An Ultrasound-Based Radiomics Nomogram for Differentiation of Triple-Negative Breast Cancer From Fibroadenoma

Yu Du

The First Affiliated Hospital of Nanjing Medical University

Hai-Ling Zha

The First Affiliated Hospital of Nanjing Medical University

Hui Wang

The First Affiliated Hospital of Nanjing Medical University

Xin-Pei Liu

The First Affiliated Hospital of Nanjing Medical University

Jia-Zhen Pan

The First Affiliated Hospital of Nanjing Medical University

Li-Wen Du

The First Affiliated Hospital of Nanjing Medical University

Min Zong

The First Affiliated Hospital of Nanjing Medical University

CUI-YING LI (✉ lcy_njmu@163.com)

The First Affiliated Hospital of Nanjing Medical University <https://orcid.org/0000-0002-8633-791X>

Research article

Keywords: Triple-negative breast cancer, Fibroadenoma, Ultrasonography, Radiomics, Nomogram

Posted Date: September 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-855924/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at The British Journal of Radiology on February 9th, 2022. See the published version at <https://doi.org/10.1259/bjr.20210598>.

Abstract

Background: To develop a radiomics nomogram that incorporates the radiomics features and ultrasound (US) conventional features and clinical findings to differentiate triple-negative breast cancer (TNBC) from fibroadenoma.

Methods: A total of 182 pathology-proven fibroadenomas and 178 pathology-proven TNBCs which underwent preoperative US examination were involved and randomly divided into training ($n = 253$) and validation cohorts ($n = 107$). The regions of interest (ROIs) of all lesions were delineated based on preoperative US examination subsequently radiomics features extracted. The least absolute shrinkage and selection operator model and the maximum relevance minimum redundancy algorithm were established for the selection of tumor status-related features and construction of radiomics signature (Rad-score). Then, multivariate logistic regression analyses were utilized to develop a radiomics model by incorporating the radiomics signature and clinical findings. Finally, the usefulness of the combined nomogram was assessed by the receiver operator characteristic curve (ROC), calibration curve, and decision curve analysis (DCA).

Results: The radiomics signature, composing of twelve selected features, achieved good diagnostic performance. The nomogram incorporated with radiomics signature and clinical data showed favorable diagnostic efficacy in the training cohort (AUC 0.986, 95% CI, 0.975-0.997) and validation cohort (AUC 0.977, 95% CI, 0.953-1.000). The radiomics nomogram outperformed the Rad-score and clinical models alone ($p < 0.05$). The calibration curve and decision curve analysis demonstrated the better clinical utility of the combined radiomics nomogram.

Conclusions: The radiomics signature is a potential predictive indicator for differentiating TNBC and fibroadenoma. The radiomics nomogram associated with Rad-score, US conventional features, and clinical data outperformed the Rad-score and clinical models alone.

Advances in knowledge: Recent advances in radiomics-based US identified increasingly gaining ground in oncology to improve diagnosis, assessment of therapeutic response, and disease prediction. We revealed that the Rad-score is an independent predictive indicator for differentiating TNBC and fibroadenoma. The radiomics nomogram associated with Rad-score, US conventional features, and clinical data outperformed the Rad-score and clinical models alone.

Background

Triple-negative breast cancer (TNBC) is regarded as the most aggressive breast cancer, and it is immunohistochemically defined by the negative expression of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 [1]. The disease occurs in approximately 10–15% of all breast cancers and has a potential correlation with young age and mutation of the TP53 gene and a high degree of correlation with suppressed BRCA1 function [2]. TNBC is one of the most fatal tumors and shows less favorable outcomes, the possible causes for poor outcomes with TNBC include the high pathologic grade and the absence of effective targeted therapies such as endocrine treatment or trastuzumab [1, 3]. The immediate priorities are early detection and personalized treatment.

Conventional breast ultrasonography is an indispensable imaging modality for detection and evaluation of masses in the breast. The predominant ultrasound (US) presentation found that approximately 21%–40% TNBC present oval or round in shape, with a circumscribed margin, a parallel orientation, and posterior acoustic enhancement; these morphological characteristics are also typically encountered in fibroadenoma [3–5]. Early studies emphasized that

these benign clinical features may be impaired the diagnostic performance of the ultrasound [3]. Although US is a preoperative assessment diagnosing TNBC with noninvasiveness, reproducibility, and no additional cost, the specificity and accuracy should be improved.

Radiomics is an emerging methodology concerned with the quantitative analysis of large-scale medical imaging data, and it can automatically filter comprehensive data obtained from an image [6, 7]. Recent advances in radiomics-based US identified increasingly gaining ground in oncology to improve diagnosis, assessment of therapeutic response, and disease prediction [8–12]. Luo et al. presented that the nomogram incorporating radiomics score and BI-RADS exhibited a potential diagnostic capacity in benign and malignant masses, especially in the BI-RADS category 4 or 5 [11]. Youk et al. demonstrated that textural features based on US images have a potential diagnostic performance to distinguish benign and malignant breast cancer [13]. To date, whether radiomics features can differentiate TNBC and fibroadenoma and combined conventional US features and clinical findings can improve the predictive values for distinguishing TNBC from fibroadenoma have not been systematically investigated. Therefore, the objective of this study was to construct a radiomics nomogram based on conventional US morphological characteristics and clinical findings to differentiate TNBC from fibroadenoma.

Methods

Patients

This retrospective study was approved by the ethics committee of the First Affiliated Hospital of Nanjing Medical University (Nanjing, China), and patient informed consent was waived. This research complied with the Declaration of Helsinki. Between April 2016 and January 2021, we identified pathology-proven fibroadenomas and pathology-proven TNBCs which underwent preoperative US examination. The inclusion criteria included the following: (1) verified lesions after inpatient surgery; (2) complete US and immunohistochemical data; (3) US images acquired from the same ultrasonic instrument; (4) US images stored by Digital Imaging and Communications in Medicine. The exclusion criteria included the following: (1) patients who had undergone neoadjuvant chemotherapy; (2) a diffuse extensive intraductal component on the US images; (3) patients receiving biopsy before US examination; (4) tumor size not fully included in the same plane. Finally, 360 patients involved in our study were randomly separated into training (n = 253) and testing (n = 107) groups. The recruitment process flowchart is shown in Figure 1A.

US Examination and Region-of-Interest (ROI) Segmentation

The flowchart of radiomics is summarized in Figure 1B and can be separated into three parts: imaging acquisition, ROI segmentation, feature extraction. In addition, the subsequent study flowchart included feature selection, model analysis, and model evaluation. All the US images from the institution were collected using the same US instrument (MyLab Twice, Esaote, Italy) equipped with a 5–13 MHz linear array transducer. All of the patients were examined in a supine position with full exposure of the breast, including the area closest to the chest wall for better detection of the breast lumps.

To test the inter-observer and intra-observer about the repeatability and reproducibility based on the features extracted from the ROIs. We randomly selected 60 patients for ROI segmentation, two radiologists with 3 years (Y.D.) and 15 years (H.W.) of experience in breast US scans who were unaware of the pathological results. The maximal-diameter plane was selected according to the US images of each breast lesion, and then the two radiologists drew an ROI along the mass boundary using an open source application called ITK-SNAP (<http://www.itksnap.org>). Then radiologist (H.W.) repeated the same workflow after one month. Features with intraclass correlation

coefficients (ICCs) better than 0.80 in the inter-observer and intra-observer agreement extraction were considered for subsequent analysis. The remaining ROI segmentation of the images was accomplished by a radiologist (H.W.).

Clinical Information and US conventional features

Clinical characteristics such as age were acquired by reviewing the medical records. US features and the assessment of the BI-RADS category were retrospectively reviewed by two investigators (H.W and Y.D), who have 15 years' and 3 years' experience in breast US examination. Neither of the investigators was involved in the US examination, and both were unaware of the clinical information and the pathology data of the patients. They were asked to evaluate and record the imaging features of all the patients. The following image features of breast lesions were recorded and divided into various categories: 1) size (maximum tumor diameter: <3 or >3 cm); 2) shape (oval or round/irregular); 3) margin (well- or non-circumscribed); 4) orientation (parallel or anti-parallel); 5) echotexture (hypoechoic or heterogeneous); 6) posterior echo feature (none or enhancement); 7) BI-RADS category (3, 4A, 4B, 4C, or 5). In cases of discordance, a consensus reading was performed, and the consensus data were used for the following analysis.

Radiomics analysis and radiomics construction

Pyradiomics package 2.1.2 was used for the extraction of radiomics features from the ROIs [14]. We used the maximum relevance minimum redundancy (mRMR) algorithm to develop TNBC-related radiomics signatures. The most appropriate feature with nonzero coefficient in the training group among the 360 breast images features was selected using the least absolute shrinkage and selection operator (LASSO) logistic regression algorithm (Figure 2A and 2B). Then, penalty parameter tuning adjusted by 10-fold cross-validation was used to access the robust and nonredundant features from the initial cohort. Each patient with a Rad-score was created by a linear combination of selected features that were weighted by their respective coefficients [15]. The corresponding Rad-score was generated with the selected features and calculated with all the patients in the training and validation groups.

A total of 1283 features were computed and divided into three categories: (1) 9 shape-based features: used to evaluate typical morphological features, such as shape and size information about the mass; (2) 216 first-order statistics: first-order parameters that are used to calculate the distribution of individual voxel intensities through commonly used and basic metrics and ignore the spatial information within the tumors; among the first-order statistics, entropy is consistently considered one of the most stable features; (3) 1058 second-order features: generally known as textural features, encode valuable information about the scene, and have a relative spatial arrangement of intensity values in a medical image; in the second-order features we extracted, including Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Dependence Matrix (GLDM) features, Gray Level Size Zone Matrix (GLSZM) features, Gray Level Run Length Matrix (GLRLM) features, Neighbouring Gray Tone Difference Matrix (NGTDM) features and wavelet-based features [16-19].

Establishment and validation of the radiomic nomogram

Among our study, we separated all the patients into training (n = 253) and validation (n = 107) datasets using a statistical software. One-way analysis of variance was used to compare demographic characteristics and continuous variables as appropriate. The χ^2 test or Fisher's exact test was used to establish significant differences in categorical variables as appropriate. To determine which clinical characteristics could serve as a candidate predictor for TNBC, we used univariate analysis to variable screen, those variables with $p < 0.1$ were subjected to the

subsequent analysis. Therewith, a clinical predictive was developed through multivariate logistic regression analysis. Regarding the multivariate logistic regression analysis, we applied the variance inflation factor (VIF) to estimate the collinearity diagnostics. The characteristic features were integrated with radiomics signature, then a radiomics nomogram was developed with multivariate analysis and provided an individualized and visual model tool for distinguishing TNBC from fibroadenoma.

The potential predictive ability of the established models was assessed by the receiver operating characteristic (ROC) curve analysis with the area under the curve (AUC). The model performance with statistical difference of AUC was compared with the Delong algorithm ($p < 0.05$). The nomogram model was well-adjusted in accordance with a calibration curve evaluated by the Hosmer–Lemeshow test and used to assess the prediction capability of the nomogram. The discrimination diagnostic performance of the radiomics nomogram model was quantified with the access of concordance index (C-Index). The range of the C-index is defined 0.5 to 1.0, with 1.0 corresponds to the best model prediction, and 0.5 represents a random prediction [20]. Bootstraps with 1000 resamples were calculated for the prediction model relatively corrected C-index, and a corrected model was developed. The model internal performance was used on all patients in the validation cohort.

Clinical utility of the radiomics nomogram

The discrimination and predictive performance of the established models were assessed based on the decision curve analysis (DCA) of the training and validation datasets. DCA was developed to ascertain the clinical utility of the nomogram by quantifying the net benefits at different threshold probabilities in the whole cohort [21].

Statistical analysis

Statistical analyses were conducted with SPSS (version 26.0) and R statistical software (version 4.0.5). An independent sample T-test was used to compare continuous data as appropriate. Categorical variables were compared using Fisher's exact test or Chi-square test. The LASSO logistic regression analysis was running by "glmnet" package. The "mRMRe" package was applied to implement mRMR algorithm. The VIF values were computed using the "car" package. The ROC curves were plotted using "pROC" package. The nomogram and calibration curve was constructed using the "rms" package. Decision curve was plotted using the "dca.r" package. A two-tailed p -value of < 0.05 indicated statistical significance.

Results

Clinical Characteristics and Clinical Model

The study flowchart is summarized in Figure 1. Table 1 presents the comparisons of US-based morphology characteristics and clinical data, respectively. The results demonstrated no statistical difference between the training and validation cohorts in terms of lesion size, age, and US morphological features ($p = 0.061$ - 0.762). Univariate analysis showed that age, tumor size, shape, margin, orientation, posterior echo feature, BI-RADS category, and Rad-score were associated with well statistical significance in the training cohort (Table 2). According to the multivariate analysis, the patients' age (odds ratio [OR]: 1.16; 95% CI: 1.09-1.24; $p < 0.001$), a BI-RADS final assessment category of 3 (OR: 0.04; 95%CI: 0.01-1.24; $p = 0.001$) or 4 (OR of 4B: 7.43; 95%CI: 1.72-32.142; $p = 0.007$; OR of 4C: 17.84; 95%CI: 2.84-112.23; $p = 0.002$) or 5 (OR: 23.85; 95%CI: 1.05-539.60; $p = 0.001$), and Rad-score (OR: 1.06; 95%CI: 1.18-1.32; $p < 0.001$) were considered as independent predictors of the predictive model associated with TNBC. The clinical model was constructed with age and BI-RADS category, which was

showed favorable discrimination with an AUC of 0.942 (95%CI: 0.914-0.971) in the training datasets and 0.943 (95% CI: 0.901-0.984) in the testing datasets, respectively. The specific number of TNBC and fibroadenoma lesions in the training and validation cohort were illustrated in Table S2.

Features Selection and Rad-score Building

A total of 1283 radiomics features were selected, and 26 features were extracted by the mRMR associated with TNBC. The intra-observer ICCs ranged from 0.824 to 0.988 and the inter-observer ICCs ranged from 0.807 to 0.973, demonstrating a mark of satisfactory reproducibility in ROI segmentation. Finally, the features were reduced to 12 potential predictive factors from features with nonzero coefficients by the LASSO regression algorithm (Figure 3). More information about the features can be found in Supplementary A1 and Table S1. The formula calculated by the Rad-score associated with the 12 radiomic features was as follows:

Rad-score = $-0.107 \times \text{exponential_glszm_ZoneEntropy} + 0.444 \times \text{square_glszm_LargeAreaLowGrayLevelEmphasis} + 0.551 \times \log\text{-sigma-3-0-mm-2D_firstorder_Skewness} + 0.007 \times \text{squareroot_ngtdm_Coarseness} + 0.326 \times \text{square_ngtdm_Contrast} + 0.195 \times \log\text{-sigma-5-0-mm-2D_glcm_lmc1} + 0.284 \times \text{lbp-2D_glszm_ZoneVariance} + 0.35 \times \log\text{-sigma-5-0-mm-2D_glszm_SmallAreaEmphasis} + 0.305 \times \log\text{-sigma-5-0-mm-2D_glrlm_ShortRunLowGrayLevelEmphasis} + 0.233 \times \text{gradient_firstorder_Kurtosis} + 0.579 \times \text{wavelet-H_firstorder_Skewness} + 0.131 \times \text{exponential_gldm_DependenceEntropy} + -0.071$

The Rad-score presented an acceptable diagnostic accuracy in predicting TNBC with an AUC of 0.890 (95% CI: 0.851-0.929) in the training set and 0.886 (95% CI: 0.825-0.948) in the validation set, respectively.

Construction and Validation of Radiomics Nomogram

Clinical and Rad-score models were constructed in accordance with the above methods, and both ascertained the independent predictors of TNBC in the training and testing groups. The radiomics nomogram model incorporated significant clinical morphological predictors with the fusion Rad-score of US images. Among the multivariate logistic regression analysis, age, BI-RADS, and Rad-score were identified as potential predictors associated with TNBC. The VIF values of the three predictors ranged from 1.13 to 1.19, indicating that there was no collinearity in the collinearity diagnosis. A nomogram incorporated with the predictors was established (Figure 4).

The combined model also showed improved diagnostic efficiency in the ROC analysis compared with the Rad-score and clinical models alone (Figure 5A and 5B). By comparing the models, the combined model presented an optimal performance and the best AUC value (0.986; 95% CI: 0.975-0.997) in the training group. The model also presented the highest AUC (0.977; 95% CI: 0.953-1.000) in the testing cohort. Therefore, we suggested that the nomogram integrates with Rad-score and clinical model for TNBC prediction, with a satisfactory discrimination ability achieved. Figure 5C illustrated the calibration curve of the nomogram in the training cohort, the Hosmer-Lemeshow test yielded a non-significant statistic ($p = 0.732$). The calibration curve of the nomogram in the validation cohort and the Hosmer-Lemeshow test showed favorable calibration (Figure 5D) with a non-significant statistic ($p = 0.807$). The nomogram we recommended was integrated with Rad-score and clinical model, with favorable discrimination achieved (C-index, 0.984 in training cohort, and 0.985 in validation cohort, respectively). Hence, our nomogram demonstrated a favorable precision both in the training and validation cohorts.

In addition, the combined model outperformed the Rad-score model (AUC values of 0.986 versus 0.890; $p < 0.001$) and clinical model (AUC values of 0.986 versus 0.942; $p < 0.05$) in the training cohort according to the Delong test. In the validation cohort, the combined model also exhibited a better predictive capability than the Rad-score (AUCs of 0.977 versus 0.886; $p < 0.001$) and clinical models (AUCs of 0.977 versus 0.943; $p < 0.001$).

The Clinical Usefulness of the Nomogram

The decision curves for the clinical model, Rad-score model, and combined nomogram in the training cohort were presented in Figure 6A. The results showed that using the nomogram to predict the risk of TNBC adds more net benefit than the “treat all” or “treat none” strategies and achieves the most net benefit when threshold probabilities range from 0.1 to 1.0. Similar results could be found in the testing cohort (Figure 6B).

Discussion

In this study, we developed a nomogram incorporated radiomics signature, US conventional features, and clinical findings to differentiate TNBC from fibroadenoma. The Rad-score was not only a potential predictive indicator in differentiating TNBC and fibroadenoma but also showed a better performance when combined with the clinical findings.

Radiomics is a computer-aided diagnosis model that uses an automated high-throughput extraction of large amounts of quantitative phenotypic characteristics of medical images and converts features into mathematical models that can provide more and better information than a radiologist [22-24]. The evaluation of criteria and reporting guidelines needs to be completed to develop radiomics analysis and mature as a discipline [25]. In previous studies, texture analysis demonstrated a favorable performance in the clinical diagnosis based on US morphological features, distinguishing benign and malignant lesions, differentiating between breast cancer subtypes, and therapeutic responses to neoadjuvant chemotherapy which illustrated the clinical application of the radiomics [9, 10, 13]. Lee et al. demonstrated that the Rad-score based on US texture presented a favorable diagnostic performance for differentiating TNBC and fibroadenoma with an AUC of 0.91 [13]. Additionally, Moon et al. designed US-based texture features of computer-aided diagnosis that can be useful to differentiate TNBC from fibroadenoma with an AUC of 0.84. However, 169 patients were involved in this study; thus, the application of this CAD system in every US examination is impossible, and computer feature sets may waste a considerable amount of time when getting into trouble in the multiresolution gray-scale invariant texture [26]. In our present research, we developed a Rad-score consisting of 12 radiomic features to differentiate TNBC from fibroadenoma, and the Rad-score demonstrated sufficient discrimination performance between the TNBC and fibroadenoma, which further emphasizes the robust capability of radiomics features. Moreover, the Rad-score indicated was in accordance with that of previous studies and can serve as an independent biomarker for predicting TNBC and fibroadenoma.

US morphological features normally depend on conventional clinical practices for differentiating variable subtypes of breast cancer. Conventional sonographic features are intuitive, and TNBC has atypical features that are an underestimation of the diagnostic accuracy. Yoon et al. TNBC presented more suspicious US characteristics with posterior echo enhancement, irregular shape, and a non-circumscribed margin than fibroadenoma [4]. Yeo et al. reported a study predicting TNBC and fibroadenoma based on US alone of 131 patients, showing dissatisfactory results with an AUC of 0.65 [27]. This conclusion on TNBCs can be confused with diagnosis due to several distinct sonographic criteria and is more likely to be associated with benign tumor such as fibroadenoma. In our study, the univariate analysis indicated that most US characteristics showed a

good diagnostic ability in distinguishing TNBC and fibroadenoma. Multivariate analysis presented that age and BI-RADS category were associated with a predictive clinical model, which can be an effective tool for predicting TNBC and fibroadenoma. Thus, US characteristic is essential in constructing a clinical predictive model. However, image acquisition depends on the experience of the radiologist and different radiologists are skilled in different fields, such as breast, thyroid, and liver. This condition may well explain why the AUC of the clinical model was inferior to the radiomics nomogram in this study.

Rapid nomograms are widely used to the response of treatment and prediction of treatment in oncology [28-30]. To our knowledge, this research is the first study that is developed and validated to predict breast lesions with TNBC and fibroadenoma using a radiomics signature-based nomogram. Substantial overlap exists in the conventional US features of TNBC and fibroadenoma. Taking a further step, previous studies have constructed a US-based radiomics signature and proved a radiomics nomogram which was dedicated to distinguish benign and malignant lesions and preoperative prediction of in TNBC [25, 31, 32]. Luo et al. reported a nomogram combined BI-RADS and radiomics, which showed potential application value for the diagnosis and discrimination of category 4 or 5 lesions in BI-RADS [11]. In our study, we constructed and identified radiomics nomogram-incorporated Rad-score and clinical findings to better and easily distinguish TNBC and fibroadenoma. The nomogram presented an enormous potential for differentiating such lesions with a satisfactory C-index and well calibration. First, we observed that the radiomics signature has a good accuracy to differentiate TNBC and fibroadenoma. Next, we used multivariate logistic regression analysis to construct a combined nomogram. We observed its excellent calibration and attempted to unravel its predictive capacity for differentiating TNBC and fibroadenoma. The nomogram showed a better diagnostic discrimination performance than the Rad-score or clinic model alone, and it can be more intuitive and easily used in distinguishing breast lesions with TNBC and fibroadenoma. Future directions can be associated with the prognostic indicator and treatment schedule and identify the riskiest radiomics signature.

Our present study possess several limitations. First, our study was retrospective and had a potential selection bias. Second, ROIs were drawn manually, and the lack of standardization due to differences in selection can be encountered although the two radiologists showed excellent inter-observer and intro-observer agreement in our study. Fully-automatic segmentation tools associated with better diagnostic accuracy should be applied in future studies to avoid such potential faults. Third, the radiomics nomogram may be difficult to generalize because all the breast lesions were examined using the same machine and US examination protocol. Further studies are required for multiple types of US and multi-center validation with a larger sample to achieve higher predictive accuracy for clinical application.

Conclusion

In conclusion, the present study applied US-based images to construct a Rad-score model and suggested it as an independent biomarker for risk prediction in patients with TNBC. In addition, our study presented a nomogram, with a combination of Rad-score and clinical findings, which exhibited a favorable performance in differentiating TNBC and fibroadenoma. The combined model may possess practicability and value of popularization to reduce the number of biopsies and provide a suitable treatment strategy in the future.

Abbreviations

US: Ultrasound; TNBC: Triple-negative breast cancer; ROI: regions of interests; BI-RADS: Breast Imaging Reporting and Data System; mRMR: Maximum relevance minimum redundancy; LASSO: Least absolute shrinkage and

selection operator; DCA: Decision curve analysis; ROC: Receiver operating characteristic curve; ICC: Intraclass correlation coefficients; VIF: Variance inflation factor; GLCM: Gray Level Co-occurrence Matrix; GLRLM: Gray Level Run Length Matrix; GLSZM: Gray Level Size Zone Matrix; NGTDM: Neighbouring Gray Tone Difference Matrix; GLDM: Gray Level Dependence Matrix.

Declarations

Acknowledgments

The authors thank Cheng Yang for his support in the statistical analysis.

Authors' contributions

YD, CYL, HLZ, XPL designed the clinical study from which patient data in this study derives and assisted with data collection. YD and MZ designed the retrospective radiomics study. YD, HW, LWD, HLZ, and JZP contributed to the analysis and interpretation of data. YD, HLZ, MZ, and CYL contributed to the writing, review, and/or revision of the manuscript. CYL, HW, MZ, and HLZ contributed to the technical or material support. MZ and CYL contributed to the study supervision. All authors read and approved the final manuscript.

Funding

None.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This retrospective study was approved by the ethics committee of the First Affiliated Hospital of Nanjing Medical University, and patient consent was waived.

Consent for publication

Not applicable.

Competing interests

None.

References

1. Foulkes WD, Smith IE, Reis-Filho JS. Triple-Negative Breast Cancer. *N Engl J Med*. 2010;363(20):1938–48.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7–30.
3. Dogan BE, Turnbull LW. Imaging of triple-negative breast cancer. *Ann Oncol*. 2012;23(Suppl 6):vi23–9.
4. Yoon GY, Cha JH, Kim HH, Shin HJ, Chae EY, Choi WJ. Sonographic features that can be used to differentiate between small triple-negative breast cancer and fibroadenoma. *Ultrasonography*. 2018;37(2):149–56.

5. Ko ES, Lee BH, Kim HA, Noh WC, Kim MS, Lee SA. Triple-negative breast cancer: correlation between imaging and pathological findings. *Eur Radiol.* 2010;20(5):1111–7.
6. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. *J Nucl Med.* 2020;61(4):488–95.
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology.* 2016;278(2):563–77.
8. Zha HL, Zong M, Liu XP, Pan JZ, Wang H, Gong HY, et al. Preoperative ultrasound-based radiomics score can improve the accuracy of the Memorial Sloan Kettering Cancer Center nomogram for predicting sentinel lymph node metastasis in breast cancer. *Eur J Radiol.* 2021;135:109512.
9. Jiang M, Li CL, Luo XM, Chuan ZR, Lv WZ, Li X, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. *Eur J Cancer.* 2021;147:95–105.
10. Gao Y, Luo Y, Zhao C, Xiao M, Ma L, Li W, et al. Nomogram based on radiomics analysis of primary breast cancer ultrasound images: prediction of axillary lymph node tumor burden in patients. *Eur Radiol.* 2021;31(2):928–37.
11. Luo WQ, Huang QX, Huang XW, Hu HT, Zeng FQ, Wang W. Predicting Breast Cancer in Breast Imaging Reporting and Data System (BI-RADS) Ultrasound Category 4 or 5 Lesions: A Nomogram Combining Radiomics and BI-RADS. *Sci Rep.* 2019;9(1):11921.
12. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749–62.
13. Lee SE, Han K, Kwak JY, Lee E, Kim E-K. Radiomics of US texture features in differential diagnosis between triple-negative breast cancer and fibroadenoma. *Sci Rep.* 2018;8(1):13546.
14. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21):e104-7.
15. Ji GW, Zhu FP, Zhang YD, Liu XS, Wu FY, Wang K, et al. A radiomics approach to predict lymph node metastasis and clinical outcome of intrahepatic cholangiocarcinoma. *Eur Radiol.* 2019;29(7):3725–35.
16. Wu Q, Wang S, Chen X, Wang Y, Dong L, Liu Z, et al. Radiomics analysis of magnetic resonance imaging improves diagnostic performance of lymph node metastasis in patients with cervical cancer. *Radiother Oncol.* 2019;138:141–8.
17. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol Biol Phys.* 2018;102(4):1143–58.
18. Rai R, Holloway LC, Brink C, Field M, Christiansen RL, Sun Y, et al. Multicenter evaluation of MRI-based radiomic features: A phantom study. *Med Phys.* 2020;47(7):3054–63.
19. Gibbs P, Onishi N, Sadinski M, Gallagher KM, Hughes M, Martinez DF, et al. Characterization of Sub-1 cm Breast Lesions Using Radiomics Analysis. *J Magn Reson Imaging.* 2019;50(5):1468–77.
20. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med.* 2011;30(10):1105–17.
21. Hu HT, Wang Z, Huang XW, Chen SL, Zheng X, Ruan SM, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *Eur Radiol.* 2019;29(6):2890–901.

22. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–6.
23. Conti A, Duggento A, Indovina I, Guerri M, Toschi N. Radiomics in breast cancer classification and prediction. *Semin Cancer Biol*. 2021;72:238–50.
24. Gillies RJ, Schabath MB. Radiomics Improves Cancer Screening and Early Detection. *Cancer Epidemiol Biomarkers Prev*. 2020;29(12):2556–67.
25. Cui X, Zhu H, Huang J. Nomogram for Predicting Lymph Node Involvement in Triple-Negative Breast Cancer. *Front Oncol*. 2020;10:608334.
26. Moon WK, Huang YS, Lo CM, Huang CS, Bae MS, Kim WH, et al. Computer-aided diagnosis for distinguishing between triple-negative breast cancer and fibroadenomas based on ultrasound texture features. *Med Phys*. 2015;42(6):3024–35.
27. Yeo SH, Kim GR, Lee SH, Moon WK. Comparison of Ultrasound Elastography and Color Doppler Ultrasonography for Distinguishing Small Triple-Negative Breast Cancer From Fibroadenoma. *J Ultrasound Med*. 2018;37(9):2135–46.
28. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16(4):e173-80.
29. Liang W, Yang P, Huang R, Xu L, Wang J, Liu W, et al. A Combined Nomogram Model to Preoperatively Predict Histologic Grade in Pancreatic Neuroendocrine Tumors. *Clin Cancer Res*. 2019;25(2):584–94.
30. Cai J, Zheng J, Shen J, Yuan Z, Xie M, Gao M, et al. A Radiomics Model for Predicting the Response to Bevacizumab in Brain Necrosis after Radiotherapy. *Clin Cancer Res*. 2020;26(20):5438–47.
31. Koh J, Lee E, Han K, Kim S, Kim D-K, Kwak JY, et al. Three-dimensional radiomics of triple-negative breast cancer: Prediction of systemic recurrence. *Sci Rep*. 2020;10(1):2976.
32. Braman N, Prasanna P, Whitney J, Singh S, Beig N, Etesami M, et al. Association of Peritumoral Radiomics With Tumor Biology and Pathologic Response to Preoperative Targeted Therapy for HER2 (ERBB2)-Positive Breast Cancer. *JAMA Netw Open*. 2019;2(4):e192561.

Tables

Table 1. Comparisons of the clinical and US morphological features in training and validation cohorts

Characteristic	Training cohort (n=253)	Testing cohort (n=107)	P value
Age*	44.23±13.14	42.71±13.75	0.323
Tumor size	205(81.0%)	89 (83.2%)	0.630
<3 cm	48 (19.0%)	18 (16.8%)	
>3 cm			
Shape	128 (50.6%)	56 (52.3%)	0.762
Oval or round	125 (49.4%)	51 (47.7%)	
Irregular			
Margin	126 (49.8%)	64 (59.8%)	0.642
Well-circumscribed	127 (50.2%)	43 (40.2%)	
Non-circumscribed			
Orientation	213 (84.2%)	88 (82.2%)	0.061
Parallel	40 (15.8%)	9 (17.8%)	
Nonparallel			
Echotexture	198 (78.3%)	91 (85.0%)	0.234
Hypoechoic	55 (21.7%)	16 (15.0%)	
Heterogeneous			
Posterior echo feature	178 (70.4%)	79 (73.8%)	0.505
None	75 (29.6%)	28 (26.2%)	
Enhancement			
BI-RADS category	71 (28.1%)	32 (29.9%)	0.673
3	66 (26.1%)	26 (24.3%)	
4A	61 (24.1%)	31 (29.0%)	
4B	34 (13.4%)	13 (12.1%)	
4C	21 (8.3%)	5 (4.7%)	
5			
Rad-score*	-0.07±2.06	0.25±1.95	0.160

NOTE: Unless otherwise noted, data are shown as number of patients, with the percentage in parentheses.

*Data are means ± standard deviations.

Table 2. Results of the univariate and multivariate analyses of the potential predictors based on the training cohort

Factors	Univariate analysis		Multivariate analysis	
	OR (95% CI)	<i>p</i> value	OR (95% CI)	<i>P</i> value
Age	0.86 (0.82 – 0.89)	< 0.001	1.16 (1.09 – 1.24)	<0.001
Tumor size	1 (ref)		NA	
<3 cm	0.58 (0.30 – 1.10)	0.092		
>3 cm				
Shape	1 (ref)		NA	
Oval or round	3.51 (2.10 – 5.90)	<0.001		
Irregular				
Margin	1 (ref)		NA	
Well-circumscribed	2.34 (1.41 – 3.87)	0.001		
Non-circumscribed				
Orientation	1 (ref)		NA	
Parallel	2.14 (1.16 – 4.32)	0.034		
Nonparallel				
Echotexture	1 (ref)		NA	
Hypoechoic	0.73 (0.43 – 1.26)	0.265		
Heterogeneous				
Posterior echo feature	1 (ref)		NA	
None	2.62 (1.55 – 4.44)	<0.001		
Enhancement				
BI-RADS category	0.19 (0.08 – 0.49)	<0.001	0.04 (0.01 – 1.24)	0.001
3	1 (ref)		1 (ref)	
4A	4.18 (1.99 – 8.80)	<0.001	7.43 (1.72 – 32.14)	0.007
4B	18.08 (4.99 – 65.49)	<0.001	17.84 (2.84 – 112.23)	0.002
4C	33.00 (4.41 – 277.38)	0.001	23.85 (1.05 – 539.60)	0.046
5				
Rad-score	1.26 (1.19 – 1.37)	<0.001	1.16 (1.08 – 1.32)	<0.001

NOTE: TNBC, triple negative breast cancer; OR, odds ratio; CI, confidence interval; ref, reference; BI-RADS, Breast Imaging Reporting and Data System

Figures

Figure 1

Workflow of the radiomics analysis and study flowchart. (A). The radiomics analysis was divided into three parts: imaging, ROI segmentation, and feature extraction. (B). In the study flowchart, LASSO logistic regression was used to construct radiomics signature and combining with the multivariate analysis of the clinical data to develop a nomogram, then the nomogram was validated in the testing set.

Figure 2

Selection of the radiomics features by using a parametric algorithm, the LASSO binary logistic regression model. (A). Selection of the tuning parameter (λ) was performed in the LASSO model by 10-fold cross-validation based on minimum criteria. A dotted vertical black line was depicted at the optimal value by using the 1 standard error of the minimum and minimum criteria. The number along the upper axis indicates the average number of predictors. The lower x-axis represents the $\log(\lambda)$ and the y-axis represents binomial deviances. The optimal λ value of 0.040 with $\log(\lambda) = -3.688$ was selected. (B). LASSO coefficients profiles through the 1283 features. The dotted vertical line was shown at the optimal value using 10-fold cross-validation in Figure 2A. Features with 12 nonzero coefficients are constructed by using the optimal λ in the plot.

Figure 3

Result of feature selection and the verified set of Rad-score. Twelve nonzero coefficients features are presented.

Figure 4

Combined radiomics nomogram constructed by Rad-score and clinical characteristics for predicting TNBC in the training cohort.

Figure 5

(A, B) Comparison of the AUCs for the clinical model, Rad-score model, and radiomics nomogram in the training and validation cohorts. (C). Calibration curves of the radiomics nomogram in the training cohort. (D). Calibration curves of the radiomics nomogram in the validation cohort. The X-axis represents the predictive probability; Y-axis denotes the observed probability. The 45° blue diagonal line represents the perfect prediction of TNBC, and the red line indicates the prediction model of the radiomics nomogram. The closer the red line fits to the ideal line, the better the discrimination of the nomogram is.

Figure 6

Decision curve for the clinical findings, Rad-score, and radiomics nomogram. The gray line represents the assumption that all patients are TNBC cases; the black line refers to the assumption that all patients are fibroadenoma cases. The x-axis denotes the high-risk threshold and the y-axis, the net benefit. The blue, orchid, and lightpink lines represent the clinical model, Rad-score model, and radiomics nomogram, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.docx](#)