of neoadjuvant chemotherapy [42]. Breast ultrasound automated diagnostic module, named S-detection technology, based on the sonographic appearances of breast masses has been launched and applied with promising clinical results [43]. Our previous studies also demonstrated that high-throughput sonographic features analysis was reliable and reproducible for breast cancers and can assist in predicting the expression of hormone receptors in invasive breast cancers [18]. A recently published study found that radiomics method can provide a high diagnostic performance in the differential diagnosis of fibroadenoma and TNBC [37].

Nowadays, the individualized treatment of TNBCs mainly focuses on the biomolecular characteristics detected by genomics and proteomics [28, 30–33, 44, 45]. TNBCs can be divided into 4–6 subgroups based on cytokeratins [33], transcriptomes [32, 45], or genomics [28, 31] which are associated with the clinical outcomes of TNBCs. Infiltrative tumor border pattern was more associated with the luminal cluster and pushing border pattern was more associated with the basal cluster which showed better clinical outcome compared with the luminal group [33]. This finding might indicate that the malignant like TNBCs might have a poorer prognosis than those with benign like appearances. This is similar as Wang et al found that vertical orientation was associated with worse outcomes and a greater chance of LNM in axilla [46]. However, these results were controversial with our finding that basal-like and immune-suppressed (BLIS) groups, presenting a higher chance to be benign-like sonographic features tended to have a poorer prognosis than other molecular subgroups such as immunomodulatory (IM), luminal androgen receptor (LAR), and mesenchymal-like (MES) TNBCs [25]. But the routine application of proteomics and genomics in clinical practice is still in challenge due to the complicated process and high expenses. Imaging features are always the initial information acquired before the treatment of breast cancers and the acquisition of genomics and proteomics. If there were associations between radiomics, proteomics, and genomics, the imaging characteristics can assist in predicting the biological properties of TNBCs. This will be valuable in treatment decision-making and prognosis prediction. This project has been initiated through collaborations with the Department of Breast Surgery at our cancer center.

Our results should be interpreted after considering the limitations. First, images used for quantitative and qualitative assessment were retrospectively retrieved from the image archives. The still images may not fully depict sonographic features of breast masses. This aspect is very important in traditional ultrasound and should be investigated. A future proposal with the automated breast ultrasonography (ABUS) system may allow to fully explore the breast with an automatic system and to keep the examination entirely [47]. Second, our results were exclusive for tumor size larger than 5 cm and non-invasive breast carcinomas. In the process of feature dimensionality reduction using PCA, the generated new features were the combination of the original dimensions, which fully summarized the information contained in the original feature space. This resulted in the poor interpretation of the generated features. Last, the AUCs for quantitative and qualitative methods were difficult to be compared as they were acquired with different algorithms.

Conclusion

High-throughput sonographic feature analysis using the combination of PCA and LASSO is superior to two-dimensional feature assessment in terms of predicting tumor biological characteristics and clinical behavior of TNBCs.

Acknowledgements This study was funded by the National Natural Science Foundation of China project (Nos. 81627804, 81830058) and Shanghai Anticancer Association SOAR Project (No. SACA-AX201905).

Funding This study was funded by the National Natural Science Foundation of China project (Nos. 81627804, 81830058) and Shanghai Anticancer Association SOAR Project (No. SACA-AX201905).

Declarations

Guarantor The scientific guarantor of this publication is Prof. Jiangang Chen.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Informed written consent was signed by all patients.

Ethical approval Institutional Review Board approval was acquired (Ref No 1802181–22-NSFC). There was no animal-related experiment in the study and therefore, no ethical approval for animal studies was applied.

Study subjects or cohorts overlap There were 252 patients in the present study. A cohort of 100 patients in this study were published in the previous study in *Scientific Reports* in 2018 (reference 7 in the present study). A cohort of 90 patients in this study were published in the *Chinese Journal of Ultrasonography* in 2019 (reference 19 in the present study).

Methodology • retrospective

• diagnostic or prognostic study
• performed at one institution

References s

1. Jiang YZ, Liu Y, Xiao Y et al (2021) Molecular subtyping and genomic profiling expand precision medicine in refractory metastatic triple-negative breast cancer: the FUTURE trial. Cell Res 31:178–186
2. Dent R, Trudeau M, Pritchard KI et al (2007) Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 13:4429–4434
3. Li CY, Zhang S, Zhang XB, Wang P, Hou GF, Zhang J (2013) Clinicopathological and prognostic characteristics of triple-negative breast cancer (TNBC) in Chinese patients: a retrospective study. Asian Pac J Cancer Prev 14:3779–3784
4. Liao HY, Zhang WW, Sun JY, Li FY, He ZY, Wu SG (2018) The clinicopathological features and survival outcomes of different histological subtypes in triple-negative breast cancer. J Cancer 9:296–303
5. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V (2007) Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer 109:1721–1728
6. Abulkhair O, Moghraby JS, Badri M, Alkushi A (2012) Clinicopathologic features and prognosis of triple-negative breast cancer in patients 40 years of age and younger in Saudi Arabia. Hematol Oncol Stem Cell Ther 5:101–106
7. Li JW, Zhang K, Shi ZT et al (2018) Triple-negative invasive breast carcinoma: the association between the sonographic appearances with clinicopathological feature. Sci Rep 8:9040
8. Wojcinski S, Stefanidou N, Hillemanns P, Degenhardt F (2013) The biology of malignant breast tumors has an impact on the presentation in ultrasound: an analysis of 315 cases. BMC Womens Health 13:47
9. Aerts H, Velazquez ER, Leijenaar RTH et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006
10. Gillies RJ, Kinahan PE, Hricak H (2015) Radiomics: Images Are More than Pictures, They Are Data. Radiology 278:563–577
11. Zheng X, Yao Z, Huang Y et al (2020) Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun 11:1236
12. Coroller TP, Grossmann P, Hou Y et al (2015) CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 114:345–350
13. Wibmer A, Hricak H, Gondo T et al (2015) Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. Eur Radiol 25:2840–2850
14. Michoux N, Van den Broeck S, Lacoste L et al (2015) Texture analysis on MR images helps predicting non-response to NAC in breast cancer. BMC Cancer 15:574
15. Dilorenzo G, Telegrafo M, La Forgia D, Stabile Ianora AA, Moschetta M (2019) Breast MRI background parenchymal enhancement as an imaging bridge to molecular cancer sub-type. Eur J Radiol 113:148–152
16. Youk JH, Jung I, Yoon JH et al (2016) Comparison of inter-observer variability and diagnostic performance of the fifth edition of BI-RADS for breast ultrasound of static versus video images. Ultrasound Med Biol 42:2083–2088
17. Hu Y, Qiao M, Guo Y et al (2017) Reproducibility of quantitative high-throughput BI-RADS features extracted from ultrasound images of breast cancer. Med Phys 44:3676–3685
18. Guo Y, Hu Y, Qiao M et al (2018) Radiomics analysis on ultrasound for prediction of biologic behavior in breast invasive ductal carcinoma. Clin Breast Cancer 18:e335–e344
19. Li JW, Fang Z, Zhou J et al (2019) The association between molecular biomarkers and ultrasonographic radiomics features for triple negative invasive breast carcinoma. Chin J Ultrasonogr 28:137–143
20. Mendelson EB, Böhm-Vélez M, Berg WA (2013) ACR BI-RADS® Ultrasound. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. American College of Radiology, Reston
21. Chen P, Chen Y, Deng Y et al (2020) A preliminary study to quantitatively evaluate the development of maturation degree for fetal lung based on transfer learning deep model from ultrasound images. Int J Comput Assist Radiol Surg 15:1407–1415
22. Zhang Q, Xiao Y, Suo J et al (2017) Sonoelastomics for breast tumor classification: a radiomics approach with clustering-based feature selection on sonoelastography. Ultrasound Med Biol 43:1058–1069
23. Goldhirsch A, Winer EP, Coates AS et al (2013) 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 24:2206–2223
24. Li JW, Zhou J, Shi ZT, Li N, Zhou SC, Chang C (2021) Sonographic features of triple-negative breast carcinomas are correlated with mRNA-lncRNA signatures and risk of tumor recurrence. Front Oncol 10:587422
25. Li JW, Li N, Jiang YZ et al (2020) Ultrasonographic appearance of triple-negative invasive breast carcinoma is associated with novel molecular subtypes based on transcriptomic analysis. Ann Transl Med 8:435
26. Baldi P, Brunak S, Cgayvin Y, Andersen CA, Nielsen H (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16:412–424
27. Boisserie-Lacroix M, Macgrogan G, Debled M et al (2013) Triple-negative breast cancers: associations between imaging and

pathological findings for triple-negative tumors compared with hormone receptor-positive/human epidermal growth factor receptor-2-negative breast cancers. Oncologist 18:802–811
28. Burstein MD, Tsimelzon A, Poage GM et al (2015) Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res 21:1688–1698
29. Jiang YZ, Ma D, Suo C et al (2019) Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. Cancer Cell 35:428-440 e5
30. Shen M, Jiang YZ, Wei Y et al (2019) Tinagl1 suppresses triple-negative breast cancer progression and metastasis by simultaneously inhibiting integrin/FAK and EGFR signaling. Cancer cell 35:64-80 e7
31. Lehmann BD, Bauer JA, Chen X et al (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 121:2750–2767
32. Liu YR, Jiang YZ, Xu XE et al (2016) Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. Breast Cancer Res 18:33
33. Elsawaf Z, Sinn HP, Rom J, Bermejo JL, Schneeweiss A, Aulmann S (2013) Biological subtypes of triple-negative breast cancer are associated with distinct morphological changes and clinical behaviour. Breast 22:986–992
34. Zhao S, Ma D, Xiao Y et al (2020) Molecular subtyping of triple-negative breast cancers by immunohistochemistry: molecular basis and clinical relevance. Oncologist 25:e1481-1491
35. Yang Q, Liu HY, Liu D, Song YQ (2015) Ultrasonographic features of triple-negative breast cancer: a comparison with other breast cancer subtypes. Asian Pac J Cancer Prev 16:3229–3232
36. Uematsu T, Kasami M, Yuen S (2009) Triple-negative breast cancer: correlation between MR imaging and pathologic findings. Radiology 250:638–647
37. Lee SE, Han K, Kwak JY, Lee E, Kim EK (2018) Radiomics of US texture features in differential diagnosis between triple-negative breast cancer and fibroadenoma. Sci Rep 8:13546
38. Costantini M, Belli P, Bufi E, Asunis AM, Ferra E, Bitti GT (2016) Association between sonographic appearances of breast cancers and their histopathologic features and biomarkers. J Clin Ultrasound 44:26–33
39. Tamaki K, Ishida T, Miyashita M et al (2011) Correlation between mammographic findings and corresponding histopathology: potential predictors for biological characteristics of breast diseases. Cancer Sci 102:2179–2185
40. Yip SS, Aerts HJ (2016) Applications and limitations of radiomics. Phys Med Biol 61:R150-166
41. La Forgia D, Fanizzi A, Campobasso F et al (2020) Radiomic analysis in contrast-enhanced spectral mammography for predicting breast cancer histological outcome. Diagnostics (Basel) 10:708
42. Golden DI, Lipson JA, Telli ML, Ford JM, Rubin DL (2013) Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer. J Am Med Inform Assoc 20:1059–1066
43. Kim K, Song MK, Kim EK, Yoon JH (2017) Clinical application of S-Detect to breast masses on ultrasonography: a study evaluating the diagnostic performance and agreement with a dedicated breast radiologist. Ultrasonography 36:3–9
44. Jiang YZ, Liu YR, Xu XE et al (2016) Transcriptome analysis of triple-negative breast cancer reveals an integrated mRNA-lncRNA signature with predictive and prognostic value. Cancer Res 76:2105–2114
45. Jiang YZ, Ma D, Suo C et al (2019) Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. Cancer Cell 19:30096–30090
46. Wang H, Zhan W, Chen W, Li Y, Chen X, Shen K (2020) Sonography with vertical orientation feature predicts worse disease outcome in triple negative breast cancer. Breast 49:33–40
47. Rella R, Belli P, Giuliani M et al (2018) Automated breast ultrasonography (ABUS) in the screening and diagnostic setting: indications and practical use. Acad Radiol 25:1457–1470

Authors and Affiliations

Jia-wei Li[1,2] · Yu-cheng Cao[3] · Zhi-jin Zhao[1,2] · Zhao-ting Shi[1,2] · Xiao-qian Duan[3] · Cai Chang[1,2] · Jian-gang Chen[3]

✉ Cai Chang
changc61@163.com

✉ Jian-gang Chen
jgchen@cee.ecnu.edu.cn

[1] Department of Medical Ultrasound, Fudan University Shanghai Cancer Center, No 270, Dong'an Road, Xuhui District, Shanghai 200032, China

[2] Department of Oncology, Shanghai Medical College, Fudan University, No 270, Dong'an Road, Xuhui District, Shanghai 200032, China

[3] Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, #500 Dongchuan Rd., Shanghai 200241, China