Am J Cancer Res 2024;14(12):5628-5643
www.ajcr.us /ISSN:2156-6976/ajcr0159533

Original Article

Construction and evaluation of a multifactorial clinical model for discriminating benign and malignant breast tumors using LASSO algorithm based on retrospective cohort study

Wenting Cui[1], Ying Wu[1], Yuewei Guo[1], Wei Li[1], Chen Huang[2], Yiqun Xie[1]

[1]Department of Breast Surgery, Huangpu Branch, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200011, China; [2]Department of General Surgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China

Received July 30, 2024; Accepted December 4, 2024; Epub December 15, 2024; Published December 30, 2024

Abstract: Breast cancer is one of the malignant tumors that seriously threaten women's health, and early diagnosis and detection of breast cancer are crucial for effective treatment. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is an important diagnostic tool that allows for the dynamic observation of blood flow characteristics of breast tumors, including small lesions within the affected tissue. Currently, it is widely used in clinical practice and has been shown promising prospects. This study included a total of 1,987 patients who underwent breast surgery at Huangpu Branch, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine from January 1, 2019 to December 31, 2019. Comprehensive patient information was collected, including ultrasound, mammography findings, physical examination details, age, family history, and pathological diagnoses. The least absolute shrinkage and selection operator (LASSO) algorithm was employed to assign values to the x variables, facilitating the construction and validation of the LASSO model group. Receiver operating characteristic curves were generated using support vector machines to determine the area under the curve (AUC), as well as to assess sensitivity and specificity. There were no statistically significant differences (P>0.05) in average age, body mass index, tumor location, or tumor benignity/malignancy between the training and test sets. The AUC, sensitivity, and specificity of mammography for predicting the benignity or malignancy of breast tumors were 0.83, 86.96%, and 76%, respectively. In comparison, the AUC, sensitivity, and specificity of DCE-MRI for the same predictions were 0.91, 91.3%, and 88%, respectively. The predictive performance of DCE-MRI was significantly higher than that of mammography (P<0.05). In conclusion, both mammography and DCE-MRI demonstrated high AUC, sensitivity, and specificity in predicting the benignity or malignancy of breast tumors. However, DCE-MRI showed superior predictive performance, making it a valuable tool for the early detection of clinical breast cancer with potential for broader clinical application.

Keywords: Breast tumor, DCE-MRI, mammography, LASSO algorithm, radiomics features

Introduction

Breast cancer originates in the epithelial cells lining the ducts or lobules of the breast's glandular tissue. Initially, cancer cells grow within the ducts or lobules (in situ cancer), typically without causing symptoms and with a low risk of metastasis. Over time, these in situ cancer cells may progress, invading the surrounding breast tissue (invasive breast cancer). They can then spread to nearby lymph nodes (regional metastasis) or other organs in the body (distant metastasis) [1-4]. Breast cancer is one of the most common malignancies in women, comprising 7-10% of all malignant tumors. It is more prevalent among women with a genetic predisposition and is most commonly diagnosed in women aged 40 to 60, particularly during the peri-menopausal period. The disease typically originates in the breast glandular tissue. While the vast majority of cases occur in women, male breast cancer accounts for 0.5-1% of all breast cancer diagnoses [5].

The exact cause of breast cancer remains unclear. The breast is a target organ for various

https://doi.org/10.62347/ILIJ7959

endocrine hormones, including estrogen, progesterone, and prolactin. Among these, estrone and estradiol levels have a direct association with breast cancer development. The condition is rare before the age of 20, but its incidence rises rapidly afterward. It peaks between the ages of 45 and 50 and continues to increase after menopause, possibly due to elevated estrogen levels in older individuals [6, 7]. The age at menarche, age at menopause, nulliparity, and age at first full-term pregnancy are all factors associated with the risk of developing breast cancer. Individuals with a first-degree relative who has a history of breast cancer face a 2-3 times higher risk of developing the disease compared to the general population [8-10]. While the relationship between benign breast diseases and breast cancer remains debated, most experts agree that epithelial hyperplasia or atypical hyperplasia in breast lobules may be linked to breast cancer development [11]. Additionally, excessive nutrition, obesity, and a high-fat diet can amplify or prolong estrogen stimulation of breast epithelial cells, thereby increasing breast cancer risk. Environmental factors and lifestyle choices are also associated with the incidence of breast cancer [12].

Currently, commonly used screening methods for breast cancer include breast examination, X-ray mammography, ultrasound, magnetic resonance imaging (MRI), positron emission tomography-computed tomography (PET-CT), and breast-specific gamma imaging [13-15]. Breast examination is a routine procedure to detect any abnormalities in the breasts. The optimal time for this examination is typically around 7-10 days after the end of menstruation, as estrogen has the least effect on the breasts during this period, and the breast tissue is relatively stable, making it easier to detect abnormalities or lesions. A physician's visual and tactile examination involves visually inspecting and palpating the breasts to assess their shape, the condition of the skin, the nipples and areolas, the presence of breast lumps, nipple discharge, and other abnormalities [16]. Breast ultrasound is another widely used diagnostic tool for breast diseases and is often used in conjunction with mammography. It can help differentiate between benign and malignant, cystic and solid, and hyperplastic breast conditions. Additionally, ultrasound can assist in evaluating suspicious lesions in the breast following breast implantation. Moreover, ultrasound is non-radiating, safe, painless, non-invasive, and can be repeated, making it suitable for women of any age who are suspected of having breast diseases. Modern medicine recognizes that breast hyperplasia is often linked to endocrine disorders and ovarian dysfunction. When estrogen levels increase and there is an imbalance between estrogen and progesterone, the breast tissue undergoes excessive proliferation or incomplete involution, which can lead to the development of breast lobular hyperplasia [17, 18].

Furthermore, research has shown that excessive estrogen stimulation of breast glandular epithelial cells is a major contributor to the development of breast cancer. Additionally, doctors can detect early-stage breast cancer, which may not be palpable during clinical examination, through X-rays. This method allows for the identification of tiny calcifications and lesions smaller than 0.1 millimeters. The diagnostic rate for early-stage breast cancer is higher, and the accuracy in distinguishing between benign and malignant tumors can exceed 90% [19-21]. The new generation of digital mammography has enhanced image clarity and reliability, while also reducing radiation exposure, making it safer for the body. PET-CT is a comprehensive imaging technique that can assess the overall condition of patients diagnosed with breast cancer and identify tiny metastatic lesions. Breast-specific gamma imaging (BSGI) is an innovative diagnostic method in molecular imaging, offering advantages in the qualitative diagnosis of breast cancer. The procedure involves injecting the patient with 99Tc-MIBI (Technetium-99m methoxyisobutylisonitrile), followed by imaging with a gamma camera. The malignancy of the tumor is determined based on the concentration of gamma rays within the tumor [22]. BSGI can detect larger lesions, multiple or contralateral lesions, and hidden tumors in the breast [23]. These diagnostic methods each have their own strengths and limitations, and doctors can select a combination of approaches tailored to the patient's specific condition to provide a comprehensive diagnosis and guide subsequent treatment.

Early diagnosis and detection of breast cancer are crucial and play a significant role in the

treatment. Therefore, we aimed to construct a prediction model that includes a total of 1,987 patients who underwent breast surgery at the Huangpu Branch of Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine in 2019. The dataset includes complete records of ultrasound, mammography, physical examination, age, family history, and pathological diagnosis. The least absolute shrinkage and selection operator (LASSO) algorithm was used to assign values to the x variables, construct the LASSO model group, and perform validation and fitting. Column line charts were generated to thoroughly assess the predictive efficacy of the LASSO multifactor clinical model, based on different radiomics, in determining the malignancy of breast tumors.

Materials and methods

Subject selection

A total of 1,987 patients who underwent breast surgery in the Department of Breast Surgery at the Huangpu Branch of Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, throughout 2019 were included in this study. All patients had complete records of ultrasound, mammography, physical examination, age, family history, and pathological diagnosis. Among these patients, 718 cases were diagnosed as malignant. The patients' ages ranged from 17 to 89 years, with a median age of 50 years and an average age of 51 years. The study was approved by the Institutional Review Board and Research Ethics Committee of the Huangpu Branch, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, and was conducted in accordance with the principles of the Declaration of Helsinki.

Inclusion criteria: 1) The age of the patients ranged from 17 to 89 years old; 2) Patients had complete records of ultrasound, mammography, physical examination, age, family history, and pathological diagnosis; 3) Patients were required to have HER2+ breast cancer (3+ on immunohistochemistry and/or amplified on fluorescence in situ hybridization). Exclusion criteria: 1) Patients had breast implants or were pregnant, lactating, or planning to become pregnant within half a year; 2) Biopsy conducted at an external institution and unavailable pathologic results; 3) Previous or simultaneous presence of other tumors; 4) Patients with severe heart, kidney or nervous system complications, malnutrition, diabetes, or Karnofsky index less than 70.

Mammography with molybdenum target X-ray

Mammography was performed using a fully digital mammography system, the GE digital flat-panel mammography machine. The acquisition parameters included a detector size of 25×29 cm, an imaging matrix of 3530×4200, a spatial resolution of 7.25 LP/mm, and a standard pixel size of 60 μm. Images of the breasts in craniocaudal, mediolateral, and oblique views were obtained.

MRI imaging method

Routine T2-weighted imaging (T2WI) and dynamic contrast-enhanced MRI (DCE-MRI) sequences were performed using a Philips Ingenia 3.0T fully digital MRI system. A bolus injection of 0.3 mmol/kg contrast agent was administered through the elbow vein, with an echo time of 2.5 ms, a repetition time of 3.5 ms, an inversion angle of 10°C, a matrix size of 562×562, and a slice thickness of 1.5 mm. The acquired images were transferred to a workstation for further processing.

Image segmentation

Three senior physicians independently performed image segmentation on the selected images. Each physician was unaware of the clinical pathological diagnosis of the patients. The 3D Slicer software was used to cover the entire tumor area with the selected region of interest. The intensity of the tumor boundary region was chosen as the threshold for semi-automatic threshold segmentation. The segmented images retained only the region of interest, which represents the tumor area and serves as input for extracting quantitative radiomics features.

MATLAB R2016b software was used to extract four groups of image features from the region of interest. These features include first-order gray-level histogram features (energy), texture features (area, perimeter, compactness, elongation ratio), shape features, and wavelet transform features. Among them, energy, entropy, kurtosis, and homogeneity can be represented as follows:

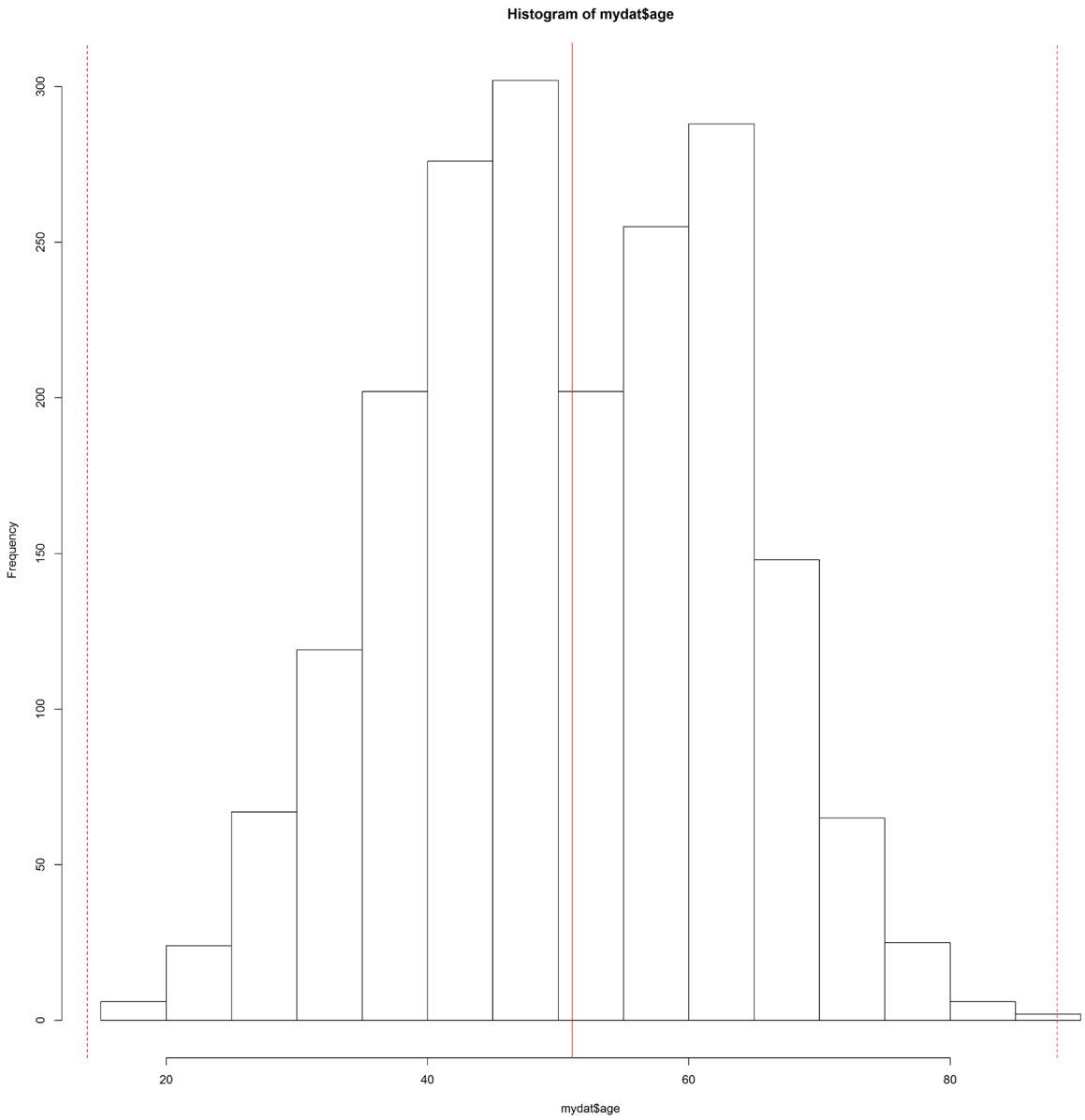

Figure 1. Histogram of age distribution.

$$energy = \sum_{i}^{m} Ki^{2} \quad (1)$$

$$entropy = \sum_{i=1}^{m} I(i)\log_{2} I(i) \quad (2)$$

$$kurtosis = \frac{1}{M}\sum_{i=1}^{m} K(i) - \overline{K}^{4} \times \sqrt{\frac{1}{M}\sum_{i=1}^{m} K(i) - \overline{K}^{2}}^{-2} \quad (3)$$

$$homogeneity = \sum_{i=1}^{m} I(i)^{2} \quad (4)$$

Where, m represents the total number of pixels, K represents the two-dimensional pixel data, I represents the first-order gray-level value, and M represents the number of gray levels.

Then, the maximum-minimum normalization method [24] was applied to normalize all the obtained features. The normalization formula is as follows:

$$c = \frac{y_i - min(y)}{max(y) - min(y)} \quad (5)$$

Where, c represents the normalized data, y represents the original data, max(y) represents the maximum value among the original data, and

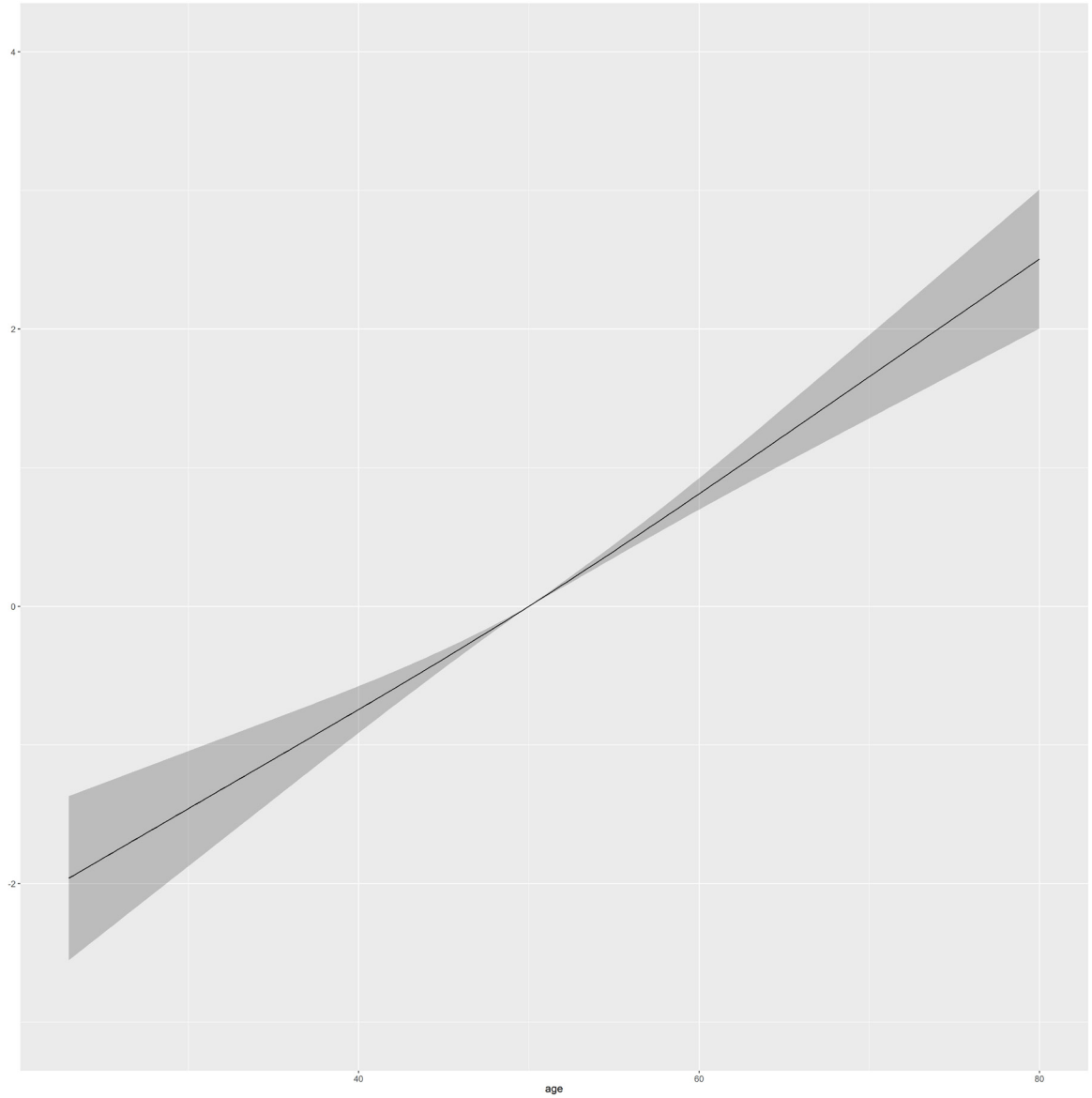

Figure 2. Spline function of diagnostic results and age.

min(y) represents the minimum value among the original data.

Preprocessing of predictive factors

To facilitate the application of the final model, we included ultrasound (BUS), mammography, palpation, and family history as binary variables in the model. BUS and mammography were categorized as 0-4a and labeled as 0, while palpation was categorized as no mass or possibly benign mass and labeled as 0. No family history of breast or ovarian cancer was labeled as 0, while the presence of family history was labeled as 1.

Since age is a continuous variable, we plotted a histogram to show its approximate distribution, and no outliers were observed (**Figure 1**). We also plotted a graph by fitting the diagnostic results with age using a spline function (**Figure 2**), and a monotonic relationship between age and diagnostic results was determined.

Age was included as a continuous variable in the model, labeled as (age1). Additionally, based on clinical experience, mammography has a higher diagnostic value in individuals aged 40 and above. Therefore, the continuous variable age was transformed into a binary variable based on <40 years old and ≥40 years old,

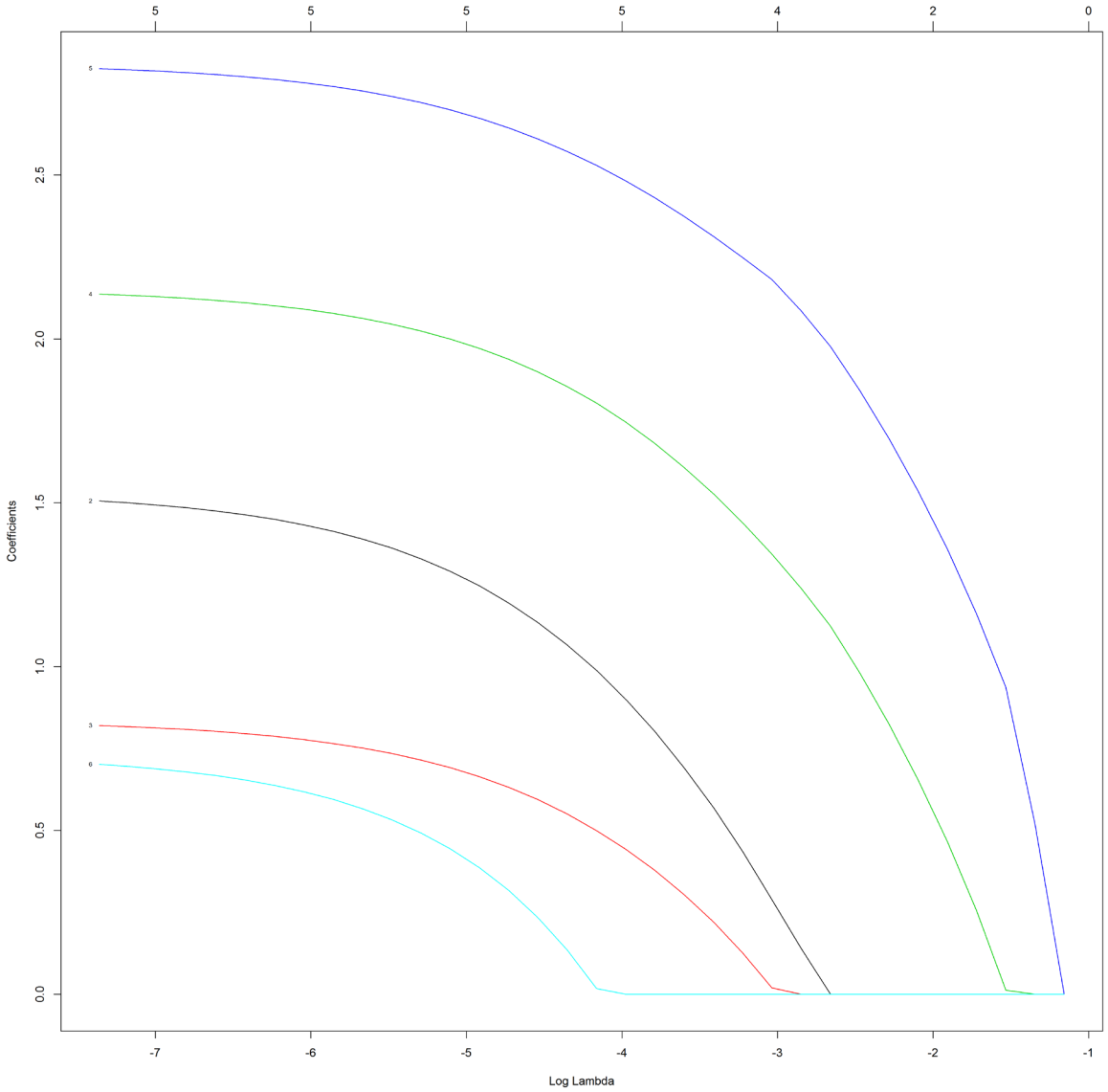

Figure 3. LASSO model set.

labeled as (age2). Patients under 40 years old were labeled as 0 (375 cases), while patients aged 40 and above were labeled as 1 (1612 cases).

LASSO variable selection and development of the nomogram

The LASSO algorithm is a regression analysis method that performs both variable selection and regularization using R software (3.8.0) with glmnet package. In the current study, the LASSO algorithm was used to select the most significant features. The LASSO model set was established, as shown in **Figure 3**. Through cross-validation, the optimal λ (lambda) value was determined as 0.0006364417. If the variable "family history", with the least impact, is removed, the optimal λ value becomes 0.01875626. The optimal and simplest acceptable LASSO model was fitted, including age 2, palpation, mammography, BUS, and family history. Afterward, the nomogram was constructed using the independent predictors selected by LASSO to generate a model for breast cancer prediction.

Validation of the nomogram

Our nomogram was internally validated by using 100 bootstrap samples. Additionally, three main aspects of validation of the nomo-

Table 1. Baseline characteristics of participants

Parameters	Tumor-negative group (n=795)	Tumor-positive group (n=1,192)	t/χ^2	P
Age (years)	51.32±1.27	51.19±1.13	2.235	0.026
Body Mass Index (kg/m^2)	22.75±1.92	22.74±1.31	0.194	0.846
Education Level [n/(%)]			0.014	1.000
Illiterate	2 (0.25%)	3 (0.25%)		
Primary school	116 (14.59%)	176 (14.77%)		
Secondary school	471 (59.25%)	706 (59.23%)		
University degree	206 (25.91%)	307 (25.76%)		
Smoking history [n (%)]	276 (34.72%)	415 (34.82%)	0.002	0.964
Drinking history [n (%)]	431 (54.21%)	650 (54.53%)	0.019	0.890
Employment [n (%)]	576 (72.45%)	875 (73.41%)	0.220	0.639
Family History [n (%)]	101 (12.7%)	392 (32.89%)	104.124	<0.001
Largest tumor size (cm)	2.28±0.61	2.29±0.34	0.215	0.830
Number of tumors [n (%)]			0.002	0.964
1	531 (66.79%)	795 (66.69%)		
>1	264 (33.21%)	397 (33.31%)		

Table 2. Optimal parameters of the model

Factor	Coefficient	HR (95% CI)	*p*-value
Age2	1.53	4.63 (3.02, 7.09)	<0.0001
Palpation	0.84	2.31 (1.73, 3.08)	<0.0001
Mammography	2.16	8.62 (6.07, 12.25)	<0.0001
BUS	2.84	17.14 (12.17, 24.15)	<0.0001
Family History	0.73	2.08 (1.06, 4.09)	0.03

Note: BUS, breast ultrasound.

gram were discrimination, calibration, and clinical usefulness. The analysis of the area under the receiver operating characteristic (ROC) curve (AUC) was conducted to assess the discriminative performance of the nomogram. Further, a calibration curve was also plotted to evaluate the deviation between the estimated and actual probability. Lastly, decision curve analysis (DCA) was performed to demonstrate the clinical usefulness of the nomogram by quantifying the net benefit at different probability thresholds.

Statistical methods

Data processing in this study was performed using SPSS 19.0 statistical software. Continuous variables are presented as mean ± standard deviation and analyzed using t test, and categorical variables are presented as percentages (%) and analyzed using chi-square test. Two-tailed tests with P<0.05 were considered statistically significant. Column chart was made by ggplot package in R software (3.8.0).

Results

LASSO variable selection and Nomogram development

Variable selection and nomogram development: We first used the LASSO variable selection method to develop the Logistic proportional hazard model. Among the multiparametric features (**Table 1**), five predictors were associated with breast malignant tumors according to the LASSO regression. The candidate variables were age2, palpation, mammography, BUS, and family history. The factors, beta coefficients, hazard ratios for factors, and *P* values of Z tests are provided in **Table 2**.

The results showed that age (>40 or not), palpation, mammography, BUS, and family history of breast cancer all exhibited independent diagnostic value for distinguishing between benign and malignant breast diseases. Among them, BUS has the highest diagnostic value, while family history has the lowest. Patients with BUS indicating 4B or above showed a 16.14 times higher possibility of having breast cancer. Patients with mammography indicating 4B or above manifested a 7.62 times higher possibility of having breast cancer. Patients aged 40 and above exhibited a 3.63 times higher possibility of having breast cancer. Patients with positive palpation findings had a 1.31 times higher possibility of having breast cancer. Patients with a family history of breast cancer

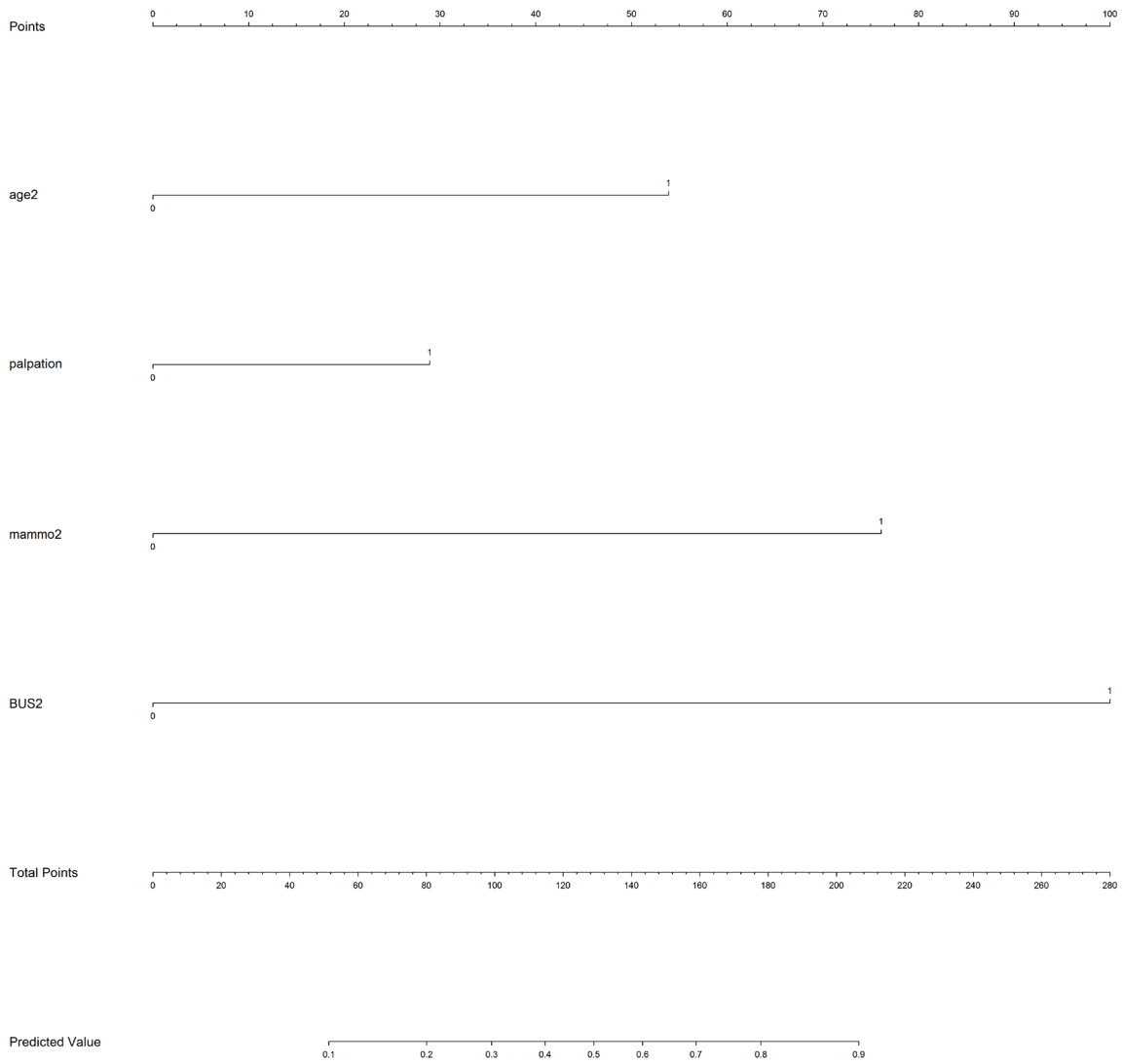

Figure 4. Age, palpation, mammography, BUS nomogram for predicting benign and malignant breast diseases. Note: BUS, breast ultrasound.

showed a 1.08 times higher possibility of having breast cancer. It can be seen that BUS has the highest value in the preoperative diagnosis of breast cancer, followed by mammography, age, palpation, and family history.

Afterward, the nomogram was developed based on the above five risk factors of breast cancer (**Figure 4**). **Figure 4** depicts the model column chart created in this study, revealing that BUS2 contributes the most to breast tumors and time, followed by mammography and age2, while family history contributes the least. The web presentation of the nomogram is shown at https://jyhprx.shinyapps.io/a_demo_of_cpm/.

Model discrimination and calibration: The nomogram was internally validated by 100 bootstrap samples to decrease the overfitting bias (C index =0.89). Therefore, our model showed good discrimination. Calibration curves were generated to visualize the discrimination between the predicted and actual probabilities (**Figure 5**). These results indicated that our extended Cox prognostic model fit well.

Distribution of feature labels

The feature coefficients and compensation parameters obtained using the LASSO algorithm to construct corresponding radiomics labels are shown in **Figure 6**. The x-axis repre-

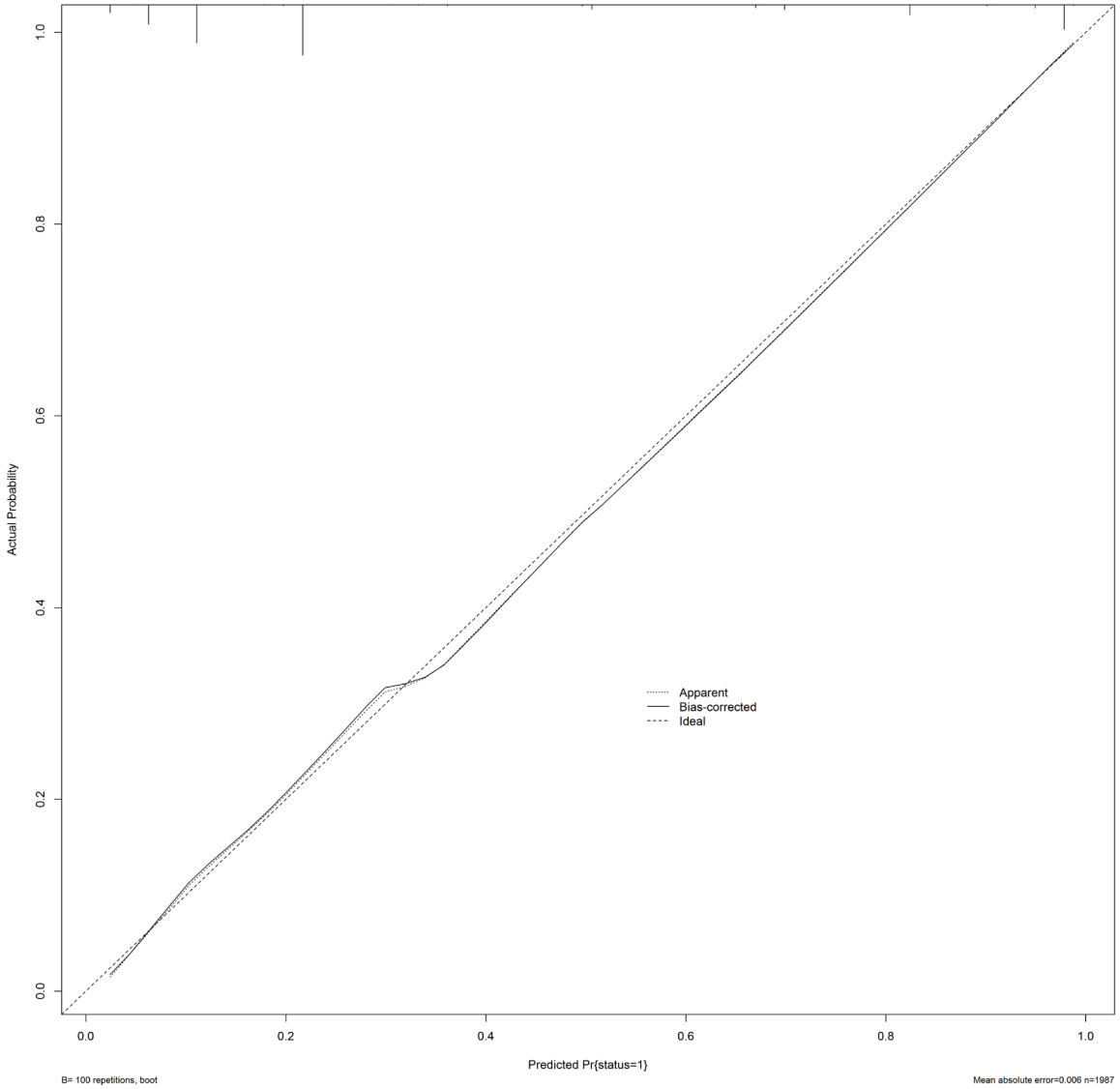

Figure 5. Enhanced bootstrap internal validation, with N=100 and C index =0.89.

sents patient numbers, while the y-axis represents the values of the radiomics feature labels. Red indicates tumor-negative cases, while green indicates tumor-positive cases.

Predictive performance of DCE-MRI and mammography models

As shown in **Figure 7**, the predictive model constructed using the training set data was applied to the actual testing set data. The AUC for the mammography model in predicting breast tumor malignancy was 0.83, with a sensitivity of 86.96% and specificity of 76%. The AUC area for the DCE-MRI model in predicting breast tumor malignancy was 0.91, with a sensitivity of 91.3% and specificity of 88%. Comparative analysis revealed that the DCE-MRI model exhibited significantly higher AUC, sensitivity, and specificity in predicting breast tumor malignancy compared to the mammography model, with statistical significance (P<0.05), as shown in **Figure 8**.

Case presentation with imaging

As shown in **Figure 9**, a 54-year-old female patient was incidentally found to have a lump in the breast. The mammographic image revealed a high-density mass with a diameter of approximately 26.3 mm, irregular margins, and calcifications in the lower central region of the left breast. Localized skin retraction was observed.

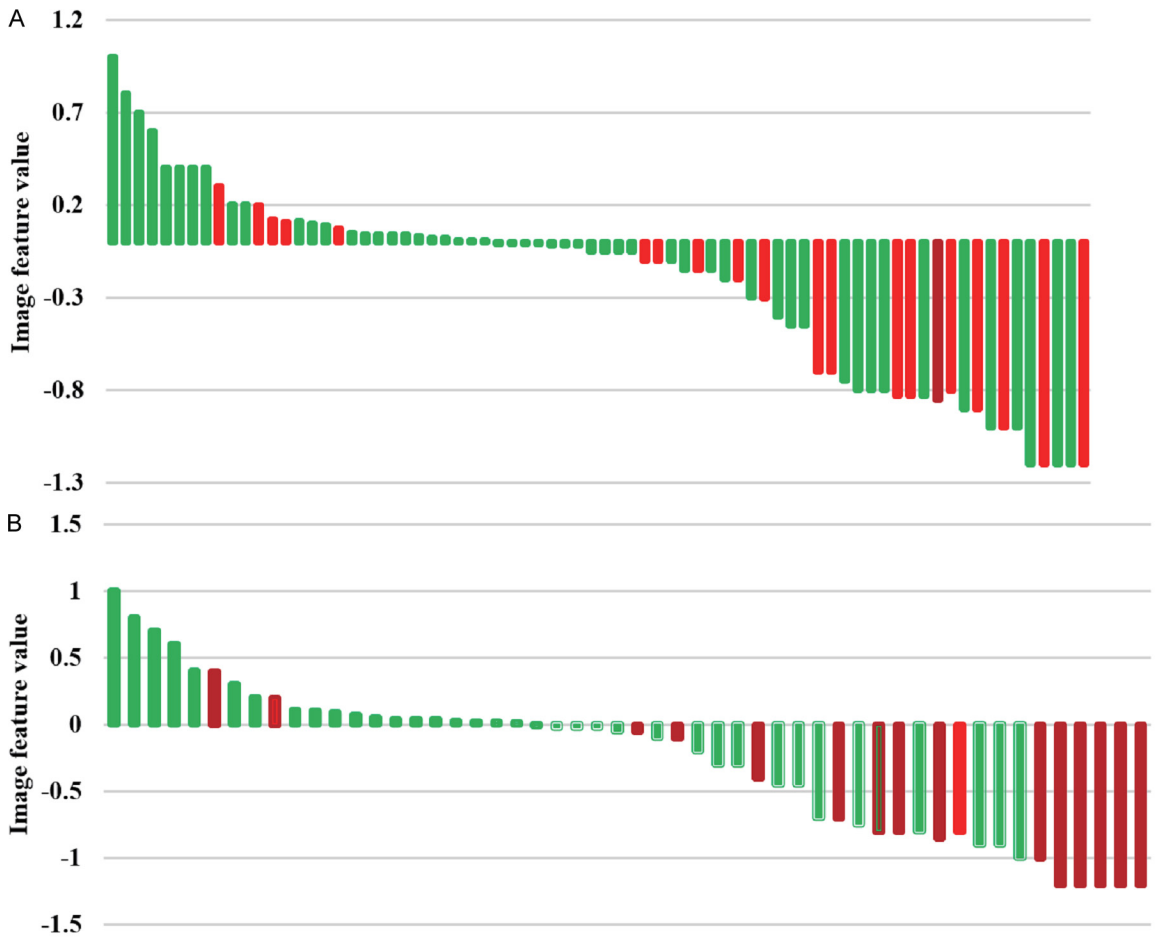

Figure 6. Distribution of feature labels in the training and testing sets (A represents the training set, and B represents the testing set).

Calcifications were also detected in both breasts.

As shown in **Figure 10**, the 54-year-old female patient exhibited an abnormal signal mass in the posterior areola of the left breast, measuring approximately 2.5 cm×1.8 cm. The mass had an irregular shape with lobulated margins and spiculated shadows. It appeared as a low signal on STIR (short-tau inversion recovery) and a high signal on DWI (diffusion weighted imaging), with an ADC (apparent diffusion coefficient) value of 0.74×10^{-3} mm^2/s. Abnormal signal nodules and strands were observed around the mass, connected to the areola of the left breast. Thickening and retraction of the left areolar skin were also noted. The enhanced images showed heterogeneous enhancement within the lesion, rapid early enhancement, and a plateau-like pattern in the delayed phase. The MIP image displayed significant thickening of the surrounding blood vessels.

Prediction model formula

Using LASSO regression analysis, five independent predictive factors associated with breast malignant tumors were selected: age2 (>40 years old or not), palpation, mammography, BUS, and family history. The model formula is: Logit (P) = β_0 + β_1·Age2 + β_2·Palpation + β_3·Mammography + β_4·BUS + β_5·Family History, where Logit (P) represents the logarithmic probability of breast tumor malignancy; β_0 is the intercept term; β_1, β_2, β_3, β_4, and β_5 are the regression coefficients of each predictive factor, respectively; Age2 (binary variable: >40 years old is 1, otherwise it is 0); Palpation (binary variable: positive palpation is 1, otherwise 0); Mammography (binary variable: molybdenum target photography anomaly is 1, otherwise 0); BUS (binary variable: BUS exception is 1, otherwise 0); Family History (binary variable: family history of breast cancer is 1, otherwise 0). The

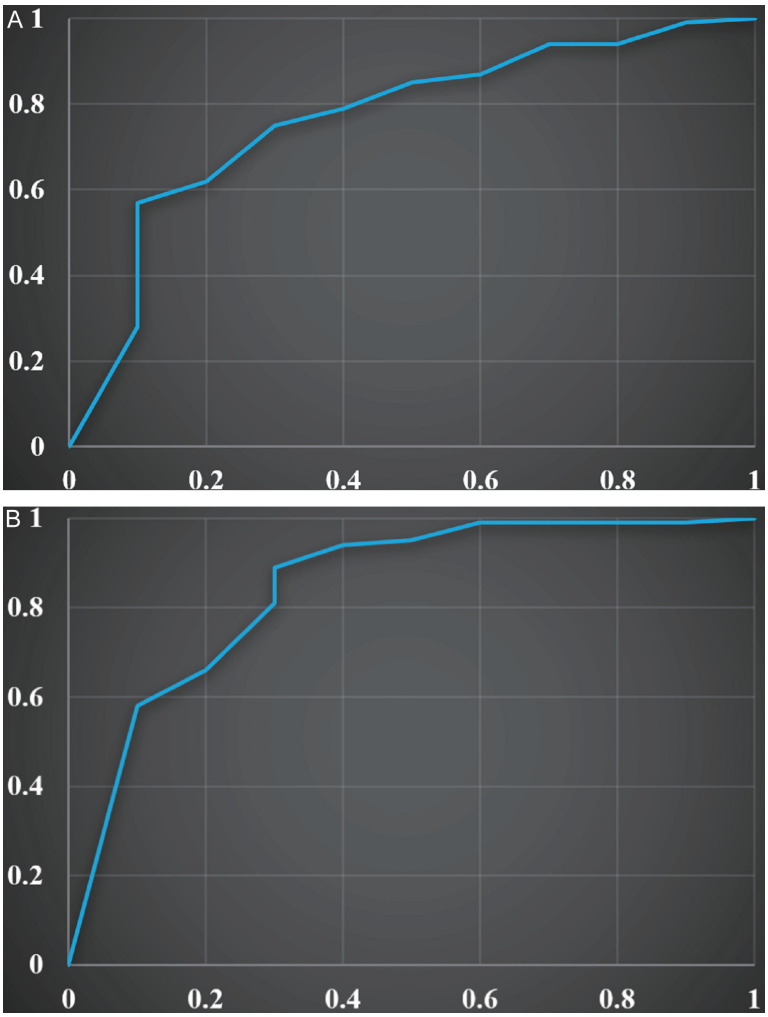

Figure 7. Predictive performance of the DCE-MRI and mammography models for breast tumor malignancy. A: Mammography model; B: DCE-MRI model.

specific intercept term β_0 needs to be calculated based on the training dataset in practical applications. The following is the specific model formula: Logit (P) = β_0 + 1.31·Palpation + 3.63·Age2 + 7.62·Mammography + 16.14·BUS + 1.08·Family History.

Discussion

Breast cancer is one of the malignant tumors that seriously threatens women's health. According to the latest national cancer report released by the National Cancer Center, the incidence rate of breast cancer in China still ranks first among female malignant tumors, reaching 30.6 per 100,000, with approximately 306,000 new cases each year [25, 26]. Breast cancer screening aims to detect, diagnose, and treat the disease early through effective, simple, and cost-effective approaches, ultimately reducing the mortality rate of breast cancer. B-ultrasound is widely used in the examination of breast cancer, capable of detecting tumors larger than 1 cm and distinguishing between cystic and solid masses with high accuracy [27]. Ultrasound examination is economical, safe, non-invasive, and allows for repeated examinations. It can also guide the puncture examination of breast lesions. Mammography, using X-ray imaging, is another a commonly used method for breast cancer screening and follow-up. Tumors appear as high-density, irregular, spiculated, or star-like changes on mammograms, but its effectiveness is limited in dense breasts [28, 29]. Breast MRI has a strong ability to detect lesions, including tiny ones, and to differentiate between benign and malignant tumors. However, MRI has limited qualitative diagnostic capabilities for lesions. Different imaging techniques have their own advantages and disadvantages, so it is crucial to choose the appropriate examination method for the diagnosis of breast cancer [30].

Emanuelli [31] found that molybdenum target photography had lower sensitivity in dense mammary glands, and our study confirms the advantage of DCE-MRI in this regard.

DCE-MRI can provide more detailed hemodynamic information, which helps identify small lesions and distinguish between benign and malignant tumors. Compared with traditional mammography, DCE-MRI can offer more dynamic enhanced features, which are of great significance for the early diagnosis of breast cancer [32]. In this study, a total of 1,987 patients who underwent breast surgery at the Huangpu Branch of Shanghai Ninth People's

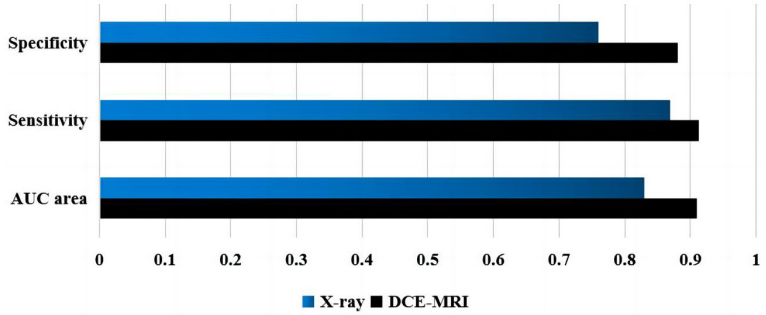

Figure 8. Comparison of the AUC, sensitivity, and specificity of the DCE-MRI and mammography models in predicting breast tumor malignancy. *, statistically significant differences (P<0.05) between the DCE-MRI and mammography models; AUC, area under the curve; X-ray, X-radiation; DCE-MRI, dynamic contrast enhanced magnetic resonance imaging.

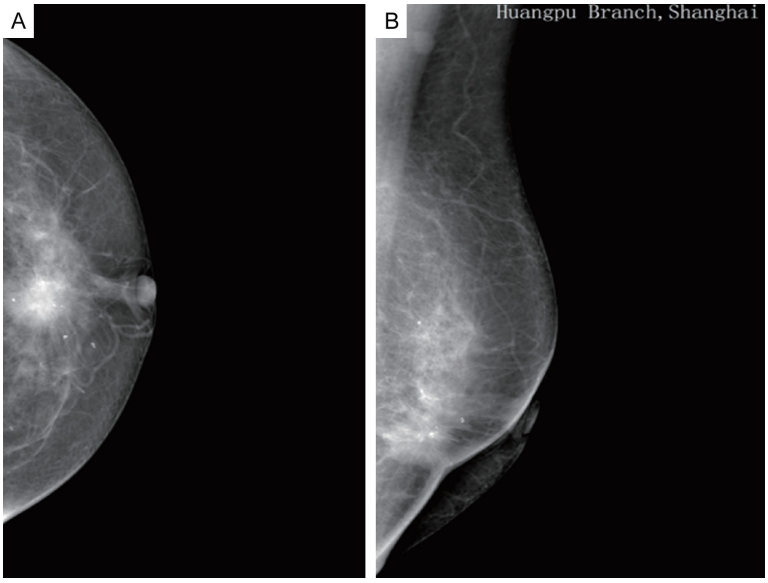

Figure 9. The mammographic image data of a 54-year-old patient. A: LCC, left craniocaudal; B: LMLO, left mediolateral oblique.

forms of calcifications. It has the advantages of being relatively painless, easy to perform, high resolution, and good repeatability [33]. DCE-MRI is also an important examination method. By using continuous, repetitive, and rapid imaging techniques, it can obtain images before and after contrast agent injection. Through a series of calculations and analyses, semi-quantitative or quantitative parameters can be obtained. DCE-MRI allows for dynamic observation of the overall blood flow characteristics of breast tumors at a micro level. It is currently widely used in clinical practice and has good prospects for future use [34]. From the imaging results, DCE-MRI showed a mass with spiculation on the outer side of the right breast and an occupying lesion behind the nipple. Mammography showed a mass above the right breast with lobulation, spiculation, punctate calcifications, and density behind the nipple. This indicates that both DCE-MRI and mammography can clearly display the characteristics of breast tumor lesions, assisting physicians in evaluating the patient's condition.

Hospital, Shanghai Jiaotong University School of Medicine in 2019 were included. The LASSO algorithm was used to assign values to the x variables and construct the LASSO model group. The model was validated, fitted, and a column chart was plotted. The support vector machine was used to draw the ROC curve, and AUC, sensitivity, and specificity were calculated.

Mammography, using X-ray imaging, is one of the most commonly used examination methods for clinical breast tumor patients. It can clearly display the tissue architecture and plays an important role in distinguishing different

The LASSO algorithm is a linear regression method that uses L1 regularization to assign feature weights. The L1 regularization can make some learned feature weights equal to zero, achieving sparsity and feature selection [35]. In this study, the feature coefficients and compensation parameters obtained by the LASSO algorithm were used to construct corresponding radiomics labels. From the generated chart, the x-axis represents the patient's ID, the y-axis represents the radiomics feature label value, red represents tumor-negative, and green represents tumor-positive. The breast tumor positivity/negativity of patients in the training and testing sets can be clearly dis-

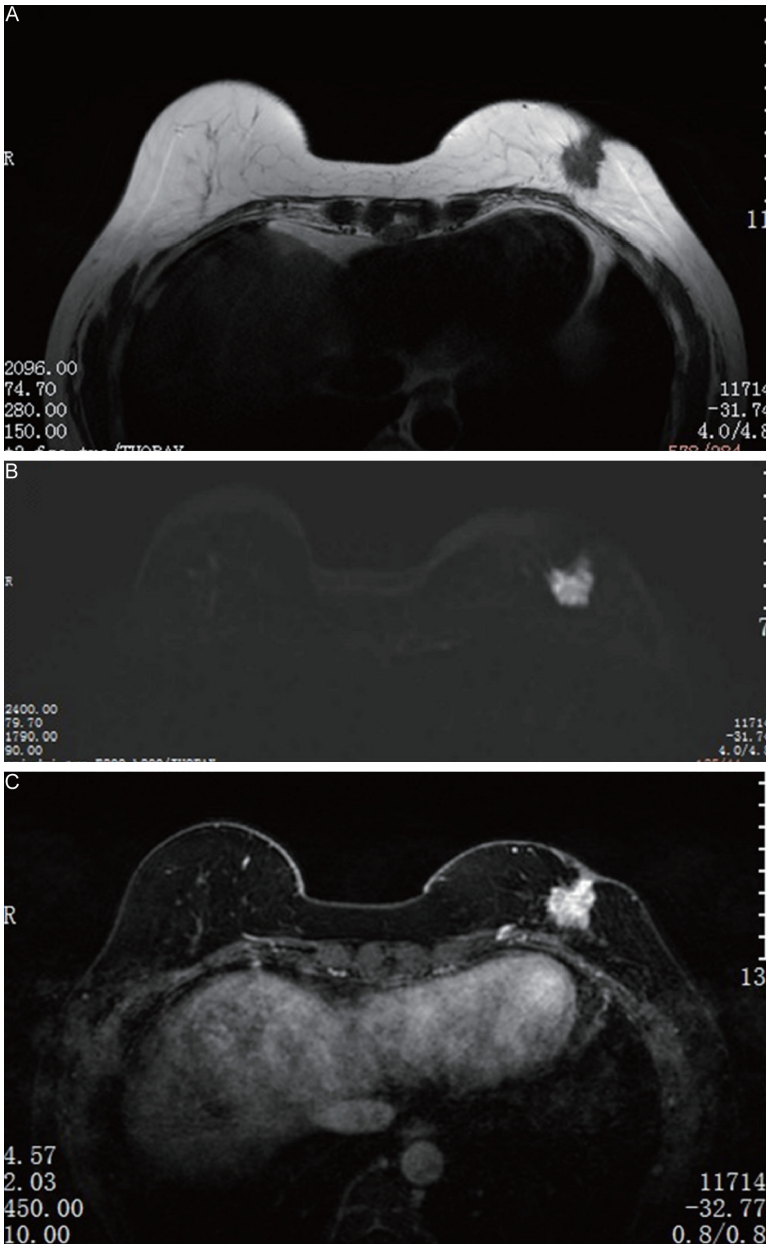

Figure 10. DCE-MRI image of a 54-year-old patient. A: STIR, short-tau inversion recovery; B: DWI, diffusion weighted imagine; C: DCE-MRI, dynamic contrast enhanced magnetic resonance imaging.

played. According to ROC curve analysis, the AUC for the prediction of the benign and malignant nature of breast tumors by the mammography model was 0.83, with a sensitivity of 86.96% and specificity of 76%. The AUC for the prediction by the DCE-MRI model was 0.91, with a sensitivity of 91.3% and specificity of 88%. This indicates that both the mammography and DCE-MRI models have high predictive performance for distinguishing benign and malignant breast tumors and have certain clinical application value [36].

Comparing the two models, the DCE-MRI model shows significantly higher AUC, sensitivity, and specificity for the prediction of breast tumor malignancy compared to the mammography model, and the differences are statistically significant ($P<0.05$). This indicates that DCE-MRI has better predictive performance for distinguishing benign and malignant breast tumors, which is helpful for the early detection of breast cancer.

This study included a total of 1987 patients who underwent breast surgery at the Huangpu Branch, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine in 2019. We included complete information, including B-ultrasound, mammography, physical examination, age, family history, and pathological diagnosis. By constructing LASSO models, it was found that both the mammography and DCE-MRI models demonstrated high AUC, sensitivity, and specificity for predicting breast tumor malignancy, with the DCE-MRI model showing better predictive performance. This has clinical value for the early detection and treatment of breast cancer. However, this study has some limitations. Due to differences in sample size, types of included patients, and physicians' diagnostic levels, we only analyzed the benign and malignant nature of tumors and did not address more in-depth aspects, such as lymph node metastasis. In future research, we will include more breast tumor cases and further explore the application value of different radiomics LASSO models in disease diagnosis. In summary, the results of this study pro-

vide a theoretical reference for the early assessment of breast tumor malignancy.

Acknowledgements

This study was supported by project of Shanghai Huangpu District Science and Technology Commission (Nos. HLM202212, HLM202214), and project of Science and Technology Commission of Shanghai Municipality (STCSM) (No. 14411972400).

Disclosure of conflict of interest

None.

Address correspondence to: Yiqun Xie, Department of Breast Surgery, Huangpu Branch, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, No. 58 Puyu East Road, Shanghai 200011, China. E-mail: duolakuangzao@126.com

References

[1] Li Y, Li M, Su K, Zong S, Zhang H and Xiong L. Pre-metastatic niche: from revealing the molecular and cellular mechanisms to the clinical applications in breast cancer metastasis. Theranostics 2023; 13: 2301-2318.

[2] Jiao Y and Lv Q. Does primary tumor resection induce accelerated metastasis in breast cancer? A review. J Surg Res 2023; 283: 1005-1017.

[3] Guo L, Kong D, Liu J, Zhan L, Luo L, Zheng W, Zheng Q, Chen C and Sun S. Breast cancer heterogeneity and its implication in personalized precision therapy. Exp Hematol Oncol 2023; 12: 3.

[4] Guo L, Kong D, Liu J, Zhan L, Luo L, Zheng W, Zheng Q, Chen C and Sun S. Correction: breast cancer heterogeneity and its implication in personalized precision therapy. Exp Hematol Oncol 2024; 13: 7.

[5] Niccolai E, Baldi S, Nannini G, Gensini F, Papi L, Vezzosi V, Bianchi S, Orzalesi L, Ramazzotti M and Amedei A. Breast cancer: the first comparative evaluation of oncobiome composition between males and females. Biol Sex Differ 2023; 14: 37.

[6] Zhang YN, Xia KR, Li CY, Wei BL and Zhang B. Review of breast cancer pathologigcal image processing. Biomed Res Int 2021; 2021: 1994764.

[7] Li Z, Wei H, Li S, Wu P and Mao X. The role of progesterone receptors in breast cancer. Drug Des Devel Ther 2022; 16: 305-314.

[8] Arafat HM, Omar J, Muhamad R, Al-Astani TAD, Shafii N, Al Laham NA, Naser I, Shamallakh OM, Shamallakh KM and Jebril MAAR. Breast cancer risk from modifiable and non-modifiable risk factors among Palestinian women: a systematic review and meta-analysis. Asian Pac J Cancer Prev 2021; 22: 1987-1995.

[9] Wang H, MacInnis RJ and Li S. Family history and breast cancer risk for Asian women: a systematic review and meta-analysis. BMC Med 2023; 21: 239.

[10] Lu N, Zhang C, You H, Ma Z, Zhu P and Cheng F. Factors affecting breast screening behavior of first-degree relatives of breast cancer patients in China: a cross-sectional study. Cancer Nurs 2024; 47: 271-280.

[11] Turkyilmaz Z, Sarisik E, Ozkurt E, Tukenmez M, Emiroglu S, Emiroglu B, Onder S, Yilmaz R, Muslumanoglu M, Igci A, Ozmen V and Cabioglu N. Evaluation of benign breast diseases with or without atypical epithelial hyperplasia accompanying radial scars. Eur J Breast Health 2023; 19: 166-171.

[12] Desai P and Aggarwal A. Breast cancer in women over 65 years - a review of screening and treatment options. Clin Geriatr Med 2021; 37: 611-623.

[13] Jokar N, Velikyan I, Ahmadzadehfar H, Rekabpour SJ, Jafari E, Ting HH, Biersack HJ and Assadi M. Theranostic approach in breast cancer: a treasured tailor for future oncology. Clin Nucl Med 2021; 46: e410-e420.

[14] Nigdelis MP, Karamouzis MV, Kontos M, Alexandrou A, Goulis DG and Lambrinoudaki I. Updates on the treatment of invasive breast cancer: Quo Vadimus? Maturitas 2021; 145: 64-72.

[15] Zhang J, Hou S, You Z, Li G, Xu S, Li X, Zhang X, Lei B and Pang D. Expression and prognostic values of ARID family members in breast cancer. Aging (Albany NY) 2021; 13: 5621-5637.

[16] Wekking D, Porcu M, De Silva P, Saba L, Scartozzi M and Solinas C. Breast MRI: clinical indications, recommendations, and future applications in breast cancer diagnosis. Curr Oncol Rep 2023; 25: 257-267.

[17] Mosconi L, Jett S, Nerattini M, Andy C, Yepez CB, Zarate C, Carlton C, Kodancha V, Schelbaum E, Williams S, Pahlajani S, Loeb-Zeitlin S, Havryliuk Y, Andrews R, Pupi A, Ballon D, Kelly J, Osborne J, Nehmeh S, Fink M, Berti V, Matthews D, Dyke J and Brinton RD. In vivo brain estrogen receptor expression by neuroendocrine aging and relationships with gray matter volume, bio-energetics, and clinical symptomatology. Res Sq [Preprint] 2023; rs.3.rs-2573335.

[18] Al-Shami K, Awadi S, Khamees A, Alsheikh AM, Al-Sharif S, Ala' Bereshy R, Al-Eitan SF, Banikhaled SH, Al-Qudimat AR, Al-Zoubi RM and Al Zoubi MS. Estrogens and the risk of breast

cancer: a narrative review of literature. Heliyon 2023; 9: e20224.

[19] Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen TH, Wang Q, Bolla MK, Yang X, Adank MA, Ahearn T, Aittomäki K, Allen J, Andrulis IL, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Auer PL, Auvinen P, Barrdahl M, Beane Freeman LE, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Børresen-Dale AL, Brauch H, Bremer M, Brenner H, Brentnall A, Brock IW, Brooks-Wilson A, Brucker SY, Brüning T, Burwinkel B, Campa D, Carter BD, Castelao JE, Chanock SJ, Chlebowski R, Christiansen H, Clarke CL, Collée JM, Cordina-Duverger E, Cornelissen S, Couch FJ, Cox A, Cross SS, Czene K, Daly MB, Devilee P, Dörk T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Engel C, Eriksson M, Evans DG, Fasching PA, Figueroa J, Fletcher O, Flyger H, Försti A, Fritschi L, Gabrielson M, Gago-Dominguez M, Gapstur SM, García-Sáenz JA, Gaudet MM, Georgoulias V, Giles GG, Gilyazova IR, Glendon G, Goldberg MS, Goldgar DE, González-Neira A, Grenaker Alnæs GI, Grip M, Gronwald J, Grundy A, Guénel P, Haeberle L, Hahnen E, Haiman CA, Håkansson N, Hamann U, Hankinson SE, Harkness EF, Hart SN, He W, Hein A, Heyworth J, Hillemanns P, Hollestelle A, Hooning MJ, Hoover RN, Hopper JL, Howell A, Huang G, Humphreys K, Hunter DJ, Jakimovska M, Jakubowska A, Janni W, John EM, Johnson N, Jones ME, Jukkola-Vuorinen A, Jung A, Kaaks R, Kaczmarek K, Kataja V, Keeman R, Kerin MJ, Khusnutdinova E, Kiiski JI, Knight JA, Ko YD, Kosma VM, Koutros S, Kristensen VN, Krüger U, Kühl T, Lambrechts D, Le Marchand L, Lee E, Lejbkowicz F, Lilyquist J, Lindblom A, Lindström S, Lissowska J, Lo WY, Loibl S, Long J, Lubiński J, Lux MP, MacInnis RJ, Maishman T, Makalic E, Maleva Kostovska I, Mannermaa A, Manoukian S, Margolin S, Martens JWM, Martinez ME, Mavroudis D, McLean C, Meindl A, Menon U, Middha P, Miller N, Moreno F, Mulligan AM, Mulot C, Muñoz-Garzon VM, Neuhausen SL, Nevanlinna H, Neven P, Newman WG, Nielsen SF, Nordestgaard BG, Norman A, Offit K, Olson JE, Olsson H, Orr N, Pankratz VS, Park-Simon TW, Perez JIA, Pérez-Barrios C, Peterlongo P, Peto J, Pinchev M, Plaseska-Karanfilska D, Polley EC, Prentice R, Presneau N, Prokofyeva D, Purrington K, Pylkäs K, Rack B, Radice P, Rau-Murthy R, Rennert G, Rennert HS, Rhenius V, Robson M, Romero A, Ruddy KJ, Ruebner M, Saloustros E, Sandler DP, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schürmann P, Schwentner L, Scott C, Scott RJ, Seynaeve C, Shah M, Sherman ME, Shrubsole MJ, Shu XO, Slager S, Smeets A, Sohn C, Soucy P, Southey MC, Spinelli JJ, Stegmaier C, Stone J, Swerdlow AJ, Tamimi RM, Tapper WJ, Taylor JA, Terry MB, Thöne K, Tollenaar R, Tomlinson I, Truong T, Tzardi M, Ulmer HU, Untch M, Vachon CM, van Veen EM, Vijai J, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Yang XR, Yannoukakos D, Zhang Y, Zheng W, Ziogas A, Dunning AM, Thompson DJ, Chenevix-Trench G, Chang-Claude J, Schmidt MK, Hall P, Milne RL, Pharoah PDP, Antoniou AC, Chatterjee N, Kraft P, García-Closas M, Simard J and Easton DF. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 2019; 104: 21-34.

[20] Tang Y, Tian W, Xie J, Zou Y, Wang Z, Li N, Zeng Y, Wu L, Zhang Y, Wu S, Xie X and Yang L. Prognosis and dissection of immunosuppressive microenvironment in breast cancer based on fatty acid metabolism-related signature. Front Immunol 2022; 13: 843515.

[21] Wang D, Wei G, Ma J, Cheng S, Jia L, Song X, Zhang M, Ju M, Wang L, Zhao L and Xin S. Identification of the prognostic value of ferroptosis-related gene signature in breast cancer patients. BMC Cancer 2021; 21: 645.

[22] Smith KA, Hunt KN, Rauch GM and Fowler AM. Molecular breast imaging in the screening setting. J Breast Imaging 2023; 5: 240-247.

[23] Freedman S and Wickramasekera IE 2nd. Review of the international hypnosis literature. Am J Clin Hypn 2020; 62: 159-165.

[24] Liu Z, Mi M, Li X, Zheng X, Wu G and Zhang L. A lncRNA prognostic signature associated with immune infiltration and tumour mutation burden in breast cancer. J Cell Mol Med 2020; 24: 12444-12456.

[25] Chen F, Yang J, Fang M, Wu Y, Su D and Sheng Y. Necroptosis-related lncRNA to establish novel prognostic signature and predict the immunotherapy response in breast cancer. J Clin Lab Anal 2022; 36: e24302.

[26] Chen DL, Cai JH and Wang CCN. Identification of key prognostic genes of triple negative breast cancer by LASSO-based machine learning and bioinformatics analysis. Genes (Basel) 2022; 13: 902.

[27] Zhang Z, Zeng X, Wu Y, Liu Y, Zhang X and Song Z. Cuproptosis-related risk score predicts prognosis and characterizes the tumor microenvironment in hepatocellular carcinoma. Front Immunol 2022; 13: 925618.

[28] Chen H, Luo H, Wang J, Li J and Jiang Y. Identification of a pyroptosis-related prognostic signature in breast cancer. BMC Cancer 2022; 22: 429.

[29] Zhang D, Zheng Y, Yang S, Li Y, Wang M, Yao J, Deng Y, Li N, Wei B, Wu Y, Zhu Y, Li H and Dai Z.

Identification of a novel glycolysis-related gene signature for predicting breast cancer survival. Front Oncol 2021; 10: 596087.
[30] Nasser M and Yusof UK. Deep learning based methods for breast cancer diagnosis: a systematic review and future direction. Diagnostics (Basel) 2023; 13: 161.
[31] Emanuelli S, Rizzi E, Amerio S, Fasano C and Cesarani F. Dosimetric and image quality comparison of two digital mammography units with different target/filter combinations: Mo/Mo, Mo/Rh, W/Rh, W/Ag. Radiol Med 2011; 116: 310-318.
[32] Zhou J, Zhang Y, Chang KT, Lee KE, Wang O, Li J, Lin Y, Pan Z, Chang P, Chow D, Wang M and Su MY. Diagnosis of benign and malignant breast lesions on DCE-MRI by using radiomics and deep learning with consideration of peritumor tissue. J Magn Reson Imaging 2020; 51: 798-809.
[33] Mao N, Shi Y, Lian C, Wang Z, Zhang K, Xie H, Zhang H, Chen Q, Cheng G, Xu C and Dai Y. Intratumoral and peritumoral radiomics for preoperative prediction of neoadjuvant chemotherapy effect in breast cancer based on contrast-enhanced spectral mammography. Eur Radiol 2022; 32: 3207-3219.
[34] Tang W, Xu F, Zhao M and Zhang S. Ferroptosis regulators, especially SQLE, play an important role in prognosis, progression and immune environment of breast cancer. BMC Cancer 2021; 21: 1160.
[35] Xu Y, Du Y, Zheng Q, Zhou T, Ye B, Wu Y, Xu Q and Meng X. Identification of ferroptosis-related prognostic signature and subtypes related to the immune microenvironment for breast cancer patients receiving neoadjuvant chemotherapy. Front Immunol 2022; 13: 895110.
[36] Su W, Hou X and Yu B. Value of dynamic contrast-enhanced magnetic resonance imaging in combination with mammography for screening early-stage breast cancer. Afr Health Sci 2023; 23: 290-297.