

Digital breast tomosynthesis-based peritumoral radiomics approaches in the differentiation of benign and malignant breast lesions

Shuxian Niu* 

Tao Yu* 

Yan Cao 

Yue Dong 

Yahong Luo 

Xiran Jiang 

PURPOSE

We aimed to evaluate digital breast tomosynthesis (DBT)-based radiomics in the differentiation of benign and malignant breast lesions in women.

METHODS

A total of 185 patients who underwent DBT scans were enrolled between December 2017 and June 2019. The features of handcrafted and deep learning-based radiomics were extracted from the tumoral and peritumoral regions with different radial dilation distances outside the tumor. A 3-step method was used to select discriminative features and develop the radiomics signature. Discriminative clinical factors were identified by univariate logistic regression. The clinical factors with $P < .05$ were used to build a clinical model with multivariate logistic regression. The radiomics nomogram was developed by integrating the radiomics signature and discriminative clinical factors. Discriminative performance of the radiomics signature, clinical model, nomogram, and breast imaging reporting and data system assessment were evaluated and compared with the receiver operating characteristic and decision curves analysis (DCA).

RESULTS

A total of 2 handcrafted and 2 deep features were identified as the most discriminative features from the peritumoral regions with 2 mm dilation distances and used to develop the radiomics signature. The nomogram incorporating the radiomics signature, age, and menstruation status showed the best discriminative performance with area under the curve (AUC) values of 0.980 (95% CI, 0.960 to 1.000; sensitivity = 0.970, specificity = 0.946) in the training cohort and 0.985 (95% CI, 0.960 to 1.000; sensitivity = 0.909, specificity = 0.966) in the validation cohort. DCA confirmed the potential clinical usefulness of our nomogram.

CONCLUSION

Our results illustrate that the radiomics nomogram integrating the DBT imaging features and clinical factors (age and menstruation status) can be considered as a useful tool in aiding the clinical diagnosis of breast cancer.

From the Department of Biomedical Engineering (S.N., Y.C., X.J., ✉ xrjiang@cmu.edu.cn), China Medical University, Shenyang, China; Cancer Hospital of China Medical University (T.Y., Y.D., Y.L., ✉ luoyahong8888@hotmail.com), Liaoning Cancer Hospital and Institute, Shenyang, China.

*Shuxian Niu and Tao Yu contributed equally to this work.

Received 20 August 2020; revision requested 21 September 2020; last revision received 5 February 2021; accepted 24 February 2021.

Published online 15 April 2022.

DOI 10.5152/dir.2022.20664

Breast cancer is the most common cancer among females with fast-growing prevalence rates, accounting for approximately 30% of all cancer diagnoses.¹⁻³ Previous reports have demonstrated that early detection and timely treatment are key to the prognosis of breast cancer patients.⁴ In clinical practice, the most widely used early detection technique for breast tumor is standard X-ray mammography, due to its advantages of fast processing, low cost, and easy operation.⁵ However, the specificity and sensitivity of dense and heterogeneous breasts tend to significantly reduce because normal tissues can obscure the breast lesions, which cause limitations during visual examinations.⁴ In addition, structural disorder, which is also known as the early sign of breast cancer, is the most easily ignored false-negative sign in standard X-ray breast screenings, thus misleading the detection performance of early diagnosis.^{6,7} In order to differentiate occlusions from the presence of other issues, the digital breast tomosynthesis (DBT) scanning has been applied widely for early-stage breast cancer screening,⁸ which employs rotating the X-ray tubes at a limited angle, thus allowing improved identification of anomalies obscured by normal breast tissue.⁹ Early studies have shown the DBT examination can effectively reduce false-positive

You may cite this article as: Niu S, Yu T, Cao Y, Dong Y, Luo Y, Jiang X. Digital breast tomosynthesis-based peritumoral radiomics approaches in the differentiation of benign and malignant breast lesions. *Diagn Interv Radiol.* 2022;28(3):217-225.

and false-negative rates compared with traditional X-ray mammography on all breast density subtypes, especially on heterogeneously dense breasts,^{10,11} and demonstrated great clinical values of DBT for breast cancer patients.^{12,13} The wide applications of DBT screenings have some limitations since well-experienced radiographers are needed during visual examination of DBT data to give clinical decisions by analyzing the morphological changes in the DBT image.¹⁴

In recent years, radiomics has drawn increasing attraction in the fields of disease diagnosis, staging, and therapeutic response prediction because it makes precision medicine possible in a non-invasive manner by extracting and selecting high volumes of quantitative imaging features and revealing the features and the underlying pathophysiology.¹⁵⁻¹⁹ Previous studies have demonstrated improvements in the detection performance of breast cancer using radiomics compared with those using visual examination by radiologists.²⁰⁻²² Recent reports have also highlighted the importance of radiomics in breast cancer diagnosis by breast mammography,^{23,24} ultrasound,²⁵ magnetic resonance imaging,²⁶ and positron emission tomography/computed tomography,²⁷ although the reports on radiomics in breast DBT data are still limited.^{13,28-30} The published studies have inherent limitations with limited feature types, limited number of patients, and lack of correlating their findings with clinical evaluation, which may further limit the clinical applicability.

In this study, we evaluated the tumoral and peritumoral regions in the breast DBT image in differentiating malignant from

benign lesions using handcrafted and deep features and developed a radiomics nomogram by integrating the radiomics signature and important clinical factors for facilitating early diagnosis of breast cancer.

Methods

Patients

In this study, a total of 225 patients were enrolled between December 2017 and June 2019. The patients' lesions were pathologically confirmed as either benign or malignant. The retrospective research was approved by the ethics committee of our hospital (No. 2013010), and informed consent was waived. The patients (1) who were adult females older than 18 years and (2) who underwent DBT screening before surgery were included in the study. The patients (1) who underwent radiotherapy, chemotherapy, or immunotherapy treatment before the DBT examination; (2) who were in their menstruation, pregnancy, or lactation periods; (3) who had a history of breast surgery, as well as breast implants surgery before the DBT examination; and (4) whose regions of interest (ROIs) cannot be accurately segmented due to overlapping artifact in the DBT image were excluded. According to the abovementioned criteria, 40 patients were excluded and 185 patients were finally included. The numbers of patients with benign and malignant breast lesions were 85 and 100, respectively. The breast imaging reporting and data system (BI-RADS) assessment for each patient was retrieved from the electronic medical records system (EMRS) of our hospital. All patients were randomly divided into training and validation cohorts in a ratio of 2 : 1 by stratified sampling. This retrospective analysis of DBT data was approved by the institutional research ethics board of the hospital.

DBT acquisition

All patients were scanned with a breast DBT scanner (Hologic Selenia Dimensions, HOLOGIC) in our hospital. The obtained images were reconstructed into images with a thickness of 1 mm by computer to give a 3-dimensional view of the tissue, slice by slice and suitably spaced. The thickness of the compressed breast determines the number of slices. Parameters for performing DBT scan were as follows: voltage range on the X-ray tubes of 20.0-49.0

kV (step=1.0 kV), current time range of 300-400 m As, nominal power of 3.0 kW, scanning time of <4.0 s, reconstruction time of 2.0-5.0 s, and pixel size of 70 μ m. The obtained DBT images with a resolution of 2457 \times 1996 were interpreted on a HOLOGIC breast computer-aided diagnosis workstation (SecureViewDx; HOLOGIC) equipped with two 5-megapixel monitors. All the images were stored in the Picture Archiving and Communication System in our hospital in a format of Digital Imaging and Communications in Medicine.

Segmentation and mask dilation

For each patient, the tumoral regions (ROIs) were drawn by 2 radiologists with more than 10 years of work experience using the ITK-SNAP software (version 3.6.0; available from URL: <http://www.itksnap.org/>). If divergence occurred during the work, other senior radiologists and specialist breast clinicians would be invited to provide help in making final decisions. The original mask was dilated with 5 different dilatation distances with an interval of 2 mm outside the tumoral region (up to a dilation distance of 10 mm). Therefore, 5 new masks that represent the peritumoral regions outside the lesion were generated for each patient. Figure 1 shows an example of the ROI segmentation and mask dilation process. The dilated masks were then used to integrate with the original DBT images for the extraction of radiomics features.

Extraction of handcrafted features

For each patient, a total of 1967 handcrafted radiomics features were extracted using Pyradiomics as previously reported.³¹ The original handcrafted features include 18 first-order statistics features, 14 shape-based features, and 75 texture features. The texture features contain 24 gray-level co-occurrence matrix, 16 gray-level run-length matrix, 16 gray-level size zone matrix, 5 neighbor gray-tone difference matrix, and 14 gray-level dependence matrix. The original DBT images were also filtered by 8 filters to generate transformed DBT images. The first-order statistics and texture features were then calculated from the transformed images. Detailed descriptions of image preprocessing and parameter settings can be found in the Pyradiomics documentation (<https://pyradiomics.readthedocs.io/>) and Supplementary S1. Handcrafted feature

Main points

- The features of handcrafted and deep radiomics were extracted and analyzed from the digital breast tomosynthesis images of 185 patients.
- The discriminative performance of tumoral and peritumoral regions was evaluated.
- The discriminative performance of the clinical model, radiomics signature, and BI-RADS assessment was evaluated and compared. A radiomics nomogram incorporating the radiomics signature and important clinical factors (age and menstruation status) was developed for potential clinical applications.

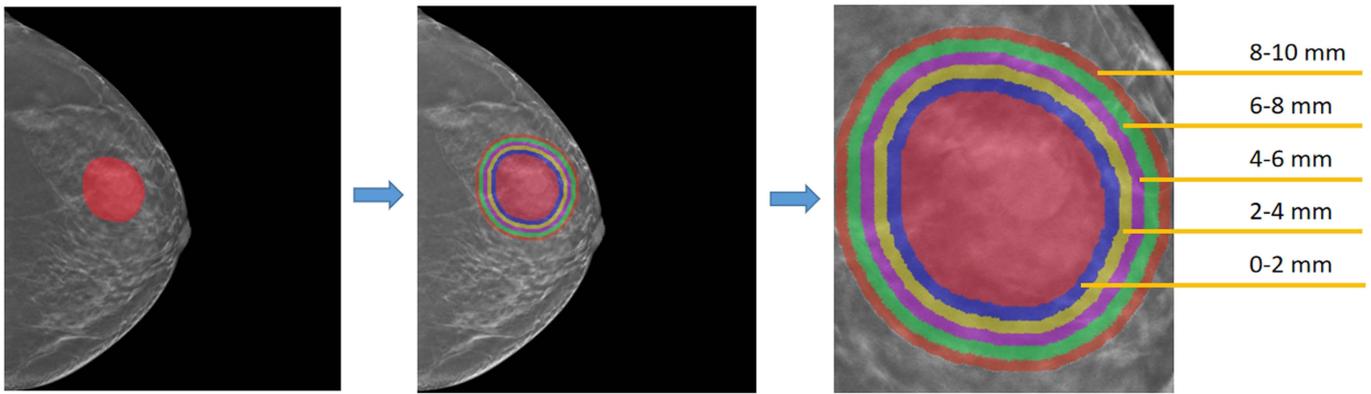


Figure 1. An example of a malignant breast lesion and the dilated masks in the digital breast tomosynthesis (DBT) image. The red-colored region represents the original tumoral region that was manually segmented by radiologists. The colored rings outside the tumoral region indicate the radially dilated regions. Each ring is 2 mm wide.

extraction methodology and filter types were described in Supplementary S2.

Extraction of deep features

The VGG16 network containing 13 convolutional layers, 3 max pooling layers, and 3 fully connected layers was used to calculate the deep features. Detailed descriptions of the architecture of the VGG16 network can be found in the previous report that proposed the VGG model.³² To adapt to our task, the fully connected layers of the original VGG16 were replaced by 3 new layers with neuron numbers of 256, 256, and 2. The transfer learning strategy³³ was used to avoid over-fitting. Our adapted VGG16 network was pretrained with 1.2 million natural images from the ImageNet database³⁴ before being fed with the DBT data. Data augmentation was also performed using the ImageDataGenerator package.³⁵ Each DBT image was shifted up and down, then rotated by 5° randomly to enlarge the data amount. The DBT images were then resized to 224 × 224 and used as inputs to train the adapted VGG16 model with cross-entropy loss function with a learning rate of 0.0001 and Adam optimizer. To calculate the deep features, the DBT slice with the largest lesions was fed into the adapted VGG16 model. The features' vectors contain 512 elements derived and reserved from the second and third to the last layer of the fully connected layers of the adapted VGG16.

Radiomics feature selection and fusion

A 3-step method was used to select the discriminative features. First, we compared the features between malignant and benign groups using the Mann-Whitney U test. The features with $P < .05$ were

considered as significant variables and reserved. Afterward, the least absolute shrinkage and selection operator (LASSO) was applied to select the best feature data set with 10-fold cross-validation for selecting the parameter lambda using the 1 standard error of the minimum criteria (the 1-SE criteria) in R language (v.3.6; available from URL: <https://www.r-project.org>).³⁶ The LASSO algorithm conducts feature selection and regularization for the improvement of model prediction accuracy. Detailed descriptions of the LASSO algorithm were shown in Supplementary S3. The features with non-zero LASSO coefficients were retained. To realize feature-level fusion of the handcrafted and deep features, the 2 features sets were combined, then further selected with the logistic regression model using Akaike information criterion as the stopping rule in R language v3.6.

Development of the clinical model, radiomics signature, and nomogram

The discriminative clinical characteristics were identified by the univariate logistic regression. The multivariate logistic regression was applied to the clinical characteristics with $P < .05$ to construct a clinical model. The radiomics signature was built based on the selected features weighted by LASSO coefficients. The “rms” package in R language v3.6 was used to develop the radiomics nomogram by integrating the radiomics signature within the second row and discriminative clinical characteristics within the third and fourth row.

Statistical analysis

Clinical characteristics (age, menstruation status, history of biopsy, and family history

of breast cancer) and the BI-RADS assessment were retrieved from the EMRS of our hospital. Mann-Whitney U test and chi-square test were performed in R language v3.6 on continuous and discrete variables, respectively. A two-sided P value of $<.05$ was considered statistically significant. The feature reliability was evaluated with intra-class correlation coefficient (ICC)³⁷ using 40 randomly selected patients, 20 with benign breast lesions and 20 with malignant breast lesions. The receiver operating characteristic (ROC) curves were plotted using the “pROC” package with the optimal cutoff value obtained using the maximum Youden index.³⁸ The ROC curves of each model were compared with the Delong test in R language v3.6. The unsupervised cluster analysis, as a widely used clustering algorithm in illustrating the discriminability of the features and distribution of all samples, was performed using the “scipy” package in Python v3.6. Calibration curves were drawn to depict the calibration of the nomogram in the training and validation cohorts. Decision curve analysis (DCA) was performed in R language v3.6 to assess the clinical usefulness of the models. Figure 2 illustrates the workflow of this study. All the R packages used in this study were listed in Supplementary S4.

Results

Table 1 shows the statistical analysis results of clinical characteristics and BI-RADS assessment. The results indicated that there were no statistical differences between the benign and malignant groups in types of family history of breast cancer and history of biopsy ($P > .25$). Age, menstruation status, and BI-RADS assessment were significantly different between the malignant and

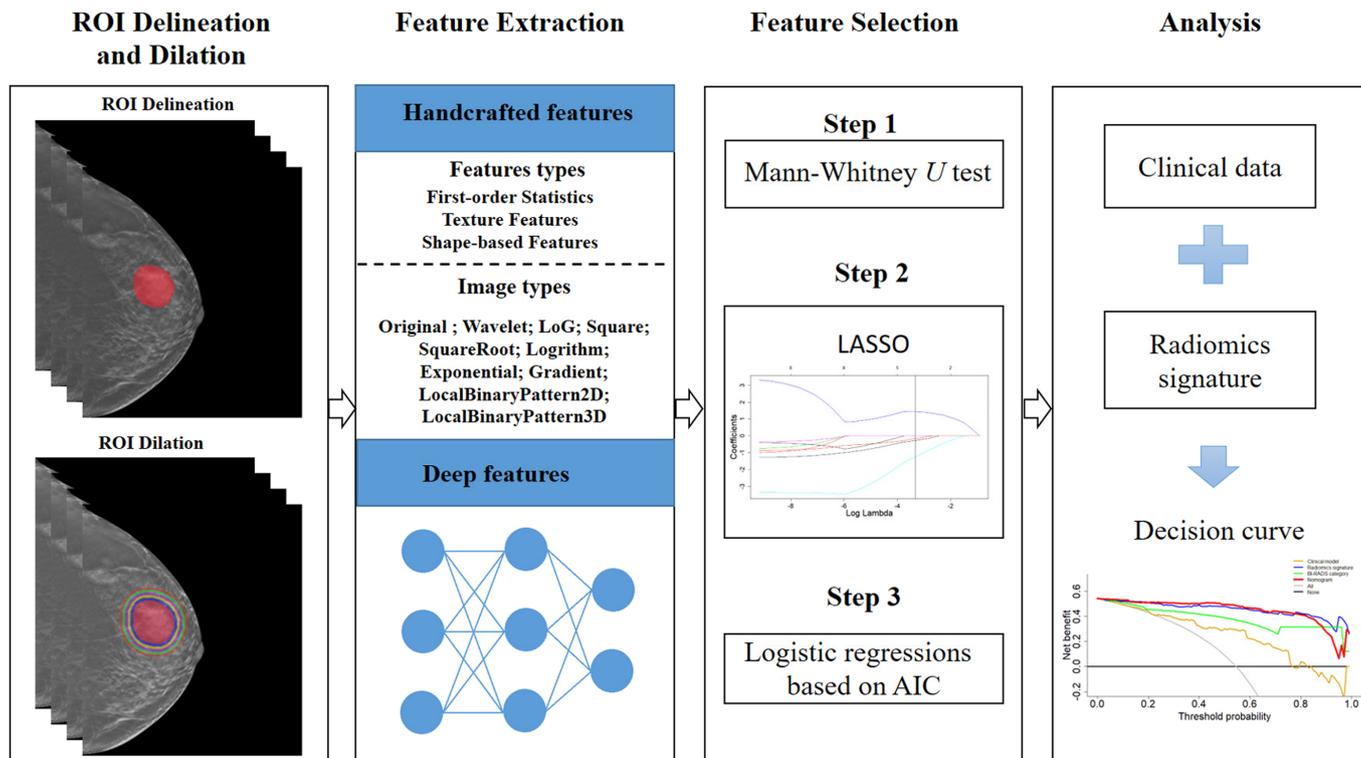


Figure 2. Overall workflow of this study. ROI, region of interest.

benign groups ($P < .001$, $P = .003$, and $P < .001$ in the training cohort, respectively; all $P < .001$ in the validation cohort).

The discriminative performance of logistic regression models based on handcrafted and deep features selected from tumoral and the peritumoral regions was

compared and listed in Supplementary Table S1. Bold type indicates the best model with low over-fitting and high AUCs. Figure 3 shows the ROC curves of each model. The results indicated that the handcrafted and deep features extracted from the peritumoral regions with 2 mm dilation

distances showed the best discriminative performance, with AUCs of 0.777 and 0.851 (handcrafted features), and 0.973 and 0.975 (deep features) in the training and validation cohorts, respectively.

Table 2 lists the 4 features selected from the peritumoral regions with 2 mm dilation distances in the DBT image, 2 of handcrafted and 2 of deep features. Supplementary Figure S1 shows the feature selection using the LASSO algorithm. All features exhibited favorable discriminative power in training and validation cohorts, with $AUC > 0.700$ and $P < .001$. The features have good consistency with ICCs all over 0.83. The results revealed that deep features were more discriminative than handcrafted features (compare the AUCs of the deep features with those of the handcrafted features in Table 2).

To evaluate the statistical distributions of the 4 selected features, boxplot analysis was performed as shown in Figure 4. All features showed statistically significant ($P < .001$) differences between the malignant and benign groups. The values of `log_sigma_3_0_mm_3D_glszm_SizeZoneNonUniformityNormalized` feature in the benign group tend to be larger than those in the malignant group, while the values of `lbp_2D_glrIm_RunEntropy` feature in the benign group tend to be smaller than

Characteristic	Training cohort		P	Validation cohort		P
	Benign (n=56)	Malignant (n=67)		Benign (n=29)	Malignant (n=33)	
Age (years), mean ± SD	44.6 ± 10.4	52.1 ± 9.0	<.001	39.9 ± 8.6	55.5 ± 8.1	<.001
Family history of breast cancer, n (%)			1.0			.26
+	4 (7.1)	5 (7.5)		2 (6.9)	6 (18.2)	
–	52 (92.9)	62 (92.5)		27 (93.1)	27 (81.8)	
History of biopsy, n (%)			1.0			.60
+	2 (3.6)	2 (3.0)		2 (6.9)	1 (3.0)	
–	52 (96.3)	65 (97.0)		27 (93.1)	32 (97.0)	
Menstruation status, n (%)			.003			<.001
+	13 (23.2)	34 (50.7)		3 (10.3)	23 (69.7)	
–	43 (76.8)	33 (49.3)		26 (89.7)	10 (30.3)	
BI-RADS, n (%)			<.001			<.001
0, 1, 2, 3	18 (32.1)	0 (0.0)		11 (37.9)	1 (3.0)	
4A, 4B, 4C	38 (67.9)	52 (77.6)		18 (62.1)	20 (60.6)	
5, 6	0 (0.0)	15 (22.4)		0 (0.0)	11 (33.3)	

SD, standard deviation; BI-RADS, breast imaging reporting and data system.

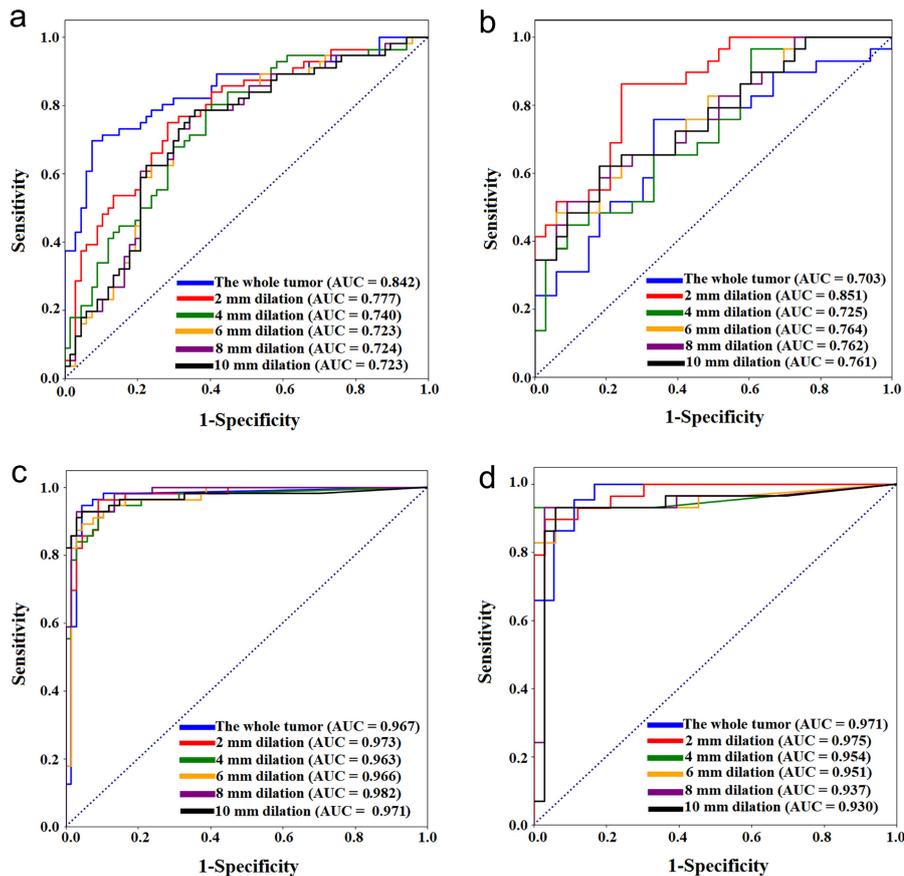


Figure 3. a-d. ROC curves of the logistic regression models based on tumoral and peritumoral regions: (a) and (b), logistic regression models using handcrafted features in the training (a) and validation (b) cohorts; (c) and (d), logistic regression models using deep features in the training (c) and validation (d) cohorts. AUC, area under the curve; ROC, receiver operating characteristic.

those in the malignant group. Unsupervised cluster analysis was performed on the selected features and patients to represent similarity and affinity between all samples as shown in Supplementary Figure S2. The 2 deep features exhibited obvious clusters between the benign and malignant groups. Benign and malignant breast lesions can be roughly separated by the deep features.

The radiomics signature was developed by incorporating the 4 selected features weighted by the non-zero LASSO coefficients and shown as follows: Radiomics signature = $-0.008193566 + FC2_24 \times 29.427074367 + FC2_44 \times -13.508175116 + \log_sigma_3_0_mm_3D_glszm_SizeZoneNonUniformityNormalized \times 3.140812585 + lbp_2D_gIrlm_RunEntropy \times -0.339913263$.

Figure 5a shows our constructed clinical radiomics nomogram, which includes the radiomics signature within the second row and age and menstrual status within the third and fourth row, respectively. Each predictor with a given value can be mapped to the point axis. The sum of the points can be referred to in the total point axis. Afterward, users can map the total points to the linear predictor axis and the probability axis at the bottom of the nomogram to obtain the values of linear predictor and the probability. The risk of being a malignant lesion can be read off the scale in the last row by vertically drawing a line from the total points.

Calibration curves for the training and validation cohorts were shown in Figure 5a, 5c, indicating the acceptable agreements between the nomogram-estimated probability and the actual outcomes of the breast lesions (i.e., the malignancy rate). The x- and y-axis represented the model-calculated and the actual probabilities, respectively. The 45° blue line and the red dotted line represent an ideal diagnosis and the performance of our nomogram, respectively. The closer the red dotted line is to the blue line, the better the discriminative performance of the nomogram.

Table 3 evaluates the diagnostic performances of the clinical model, BI-RADS assessment, radiomics signature, and nomogram. The results showed that the clinical model performed the worst among the 4 models. The radiomics signature was significantly superior to the clinical model and BI-RADS assessment in terms of AUC, accuracy, sensitivity, and specificity. The nomogram integrating the radiomics signature and important clinical factors (age and menstruation status) achieved the

Table 2. Discriminative performance of the selected handcrafted and deep features

Features	Cohorts	Mean ± SD		AUC (95% CI)	P	ICC
		Benign	Malignant			
FC2_24	Training	0.074 ± 0.030	0.017 ± 0.018	0.977 (0.953-1.000)	<.001	0.89
	Validation			0.926 (0.859-0.993)	<.001	
FC2_44	Training	0.007 ± 0.017	0.097 ± 0.056	0.973 (0.944-1.000)	<.001	0.83
	Validation			0.937 (0.873-1.000)	<.001	
log_sigma_3_0_mm_3D_glszm_SizeZoneNonUniformityNormalized	Training	0.116 ± 0.063	0.074 ± 0.036	0.733 (0.645-0.821)	<.001	0.97
	Validation			0.701 (0.568-0.836)	<.001	
lbp_2D_gIrlm_RunEntropy	Training	3.162 ± 0.228	3.371 ± 0.250	0.701 (0.568-0.836)	<.001	0.95
	Validation			0.704 (0.573-0.836)	<.001	

SD, standard deviation; AUC, area under the ROC; ICC, intraclass correlation.

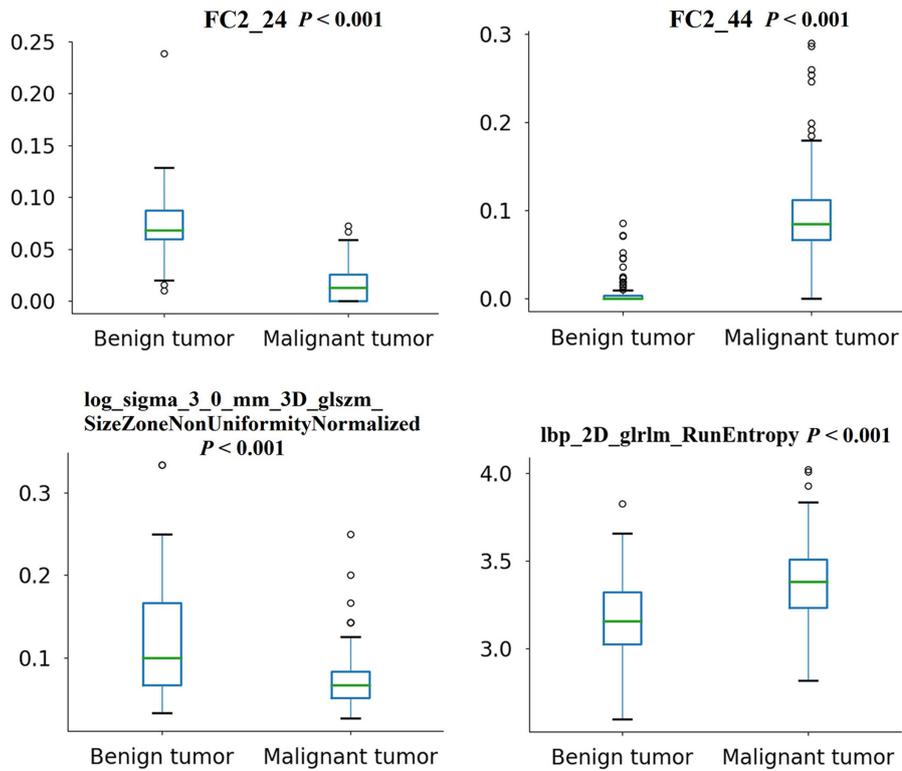


Figure 4. Boxplots of the 4 selected features from the DBT image.

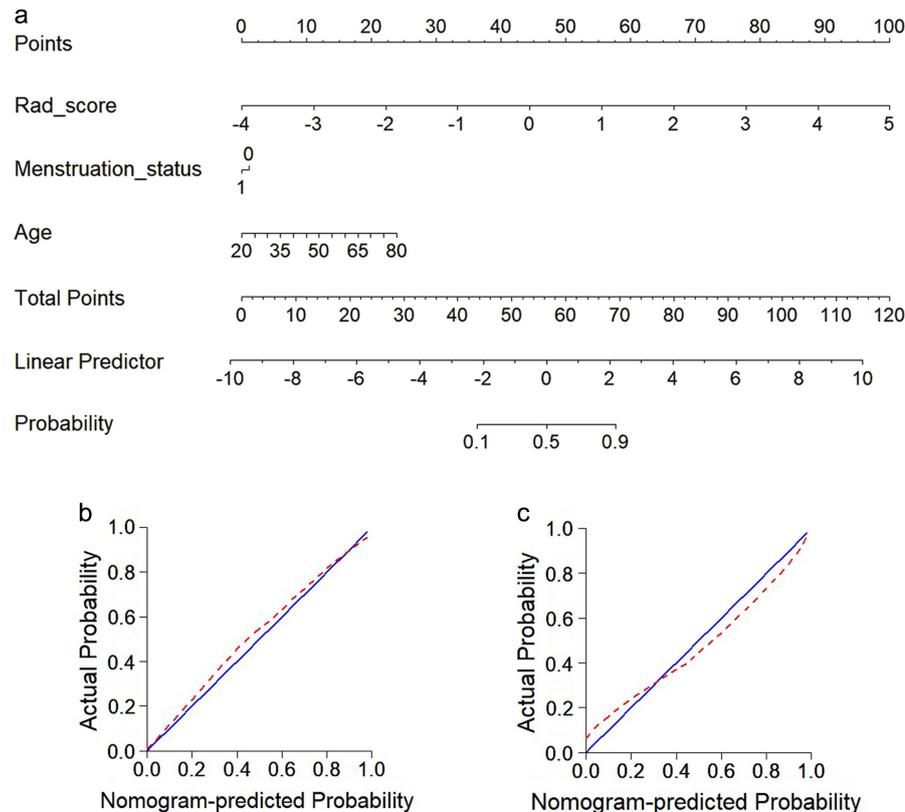


Figure 5. a-c. Development and validation of the nomogram based on DBT data: (a), the developed nomogram model; (b) and (c), calibration curves of the model in training (b) and validation (c) cohort, respectively.

best discriminative ability with an AUC of 0.980 (95% CI, 0.960 to 1.000) in the training cohort and 0.985 (95% CI, 0.960 to 1.000) in the validation cohort, respectively. Figure 6 depicts the ROC curves of each model. The ROC curve of the clinical model was significantly different from the BI-RADS, radiomics signature, and nomogram ($P = .006$, $P < .001$, and $P < .001$ in the training cohort, respectively; $P = .072$, $P < .001$, and $P < .001$ in the validation cohort, respectively) according to the DeLong test. The ROC curve of the BI-RADS category was significantly different from the radiomics signature and nomogram ($P = .021$ and $P = .067$ in the training cohort, respectively; $P = .008$ and $P = .032$ in the validation cohort, respectively). There were no significant differences in the ROC curves between the radiomics signature and nomogram ($P = .76$ in the training cohort; $P = .72$ in the validation cohort).

The decision curves of the clinical model, radiomics signature, BI-RADS assessment, and nomogram were presented in Figure 7. Greater benefit can be obtained from the nomogram when the threshold probability is more than .04. The results indicated that the nomogram had the best clinical utility for discrimination of breast cancer compared with the clinical model, radiomics signature, and BI-RADS assessment.

Discussion

In this study, we enrolled 185 patients who underwent DBT scans and whose lesions were pathologically confirmed as either benign or malignant. This was far more than early studies that enrolled 26,²⁹ 39,³⁰ and 40¹³ patients with DBT examinations. We evaluated and compared the discriminative performance of tumoral and peritumoral regions in the DBT image by calculating the radiomics features from the tumoral and peritumoral regions with different dilation distances outside the primary lesions. Our results revealed that the features from the peritumoral regions at 2 mm dilation distances achieved the best discriminative performance for both the hand-crafted and deep features (Supplementary Table S1), suggesting that the peritumoral regions in the DBT image were important for diagnosing breast cancer. The results were consistent with previous reports that demonstrated the discriminative power of peritumoral regions since tumor cells tend

Table 3. Evaluation of the clinical model, BI-RADS assessment, radiomics signature, and nomogram

	Training cohort					Validation cohort				
	AUC (95% CI)	Acc	Sen	Spe	<i>P</i>	AUC (95% CI)	Acc	Sen	Spe	<i>P</i>
M1	0.794 (0.711-0.877)	0.756	0.661	0.761		0.777 (0.711-0.877)	0.700	0.724	0.758	
M2	0.918 (0.877-0.970)	0.840	0.881	0.804		0.898 (0.822-0.976)	0.839	0.848	0.828	
M3	0.974 (0.954-0.997)	0.946	0.970	0.927		0.986 (0.954-1.000)	0.862	0.939	0.931	
M4	0.980 (0.960-1.000)	0.935	0.970	0.946		0.985 (0.960-1.000)	0.903	0.909	0.966	
M1 vs. M2					.006					.072
M1 vs. M3					<.001					<.001
M1 vs. M4					<.001					<.001
M2 vs. M3					.021					.067
M2 vs. M4					.008					.032
M3 vs. M4					.76					.72

BI-RADS, breast imaging reporting and data system assessment; AUC, area under the ROC curve; 95% CI, 95% confidence interval; Acc, accuracy; Sen, sensitivity; Spe, specificity; M1, clinical model; M2, M3, radiomics signature; M4, nomogram.

to migrate from primary tumor to peritumoral regions, which would cause morphological changes in clinical imaging.³⁹⁻⁴¹

We calculated 1967 handcrafted and 512 deep features from each patient to obtain richer information from the DBT image. This was far more than a previous study that only calculated 104 features.¹³ The results showed that the deep features exhibited more discriminative power than the handcrafted features (compare the AUC, accuracy, sensitivity, and specificity of each model in Table 2). Logistic regression models built with deep features always outperformed those built with handcrafted features. The result was consistent with previous studies that also found the deep features were superior to the traditional handcrafted features.^{42,43} This may be partially explained by considering the fact that the handcrafted features belong to low-order or simple high-order features, which contain limited information, when compared with the deep features that involve high-dimension information.⁴⁴ Among the 2479 features, a total of 4 features were identified as the most discriminative features (AUCs > 0.700 and *P* < .001), 1 of log_sigma_3_0_mm_3D_glszm, 1 of lbp_2D_glrIm, and 2 of the deep features. The 2 handcrafted features both belonged to texture feature classes. This was consistent with previous reports by Kontos's group,^{45,46} who also found that texture features were closely correlated to breast cancer in the DBT image. The lbp_2D_glrIm_RunEntropy feature reflects the gray-level change characteristics in

local texture features. The higher value of this feature indicates the more complicated texture and heterogeneity of the tumoral region, which suggests that the tumor tends to be malignant. Our results showed that the values of this feature were higher in the malignant patients than those in benign patients (Figure 4). The log_sigma_3_0_mm_3D_glszm feature, on the other hand, quantifies the changes of gray level in connected regions of image edges, which reflects the characteristics of tumor edges. Our results revealed that the benign lesions tend to have more clear edges since the values of this feature were bigger in benign patients compared to those in malignant patients (Figure 4). This was consistent with a previous report that also suggested the edge characteristics of breast lesions were important in the differentiation between benign and malignant lesions.⁴⁷

The 4 selected features were used to build the radiomics signature and generated good accuracy to differentiate malignant lesions with AUCs of 0.974 and 0.986 in the training and validation cohorts, respectively. The age and menstruation status were identified as the most important clinical characteristics in differentiating the benign and malignant lesions. Then, a clinical model integrating age and menstruation status was built and showed poor discriminative performance with AUCs of 0.794 and 0.777 in the training and validation cohorts, respectively. To facilitate clinicians on potential utilization of our radiomics methods, a radiomics nomogram incorporating

the radiomics signature and important clinical factors (age and menstruation status) was developed, and it achieved the highest discriminative performance, which outperformed the results of the BI-RADS assessment (AUCs, nomogram vs. BI-RADS, 0.980 vs. 0.918 and 0.985 vs. 0.898 in training and validation cohorts, respectively). The nomogram was also superior to the results of the BI-RADS assessment in terms of missed diagnosis rates and misdiagnosis rates (Table 3). The results indicated that our nomogram had better diagnostic ability than the classical mammographic-DBT evaluation method. The DCA demonstrated that more net benefit can be obtained for the patients by utilizing our nomogram compared with the clinical model and BI-RADS assessment, which indicates good potential in clinical uses. Therefore, we suggest that our nomogram integrating the radiomics signature and important clinical factors (age and menstruation status) can be considered as a useful tool in assisting clinicians in the early screening of breast cancer using DBT scans. To use our nomogram, clinicians should draw the ROIs on each DBT slice and calculate the probability of a patient having benign or malignant lesions based on the radiomics signature formula and values of the patient's age and menstruation status. Afterward, clinicians can incorporate the obtained probability values with other clinical manifestations of the patient to make comprehensive decisions.

There were some limitations in our study. First, the enrolled patients were from the

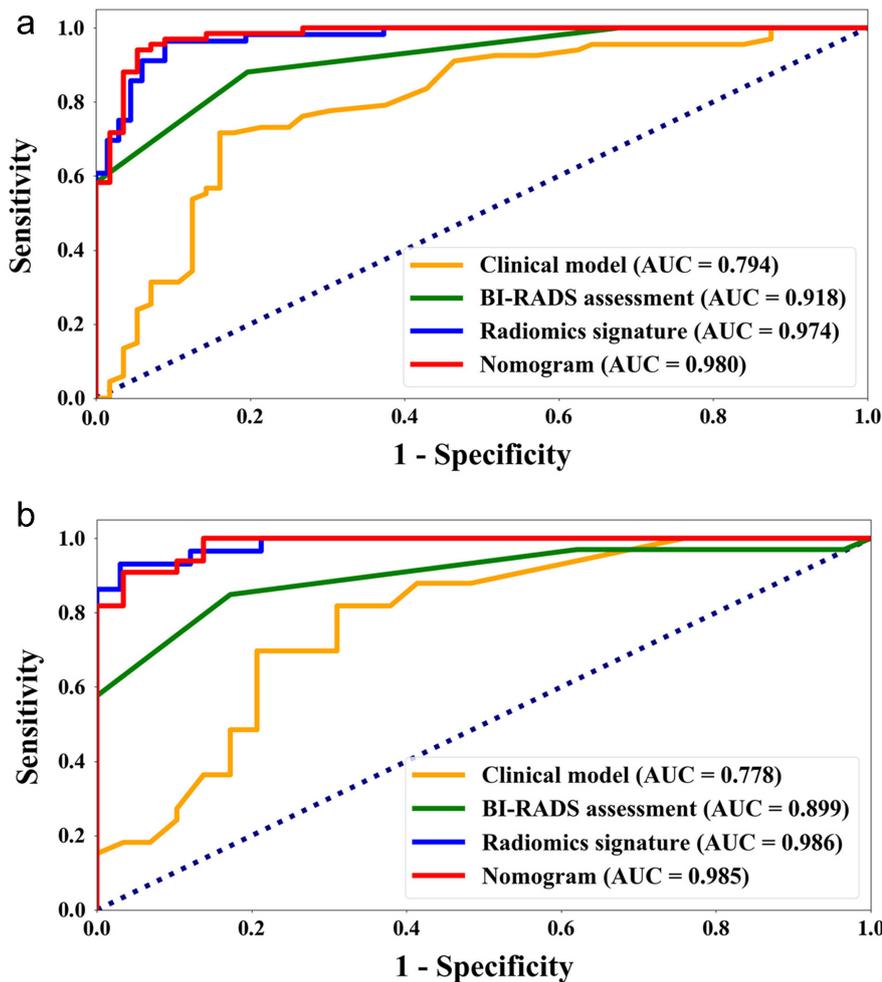


Figure 6. a, b. ROC curves of the clinical model, BI-RADS assessment, radiomics signature, and nomogram in the training (a) and validation (b) cohorts. The orange line indicates the clinical model. The green line represents BI-RADS assessment. The blue line indicates the radiomics signature. The red line represents the nomogram. BI-RADS, breast imaging reporting and data system. AUC, area under the curve.

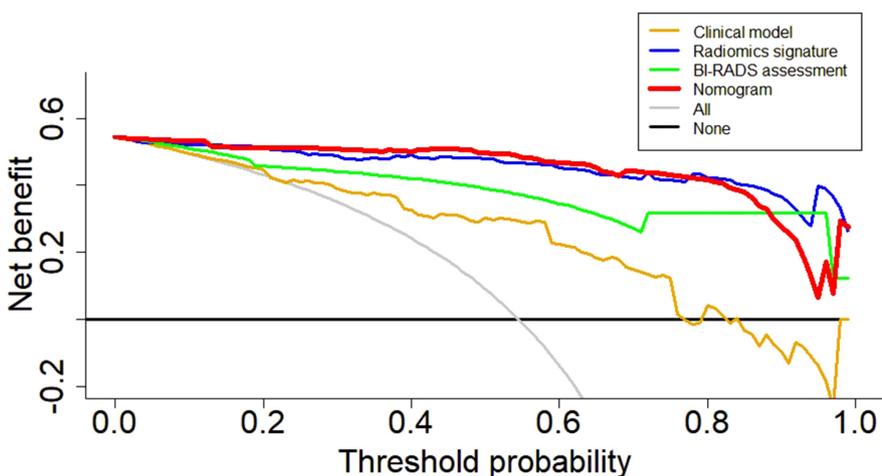


Figure 7. Decision curve analysis for the clinical model, radiomics signature, BI-RADS assessment, and nomogram. The black line shows the hypothesis that the patients were all benign and not treated. The gray line shows the assumption that the patients were all malignant and all treated. The orange line indicates the treatment decision made by the clinical model. The green line indicates the treatment decision made by the BI-RADS assessment. The blue line indicates the treatment decision made by the radiomics signature. The red line indicates the treatment decision made by the nomogram.

same hospital. We are working with multicenter research teams to enroll patients from other hospitals to assess the robustness of the developed models. Second, the assessment of the effects of breast density on peritumoral radiomics was missing. This is because the study was conducted in China with most samples having dense breasts. Patients with different breast densities would be enrolled in our future study. Third, the ROIs on each slice were manually segmented, which increased the workload. Further studies are required to build deep learning-based segmentation methods to segment the lesions automatically in the DBT image.

In conclusion, our findings demonstrated the discriminative values of tumoral and peritumoral regions in the breast DBT image. The deep features were important in addition to the traditional handcrafted features in the differentiation of benign and malignant lesions. Our radiomics nomogram exhibited good discriminative performance, which suggests its potential in assisting the diagnosis of DBT mammograms.

Financial disclosure

This work was supported by funding from the Youth Science and Technology Innovation Leader Support Project [RC170497], Natural Science Foundation of Liaoning Province of China [201602450], National Natural Science Foundation of China [U1708261], National Key R&D Program of Ministry of Science and Technology of China [2016YFC1303002] and Supporting Fund for Big Data in Health Care [HMB201903101], Special Fund for Research in the Public Interest of China [201402020], 2020 Key Project of Double Service for Universities in Shenyang.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- McGuire A, Brown JA, Malone C, McLaughlin R, Kerin MJ. Effects of age on the detection and management of breast cancer. *Cancers*. 2015;7(2):908-929. [CrossRef]
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7-30. [CrossRef]
- Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. *Breast Cancer Res Treat*. 2018;169(2):217-229. [CrossRef]

4. Michell MJ. Breast screening review—a radiologist's perspective. *Br J Radiol.* 2012;85(1015):845-847. [\[CrossRef\]](#)
5. Pauwels EK, Foray N, Bourguignon MH. Breast cancer induced by X-ray mammography screening? A review based on recent understanding of low-dose radiobiology. *Med Princ Pract.* 2016;25(2):101-109. [\[CrossRef\]](#)
6. Sharpe RE, Venkataraman S, Phillips J, et al. Increased cancer detection rate and variations in the recall rate resulting from implementation of 3D digital breast Tomosynthesis into a population-based screening program. *Radiology.* 2016;278(3):698-706. [\[CrossRef\]](#)
7. Giess CS, Pourjabbar S, Ip IK, et al. Comparing diagnostic performance of digital breast tomosynthesis and full-field digital mammography in hybrid screening environment. *AJR Am J Roentgenol.* 2017;9:929-934.
8. Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. *J Am Coll Radiol.* 2016;13(11S):R61-R66. [\[CrossRef\]](#)
9. Niklason LT, Kopans DB, Hamberg LM. Digital breast imaging: tomosynthesis and digital subtraction mammography. *Breast Dis.* 1998;10(3-4):151-164. [\[CrossRef\]](#)
10. Riedl CC, Luft N, Bernhart C, et al. Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density. *J Clin Oncol.* 2015;33(10):1128-1135. [\[CrossRef\]](#)
11. Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. *Clin Radiol.* 2016;71(2):141-150. [\[CrossRef\]](#)
12. Rafferty EA, Georgian-Smith D, Kopans DB, et al. Comparison of full-field digital tomosynthesis with two view conventional film screen mammography in the prediction of lesion malignancy [abstr]. *Radiology.* 2002;225:268.
13. Gaur S, Dialani V, Slanetz PJ, et al. Architectural distortion of the breast. *AJR Am J Roentgenol.* 2018;201:662-670.
14. Tagliafico AS, Valdora F, Mariscotti G, et al. An exploratory radiomics analysis on digital breast tomosynthesis in women with mammographically negative dense breasts. *Breast.* 2018;40:92-96. [\[CrossRef\]](#)
15. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006. [\[CrossRef\]](#)
16. Zhou M, Scott J, Chaudhury B, et al. Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. *AJNR Am J Neuroradiol.* 2018;39(2):208-216. [\[CrossRef\]](#)
17. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep.* 2015;5:13087. [\[CrossRef\]](#)
18. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749-762. [\[CrossRef\]](#)
19. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol.* 2019;25(6):485-495. [\[CrossRef\]](#)
20. Mavroforakis M, Georgiou H, Dimitropoulos N, Cavouras D, Theodoridis S. Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines. *Eur J Radiol.* 2005;54(1):80-89. [\[CrossRef\]](#)
21. Verma B, Mcleod P, Klevansky A. Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. *Expert Syst Appl.* 2010;37(4):3344-3351. [\[CrossRef\]](#)
22. Brem RF. Clinical versus research approach to breast cancer detection with CAD: where are we now? *AJR Am J Roentgenol.* 2007;188:1475-1482.
23. Ma W, Zhao Y, Ji Y, et al. Breast cancer molecular subtype prediction by mammographic radiomic features. *Acad Radiol.* 2019;26(2):196-201. [\[CrossRef\]](#)
24. Hai J, Tan H, Chen J, et al. Multi-level features combined end-to-end learning for automated pathological grading of breast cancer on digital mammograms. *Comput Med Imaging Graph.* 2019;71:58-66. [\[CrossRef\]](#)
25. Lee SE, Han K, Kwak JY, Lee E, Kim EK. Radiomics of US texture features in differential diagnosis between triple-negative breast cancer and fibroadenoma. *Sci Rep.* 2018;8(1):13546. [\[CrossRef\]](#)
26. Xiong Q, Zhou X, Liu Z, et al. Multiparametric MRI-based radiomics analysis for prediction of breast cancers insensitive to neoadjuvant chemotherapy. *Clin Transl Oncol.* 2020;22(1):50-59. [\[CrossRef\]](#)
27. Antunovic L, De Sanctis R, Cozzi L, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. *Eur J Nucl Med Mol Imaging.* 2019;46(7):1468-1477. [\[CrossRef\]](#)
28. Samala RK, Wei J, Chan HP, et al. First and second-order features for detection of masses in digital breast tomosynthesis. *Med Imaging Computer-Aided Diagn.* 2016;9:785.
29. Chan HP, Wei J, Sahiner B, et al. Computer-aided detection system for breast masses on digital tomosynthesis mammograms: preliminary experience. *Radiology.* 2005;237(3):1075-1080. [\[CrossRef\]](#)
30. Bevilacqua V, Brunetti A, Guerriero A, Trotta GF, Telegrafo M, Moschetta M. A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cogn Syst Res.* 2019;53:3-19. [\[CrossRef\]](#)
31. Van Griethuysen JMM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104-e107. [\[CrossRef\]](#)
32. Simonyan K, Zisserman A. Very deep convolutional net-works for large-scale image recognition. *Comput Ence.* 2014:1409.1556.
33. Lao J, Chen Y, Li ZC, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep.* 2017;7(1):10353. [\[CrossRef\]](#)
34. Huang G, Liu Z, Laurens VDM, et al. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition 2017.*
35. Ciritis A, Rossal C, Martini LVD, et al. Determination of mammographic breast density using a deep convolutional neural network. *Br J Radiol.* 2018;92:20180691.
36. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med.* 2007;26(30):5512-5528. [\[CrossRef\]](#)
37. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155-163. [\[CrossRef\]](#)
38. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden index and optimal cut point estimated from observations affected by a lower limit of detection. *Biom J.* 2008;50(3):419-430. [\[CrossRef\]](#)
39. Zhuo Y, Feng M, Yang S, et al. Radiomics nomograms of tumors and peritumoral regions for the preoperative prediction of spread through air spaces in lung adenocarcinoma. *Transl Oncol.* 2020;13(10):100820. [\[CrossRef\]](#)
40. Braman N, Prasanna P, Whitney J, 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. [\[CrossRef\]](#)
41. Pérez-Morales J, Tunalı I, Stringfield O, et al. Peritumoral and intratumoral radiomic features predict survival outcomes among patients diagnosed in lung cancer screening. *Sci Rep.* 2020;10(1):10528. [\[CrossRef\]](#)
42. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. *IEEE Signal Process Mag.* 2019;36(4):132-160. [\[CrossRef\]](#)
43. Cai H, Huang Q, Rong W, et al. Breast microcalcification diagnosis using deep convolutional neural network from digital mammograms. *Comput Math Methods Med.* 2019;2019:2717454. [\[CrossRef\]](#)
44. Hua W, Xiao T, Jiang X, et al. Lymph-vascular space invasion prediction in cervical cancer: exploring radiomics and deep learning multi-level features of tumor and peritumor tissue on multiparametric MRI. *Biomed Signal Process Control.* 2020;58. [\[CrossRef\]](#)
45. Kontos D, Ikejimba LC, Bakic PR, Troxel AB, Conant EF, Maidment AD. Analysis of parenchymal texture with digital breast tomosynthesis: comparison with digital mammography and implications for cancer risk assessment. *Radiology.* 2011;261(1):80-91. [\[CrossRef\]](#)
46. Kontos D, Bakic PR, Carton AK, Troxel AB, Conant EF, Maidment AD. Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: a preliminary study. *Acad Radiol.* 2009;16(3):283-298. [\[CrossRef\]](#)
47. Flobbe K, Bosch AM, Kessels AGH, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med.* 2003;163(10):1194-1199. [\[CrossRef\]](#)

Supplementary data

S1. Radiomics feature extraction

Normalization is based on all gray values in the DBT image. The normalizations of the gray level values in the DBT image were performed using the following function: $f(x) = s(x - \mu) / \sigma$, where: x and $f(x)$ are the original and normalized intensity, respectively. μ and σ are the mean and standard deviation of the image intensity values. The s is an optional scaling defined by scale and was set to 1 by default. A B-spline interpolator is used in resampling the DBT image and resampling to a voxel of

$0.1 \times 0.1 \times 1.0 \text{ mm}^3$. Nodule region intensity values were discretized using a bin width of 25 Hounsfield units.

S2. Radiomics feature extraction methodology and filter types

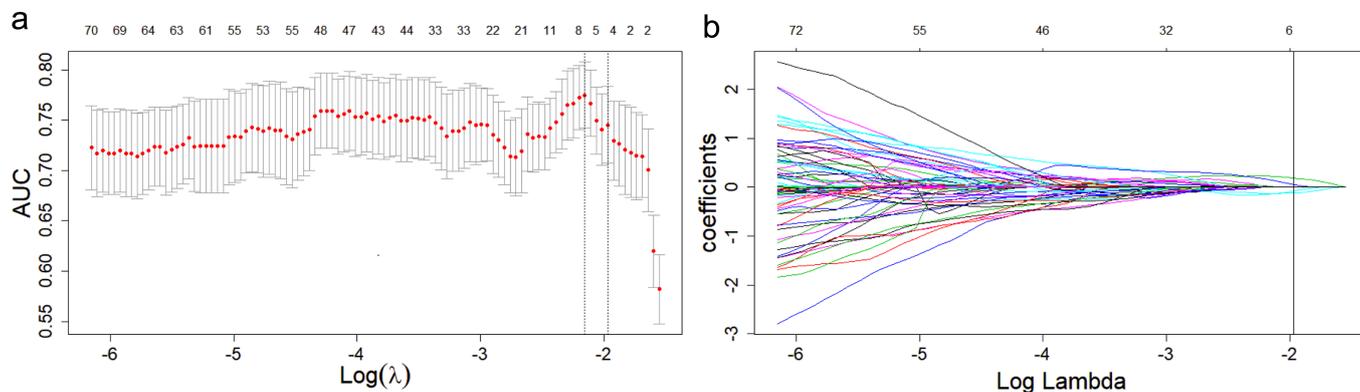
For each patient, a total of 1967 handcrafted features were calculated. The features include 3 types: (i) 18 first-order statistics features, (ii) 14 shape-based features and (iii) 75 textural features. The handcrafted features were extracted using 'Pyradiomics (v.2.2.0)'.

First-order statistics features: First-order statistics describe the distribution of voxel

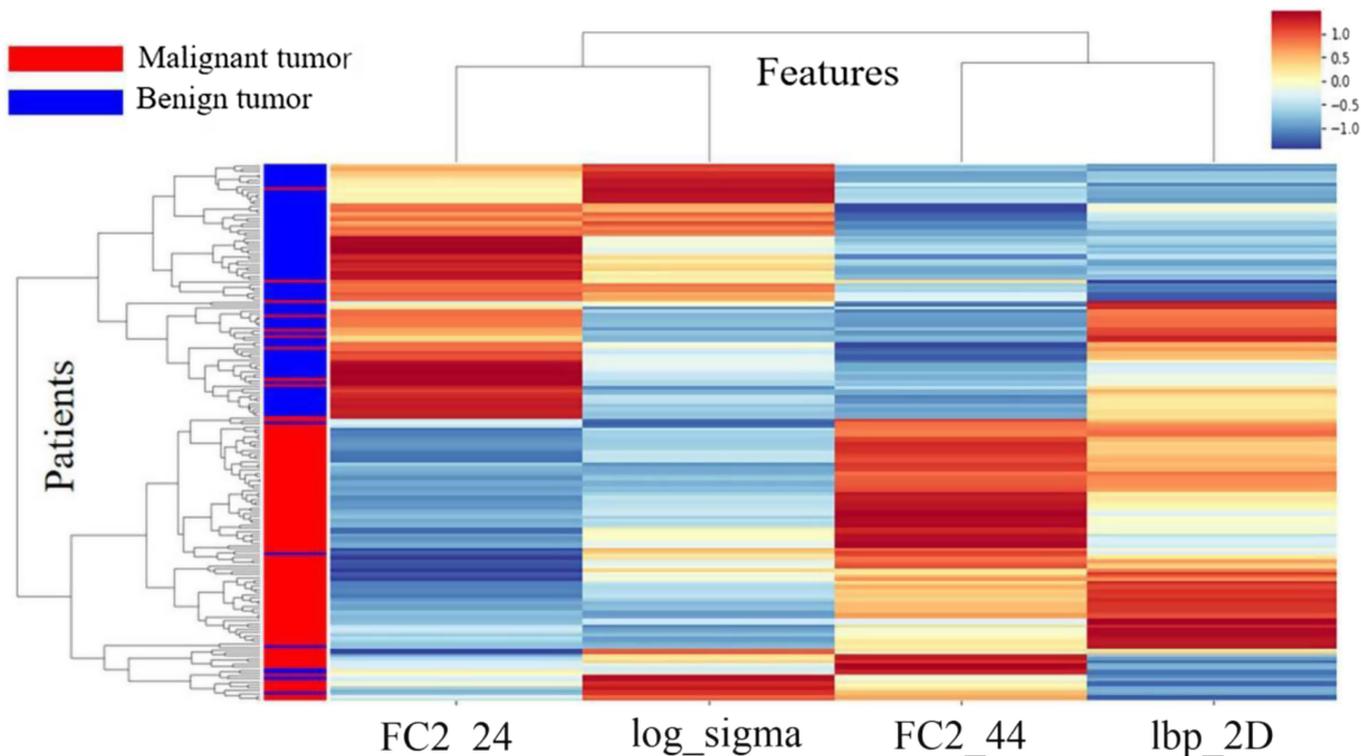
intensities within the DBT image region defined by the mask through commonly used and basic metrics.

Shape-based features: Shape-based features we included were descriptors of the 3-dimensional and 2-dimensional size and shape of the ROI. These features are independent from the gray level intensity distribution in the ROI and are therefore only calculated on the non-derived image and mask.

Textural features: Textural features represent the textural characteristics and reflect the homogeneity phenomenon in medical images. The 75 textural features include



Supplementary Figure S1. a, b. Radiomics feature selection using the least absolute shrinkage and selection operator (LASSO) logistic regression. (a) Tuning parameter (λ) selection in the LASSO model used 10-fold cross-validation via minimum criteria. (b) LASSO coefficient profiles of the radiomics features.



Supplementary Figure S2. Unsupervised cluster analysis of the selected features from peritumoral regions in the DBT (digital breast tomosynthesis) image and patients. The x-axis represents the selected features ($n=4$). The y-axis represents the patients ($n=185$). The red color represents patients with malignant tumor, while the blue color indicates the patients with benign tumor.

Supplementary Table 1. Discriminative performance of the logistic regression models based on handcrafted and deep features selected from tumoral and the peritumoral regions

Dilation Distances	Cohorts	Models based on handcrafted features				Models based on deep features			
		AUC (95% CI)	Acc	Sen	Spe	AUC (95% CI)	Acc	Sen	Spe
0 mm	Training cohort	0.842 (0.756–0.906)	0.789	0.732	0.836	0.967 (0.817–0.973)	0.943	0.929	0.955
	Validation cohort	0.703 (0.566–0.830)	0.629	0.414	0.818	0.971 (0.806–0.973)	0.935	0.977	0.833
2 mm*	Training cohort	0.777 (0.666–0.839)	0.715	0.661	0.761	0.973 (0.932–0.996)	0.927	0.946	0.910
	Validation cohort	0.851 (0.725–0.926)	0.710	0.517	0.879	0.975 (0.940–1.000)	0.920	0.897	0.939
4 mm	Training cohort	0.740 (0.652–0.827)	0.667	0.607	0.716	0.963 (0.931–0.995)	0.911	0.911	0.910
	Validation cohort	0.725 (0.598–0.850)	0.677	0.483	0.848	0.966 (0.891–1.000)	0.952	0.897	1.000
6 mm	Training cohort	0.723 (0.632–0.814)	0.683	0.625	0.731	0.966 (0.935–0.998)	0.910	0.911	0.910
	Validation cohort	0.764 (0.646–0.882)	0.694	0.586	0.788	0.951 (0.889–1.000)	0.920	0.897	0.939
8 mm	Training cohort	0.724 (0.633–0.815)	0.675	0.625	0.716	0.982 (0.965–0.999)	0.934	0.929	0.940
	Validation cohort	0.762 (0.643–0.881)	0.694	0.586	0.788	0.937 (0.863–1.000)	0.952	0.931	0.970
10 mm	Training cohort	0.723 (0.631–0.814)	0.683	0.643	0.716	0.971 (0.943–1.000)	0.919	0.929	0.910
	Validation cohort	0.761 (0.642–0.880)	0.726	0.621	0.818	0.930 (0.858–1.000)	0.919	0.897	0.939

AUC, area under the ROC curve; CI, confidence interval; Acc, accuracy; Sen, sensitivity; Spe, specificity.

* The best dilation distance with low over-fitting and high AUCs of the logistic regression model.

24 gray level cooccurrence matrix (GLCM), 16 gray level run length matrix (GLRLM), 16 gray level size zone matrix (GLSZM), 5 neighbour gray tone difference matrix (NGTDM) and 14 gray level dependence matrix (GLDM).

The features calculated from the DBT images filtered were with eight types of filters, including laplacian of gaussian, wavelet, square, square root, logarithm, exponential, gradient and local binary pattern 2D/3D filters. Detailed descriptions of the feature extraction methodology, image types and parameter settings were described in the Pyradiomics documentation (<https://pyradiomics.readthedocs.io/en/latest>).

S3. LASSO details

LASSO is a regularization method that may be used with multiple regression models. The

LASSO with logistic regression model was used in this study. In addition, LASSO has a L1 regularization which due to the penalty equal to the absolute value of the magnitude of coefficients, L1 regularization will result in a sparse model with fewer coefficients. Some coefficients can become zero and removed from the model. In this study, a total of 4 features with nonzero coefficients were selected as the most discriminative features by LASSO. LASSO regression is well-suited for models showing high-levels of variable selection. The “glmnet” package in R language v3.6 was used to perform the LASSO. In addition, a tuning parameter (lambda) was selected to maximize AUC in the total training. A total of 100 lambda values were generated, from which we finally picked the largest value of lambda such that error is within 1 standard error of the minimum criteria (the 1-SE criteria).

S4. R packages

The packages in R language v3.6 used in this study:

- 1) LASSO logistic regression was performed using the “glmnet” package.
- 2) Intra-class correlation coefficient (ICC) was performed using the “irr” package.
- 3) Multivariate logistic regression was performed using the “rms” package.
- 4) Nomogram was performed using the “rms” package.
- 5) DeLong test was performed using the “pROC” package.
- 6) Normality assessment was performed using the “nortest” package.
- 7) Decision curve analysis was performed using the “rmda” package.
- 8) Other statistical tests were performed using the “stats” package.