International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Machine Learning-Based Ultrasound Radiomics Signature: A Retrospective Analysis for Differentiating Benign from Malignant Non-Mass Lesions in Dense Breasts

Wei Chen*, Yudi Xiong*, Zirui Ke

Breast Cancer Center, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, National Key Clinical Specialty Discipline Construction Program, Hubei Provincial Clinical Research Center for Breast Cancer, Wuhan Clinical Research Center for Breast Cancer, Wuhan, Hubei, 430079, People's Republic of China

*These authors contributed equally to this work

Correspondence: Zirui Ke, Email kzr008@hotmail.com

Objective: The purpose of this study is to establish and validate a machine learning based ultrasound radiomics feature that combines clinical and ultrasound features, which can be used to identify benign and malignant non mass lesions (NML) of dense breast and evaluate its diagnostic value.

Methods: This study is a retrospective single center study. We included 619 patients with dense breast NML diagnosed by ultrasound from January 2017 to January 2023. The patients were randomly divided into training group (n=434) and validation group (n=185) according to the ratio of 7:3. About 848 radiomics features were extracted from two-dimensional ultrasound images and screened by lasso regression. The clinical model, ultrasound model, radiomics model and combined model were established. The diagnostic performance was evaluated by ROC curve, correction curve and decision curve analysis (DCA), and the model differences were compared by Delong test.

Results: There were 304 cases of malignant lesions and 315 cases of benign lesions confirmed by postoperative pathology. Multivariate logistic regression analysis showed that age, lesion length, microcalcification, surrounding structure distortion and blood flow were independent predictors of malignancy. Twelve non-zero coefficient radiomics features were selected to construct the radiomics features. The AUC of the combined model was the highest, the training set was 0.89 (95% CI: 0.86–0.92), and the validation set was 0.83 (95% CI: 0.78–0.89). Delong test showed that there were significant differences between the combined model and the other three single models (all $p<0.05$). The calibration curve showed good consistency in predicting the actual pathology, and DCA verified its best clinical application value.

Conclusion: We successfully combine ultrasound imaging, clinical and ultrasound features to build a prediction model, which has a good diagnostic effect on the differentiation of benign and malignant breast dense non-small cell lymphoma, and provides a reliable basis for clinical treatment decision-making of breast cancer.

Keywords: breast neoplasms, ultrasound radiomics, dense breasts, non-mass lesions, machine learning, diagnostic efficacy

Introduction

Non-mass lesions of the breast (NML), as a sub classification of breast cancer, account for about 9.21% of all breast abnormalities.[1,2] NML is defined by their morphological features, including diffuse echo texture changes that are different from the surrounding breast tissue, and the lack of well-defined spherical or oval masses with distinct margins, rather than the lack of measurable size.[2–4] Compared with typical mass lesions, its ultrasonic characteristics are not obvious, which brings great challenges to clinical diagnosis. Unlike well-defined mass lesions, NML is characterized by diffuse echo texture changes, which are different from the surrounding breast tissue and lack of unified classification

https://doi.org/10.2147/IJGM.S596385
Received: 15 January 2026
Accepted: 26 March 2026
Published: 16 April 2026

criteria for ultrasound imaging features.[3,5] This ambiguity leads to overlapping benign and malignant manifestations, resulting in low efficiency of clinical diagnosis.

It is worth mentioning that dense breasts, which are more prevalent in Asian women, are mainly composed of fiber and adipose tissue.[4,6] Traditional imaging modalities have inherent limitations, such as reduced sensitivity of mammography to dense breast lesions (especially noncalcified NML), while magnetic resonance imaging (MRI) may not detect low-grade ductal carcinoma in situ.[7–9] In contrast, ultrasound is not affected by breast density, can detect invisible lesions on mammograms, and has the advantage of identifying smaller, lower grade ductal carcinoma in situ, making it the preferred way to screen and diagnose dense breast lesions.[10,11] Moreover, traditional ultrasound diagnosis largely relies on the subjective experience of doctors, resulting in inconsistent judgments and insufficient specificity (previous studies ranged from 6.5% to 42.3%).[12,13] With the progress of artificial intelligence, radiomics has become a powerful tool to quantify high-dimensional imaging features to objectively assess tumor heterogeneity, providing an objective complement to subjective visual evaluation.[14,15] Although radiology based breast NML research has mainly focused on dynamic contrast-enhanced MRI, ultrasound radiology has greater clinical feasibility due to its wide availability, noninvasiveness, and cost-effectiveness.

In view of this, the purpose of this study is to construct ultrasound radiological features based on machine learning, and combine them with clinical and ultrasound features to establish a combined diagnostic model. By evaluating its performance in differentiating benign and malignant NML in dense breast, this study aims to provide a practical and reliable tool for clinical practice and solve the unmet need for accurate diagnosis of this challenging type of lesion.

Methods

Study Population

We retrospectively enrolled NML patients with dense breast admitted to Hubei Cancer Hospital from January 2017 to January 2023. The inclusion criteria were as follows: (1) ultrasound diagnosis of breast NML; (2) Breast density was confirmed by ultrasound or mammography; (3) Complete clinical data and definite pathological results were obtained by puncture biopsy, vacuum assisted biopsy or surgical resection. The exclusion criteria were as follows: (1) preoperative history of neoadjuvant chemotherapy, radiotherapy or endocrine therapy for breast disease; (2) Ultrasonic image quality is poor, too many artifacts or noise interference analysis; (3) Lesions cannot be measured on ultrasound images. Between January 2017 and January 2023, a total of 680 patients with non-mass lesions of dense breast were initially identified. According to the above criteria, 61 patients were excluded, including 18 with a history of neoadjuvant therapy, 25 with poor ultrasound image quality or severe artifacts/noise, and 18 with lesions that could not be measured on ultrasound. Finally, 619 patients were enrolled in this study.

This study was reviewed and approved by the ethics committee of Hubei Cancer Hospital (approval number: LLHBCH2026YN-026). All research procedures strictly comply with the principles of Helsinki Declaration (2013 Revision) and relevant national medical research ethics regulations. Considering the retrospective nature of the study, the ethics committee agreed to waive the written informed consent of individual patients. Before the start of the study, all patient data were de identified to delete personal identifiers (such as name, medical record number and contact information), and the data were managed and stored in a password protected secure database to ensure the privacy and confidentiality of patient information. Any access and use of the research data is limited to authorized researchers, and the data is only used for the purpose of this study and will not be disclosed to any third party. Upon reasonable request, the de identified data sets used in this study, including clinical, ultrasound and radiomics characteristics and pathological labels, can be obtained from the corresponding authors. The process of patient registration, screening and prediction model establishment is shown in Figure 1.

Imaging Acquisition and Processing

We used Ge logic E9 color Doppler ultrasound system with high-frequency linear array probe and mechanical index of 0.16. The patient was in supine position, his upper limbs were abducted, and his breast and armpit were fully exposed; Use side position if necessary. Repeat scanning from multiple angles and directions, adjust the gain, and focus on key areas. The two-dimensional

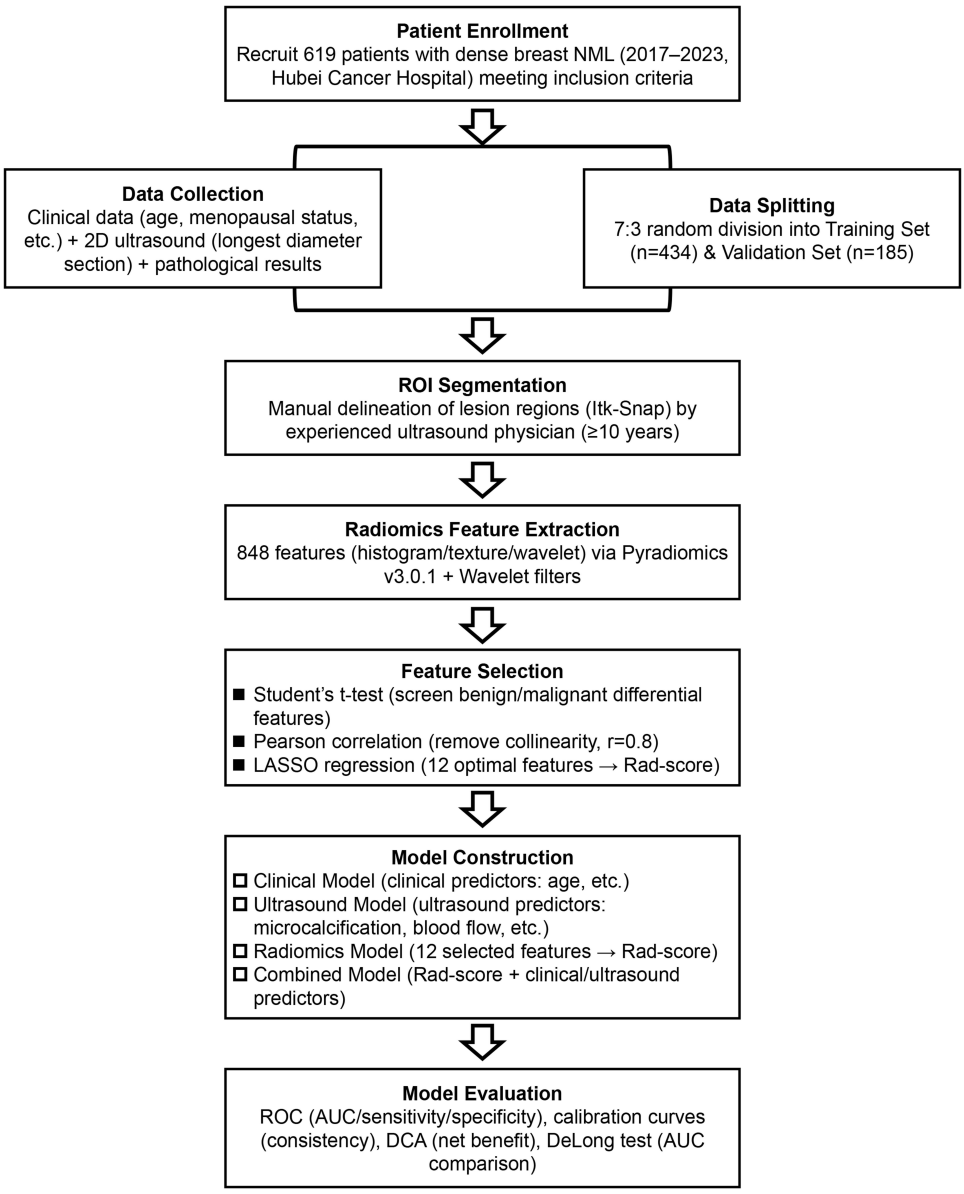

Figure 1 Flowchart of patient registration, screening and prediction model establishment.

ultrasound images of the longest diameter segment of the lesion were collected, and the ultrasonic characteristics of the lesion range, internal echo uniformity, microcalcification, posterior echo attenuation, and surrounding structure distortion were recorded. Color Doppler ultrasound was used to evaluate the blood flow characteristics.

All initial ultrasound studies were conducted by a committee certified breast ultrasound radiologist (non-technical) with at least 5 years of clinical experience in breast ultrasound imaging. Radiologists from the breast cancer center of Hubei Cancer Hospital were examined. All radiologists received unified training on intensive breast non mass lesions scanning scheme to ensure the consistency of image acquisition. All images are exported in DICOM format for subsequent analysis. Before feature extraction, Gaussian smoothing filter (sigma=0.5) was applied to all ultrasound images to reduce random speckle noise, and all images were clipped to the breast region to eliminate background artifacts. B-spline interpolation (3rd order) was used for image resampling, and the voxel size of each resampling step was $0.1 \times 0.1 \times 0.1$ mm.

Region of Interest Segmentation and Feature Extraction

Due to the complex and diverse ultrasound manifestations of NML, such as patchy hypoechogenicity, ductal patterns, microcalcifications, and glandular structural disorders, we arranged for ultrasound physicians with over 10 years of experience in breast ultrasound to manually delineate the region of interest (ROI) using Itk Snap software (version 3.8.0). Selecting the most prominent area of the lesion for ROI delineation does not rule out cystic changes, necrosis, or calcification within the lesion. ROI segmentation follows standardized standards: the ROI is delineated along the outermost boundary of ultrasound abnormalities in non mass lesions, including all areas of echo abnormalities, microcalcifications, structural distortions, or blood flow abnormalities. Use the longest diameter slice of the lesion identified by ultrasound for two-dimensional segmentation, and segment the original grayscale ultrasound images except for color Doppler images. The complete PyRadiomics configuration file (including discretization, resampling, filtering, and normalization parameters) is provided in Supplementary File 1.

The physician performing ROI segmentation does not know all the pathological results of the lesion to avoid subjective bias. For lesions with unclear boundaries (defined as the boundary where abnormal echoes mix with normal glandular tissue), the boundaries are determined by combining grayscale echo changes and color Doppler blood flow signals, and the segmentation is validated by a senior ultrasound physician with over 20 years of experience to ensure consistency. Evaluate intra observer variability in ROI segmentation: The same physician replicates segmentation of 50 randomly selected lesions every 4 weeks and calculates the intra group correlation coefficient (ICC) for all extracted radiomic features. The results showed that 98% of the features had an ICC>0.75, and 12 features had an ICC>0.85, indicating good consistency.

Data Preprocessing and Feature Selection

We used a random number generator to randomly divide the data into a training set and a validation set in a 7:3 ratio. Normalize all extracted radiomic features using the Z-score method to ensure they are on the same scale. Feature selection consists of three steps: (1) preliminary screening using Student's *t*-test for features that show significant differences between benign and malignant lesions; (2) Perform Pearson correlation analysis to remove highly correlated features (correlation coefficient threshold=0.8); (3) Based on the minimum squared error (MSE) criterion, LASSO regression is applied to select the optimal features and retain non-zero coefficient features to construct radiomics features.

All steps of feature selection, including Student's *t*-test, Pearson correlation analysis, and LASSO regression, are strictly performed only on the training set and cannot access the validation set data at any stage. The LASSO hyperparameter (λ) selection is performed using 5-fold cross validation on the training set, with the optimal λ determined by the minimum mean square error (MSE) criterion and a standard error rule to ensure model simplicity. The intensity scaling parameters (Z-score normalized mean and standard deviation) are only learned from the training set and applied to the validation set without the need for re estimation to avoid data leakage.

Model Construction

We have established four predictive models: (1) Clinical model: constructed using independent clinical predictive factors determined by multiple logistic regression; (2) Ultrasound model: construction of independent ultrasound prediction factors based on multiple logistic regression; (3) Radiomics model: Developed using LASSO logistic regression combined with selected radiomics features, calculating radiomics score (Rad score) for each patient by adding the product of each feature and its corresponding regression coefficient; (4) Combination model: Integrating Rad scores with independent clinical and ultrasound predictive factors.

Statistical Analysis

We used R software (version 4.3.2) for data statistical analysis and visualization. Quantitative data that conforms to a normal distribution are expressed as mean ± standard deviation (± s) and compared using independent sample *t*-test. Quantitative data that do not conform to a normal distribution are represented as median (interquartile range) [M (P25, P75)] and compared using Mann Whitney *U*-test. The categorical data is presented in the form of counts (percentages) and compared using chi square tests. Univariate and multivariate logistic regression analysis are used to determine

independent predictive factors for malignant tumors. The diagnostic performance of the model is evaluated through AUC, sensitivity, specificity, accuracy, balance precision, and F-score. Calibration curves are used to evaluate the consistency between predicted probabilities and actual results. Evaluate the clinical utility of DCA by calculating the net benefit at different threshold probabilities. The DeLong test is used to compare the AUC between models. Bilateral $P<0.05$ is considered statistically significant.

Results

Baseline Clinical and Pathological Characteristics of the Study Population

A total of 619 patients were ultimately enrolled, including 304 with malignant lesions (average age: 47.9 ± 9.4 years, range: 20–79 years) and 315 with benign lesions (average age: 42.2 ± 10.5 years, range: 20–76 years). Among them, the training set includes 207 malignant and 227 benign lesions, while the validation set includes 97 malignant and 88 benign lesions. As shown in Table 1, there was no significant difference in baseline clinical and pathological characteristics (including age, height, weight, menarche age, menopausal status, lesion location, and Rad score) between the training and validation sets (all $P>0.05$), indicating good homogeneity between the two groups. Overall, the research population maintains a balance between the training and validation sets, ensuring the reliability of subsequent model training and validation.

Identification of Independent Predictive Factors for Malignancy

In the training set, we used univariate logistic regression analysis to show that age, menopausal status, lesion length, lesion width, microcalcifications, peripheral structural deformation, and blood flow characteristics were potential factors associated with the malignancy of NML (all $P<0.05$); Supplementary Tables 1 and 2). Multivariate logistic regression analysis further confirmed that age (odds ratio [OR]=1.053, 95% CI: 1.027–1.080, $P<0.001$), lesion length (OR=0.701, 95% CI: 0.573–0.857, P=0.001), microcalcification (OR=8.197, 95% CI: 4.895–14.144, $P<0.001$), peripheral structural deformation (OR=3.479, 95% CI: 2.044–6.044, $P<0.001$), and blood flow characteristics (OR=1.195, 95% CI: 1.036–2.408, P=0.042) were independent predictive factors for malignant tumors. In short, age, lesion length, microcalcifications, surrounding structural deformations, and blood flow characteristics have been identified as key independent factors for predicting malignant NML, providing a basis for the construction of clinical and ultrasound models.

Table 1 Baseline Clinical and Pathological Characteristics of Patients in the Training and Validation Sets

Clinical and Pathological Characteristics	Training Set (n=434)	Validation Set (n=185)	Test Statistic	P value
Age (years, ±s)	45.1±10.4	45.1±10.7	t=0.025	0.98
Height (cm, ±s)	161.4±5.0	161.6±5.3	t=0.439	0.661
Weight (kg, ±s)	63.1±9.7	62.9±8.7	t=−0.334	0.739
Menarche age (years, ±s)	12.8±1.4	12.9±1.4	t=0.495	0.62
Rad-score (±s)	−0.007±0.209	0.018±0.207	t=1.366	0.172
Menopausal status [n (%)]			χ^2<0.001	0.977
No	335 (77.19)	143 (77.30)		
Yes	99 (22.81)	42 (22.70)		
Lesion location [n (%)]			χ^2=2.465	0.482
Upper outer quadrant	238 (54.84)	106 (57.30)		
Upper inner quadrant	90 (20.74)	38 (20.54)		
Lower outer quadrant	73 (16.82)	23 (12.43)		
Lower inner quadrant	33 (7.60)	18 (9.73)		
Pathological type [n (%)]			χ^2=1.164	0.281
Malignant	207 (47.70)	97 (52.43)		
Benign	227 (52.30)	88 (47.57)		

Notes: Rad-score = radiomics score; Data are presented as mean ± standard deviation (±s) for continuous variables and counts (percentages) for categorical variables; Test statistics include t-value for continuous variables and χ^2-value for categorical variables.

Table 2 Univariate and Multivariate Logistic Regression Analyses of Factors Associated with Malignancy of Non-Mass Lesions

Variables	Univariate Analysis		Multivariate Analysis	
	OR (95% CI)	P value	OR (95% CI)	P value
Age	1.048 (1.023–1.074)	<0.001	1.053 (1.027–1.080)	<0.001
Height (cm)	1.002 (0.985–1.019)	0.826	–	–
Weight (kg)	1.005 (0.992–1.018)	0.451	–	–
Menarche age	0.987 (0.943–1.033)	0.589	–	–
Menopausal status (Yes vs No)	3.215 (2.018–5.119)	<0.001	1.473 (0.811–2.688)	0.204
Lesion length	0.732 (0.598–0.895)	0.003	0.701 (0.573–0.857)	0.001
Lesion width	0.689 (0.552–0.861)	0.001	–	–
Microcalcification (Present vs Absent)	8.342 (5.091–13.706)	<0.001	8.197 (4.895–14.154)	<0.001
Distortion of surrounding structures (Present vs Absent)	3.618 (2.189–5.974)	<0.001	3.479 (2.044–6.044)	<0.001
Blood flow features (Grade ≥2 vs Grade <2)	1.842 (1.327–2.558)	<0.001	1.195 (1.036–1.378)	0.042

Abbreviations: OR, odds ratio; CI, confidence interval; "-" indicates variables not included in the multivariate model.

Selection of Radiomics Features and Construction of Radiomics Signature

After preliminary screening using Student's *t*-test, we retained 433 radiological features with significant differences between benign and malignant lesions. The subsequent Pearson correlation analysis excluded highly correlated features, and LASSO regression further screened out 12 non-zero coefficient features (Table 3), including 1 histogram feature, 2 texture features, and 9 wavelet features. These 12 features are used to construct radiological features and calculate Rad scores for each patient. Finally, 12 optimal radiomics features were determined through multi-step selection, forming a stable radiomics signature that can objectively quantify the imaging features of NML.

Diagnostic Performance of Individual Models

As shown in Figure 2 and Table 4, the AUC of the radiomics model in the training set is 0.782 (95% CI: 0.743–0.821), and the AUC in the validation set is 0.724 (95% CI: 0.659–0.789), with sensitivities of 74.9% and 77.3%, and specificities of 65.6% and 60.0%, respectively. It is gratifying that the ultrasound model showed better performance than the clinical model, with AUC values of 0.816 (95% CI: 0.780–0.853) and 0.754 (95% CI: 0.691–0.817) in the training and validation sets, respectively. The diagnostic efficiency of the clinical model is the lowest, with AUC values of 0.688 (95% CI: 0.644–0.732) and 0.641 (95% CI: 0.571–0.711) for the two groups, respectively. Among the three

Table 3 Selected Radiomics Features and Their Coefficients

Feature Category	Feature Name	Coefficient
Histogram feature (n=1)	Original image - First-order feature - Interquartile range	0.086881199
Texture features (n=2)	Original image - Gray-level co-occurrence matrix - Joint mean	0.036495593
	Original image - Gray-level co-occurrence matrix - Sum average	0.019147331
Wavelet features (n=9)	Wavelet transform - LLH - Gray-level dependence matrix - Large dependence emphasis	0.018194377
	Wavelet transform - LHL - First-order feature - Mean	−0.073102861
	Wavelet transform - LHL - Gray-level run length matrix - Long run low gray level emphasis	0.00804184
	Wavelet transform - HLH - Gray-level dependence matrix - Large dependence emphasis	0.059433795
	Wavelet transform - HLH - Gray-level dependence matrix - Low gray level large dependence emphasis	0.06233481
	Wavelet transform - HLH - Gray-level size zone matrix - Zone percentage	0.009762261
	Wavelet transform - HHH - Gray-level dependence matrix - Large dependence emphasis	−0.00552283
	Wavelet transform - LLL - First-order feature - Kurtosis	−0.017523206
	Wavelet transform - LLL - Gray-level dependence matrix - High gray level large dependence emphasis	−0.053530964

Abbreviations: L, low frequency; H, high frequency; LLH, LHL, HLH, HHH, LLL represent combinations of frequency information (e.g., LLH = low-low-high frequency combination).

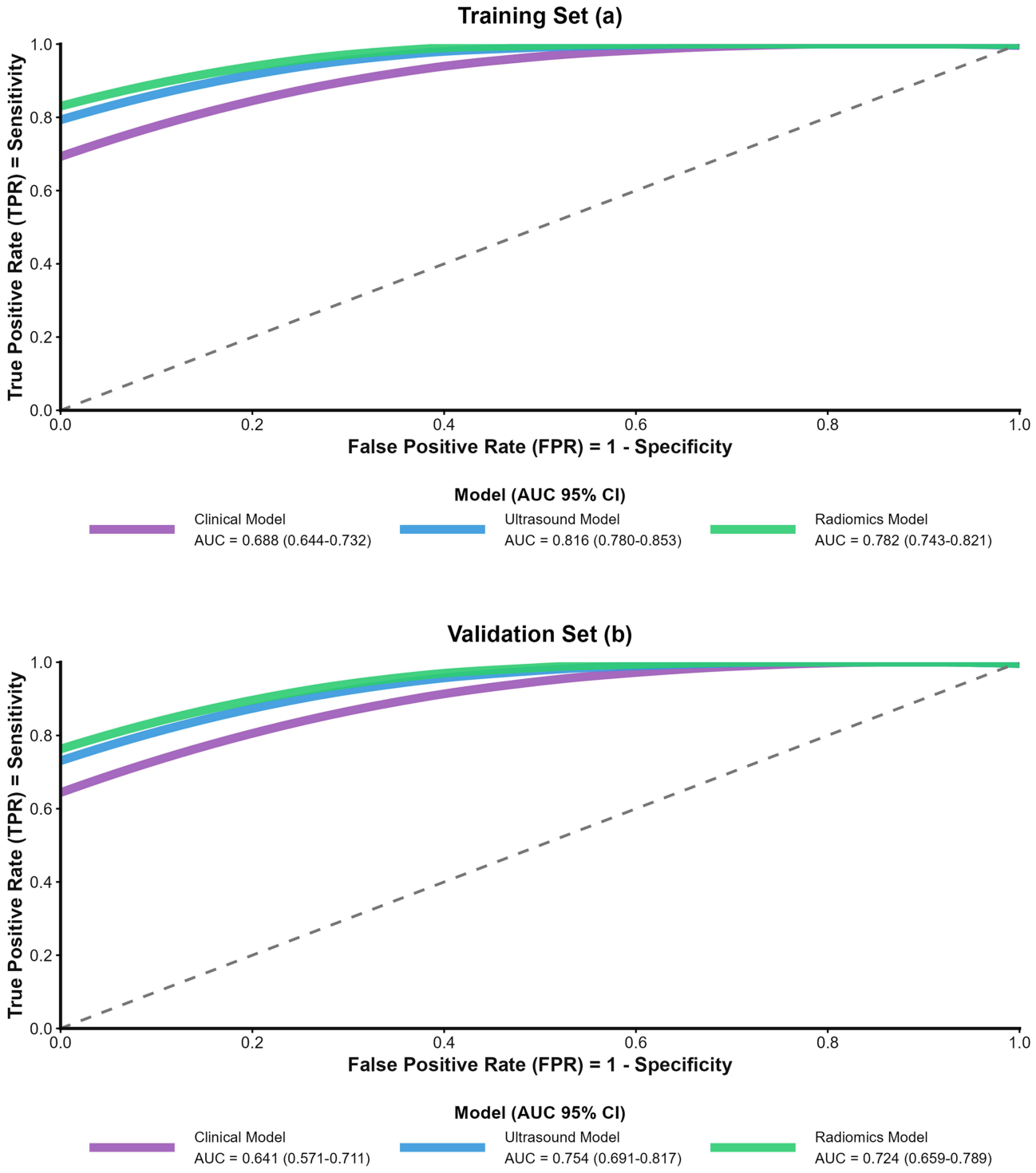

Figure 2 Receiver operating characteristic (ROC) curves of the four models in the training set (**a**) and validation set (**b**).

single models, the ultrasonic model showed the best diagnostic performance, followed by the radiological model, while the clinical model had relatively limited efficacy.

We combined the Rad score with independent clinical and ultrasound predictive factors in a combined model, and the results showed the highest diagnostic performance among all models. In the training set, the AUC of the combined model

https://doi.org/10.2147/IJGM.S596385

Table 4 Diagnostic Performance of Different Models in the Training and Validation Sets

Model	Set	Sensitivity	Specificity	Accuracy	Balanced Precision	F-Score	AUC (95% CI)
Clinical model	Training set	0.681	0.631	0.658	0.632	0.656	0.688 (0.644–0.732)
	Validation set	0.671	0.591	0.632	0.643	0.657	0.641 (0.571–0.711)
Ultrasound model	Training set	0.763	0.788	0.776	0.776	0.765	0.816 (0.780–0.853)
	Validation set	0.732	0.662	0.701	0.717	0.724	0.754 (0.691–0.817)
Radiomics model	Training set	0.749	0.656	0.704	0.665	0.705	0.782 (0.743–0.821)
	Validation set	0.773	0.6	0.686	0.675	0.721	0.724 (0.659–0.789)
Combined model	Training set	0.811	0.789	0.778	0.799	0.794	0.890 (0.860–0.920)
	Validation set	0.773	0.67	0.724	0.721	0.746	0.830 (0.780–0.890)

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval.

was 0.89 (95% CI: 0.86–0.92), sensitivity was 0.811, specificity was 0.789, accuracy was 0.778, balance accuracy was 0.799, and F-score was 0.794. In the validation set, the AUC was 0.83 (95% CI: 0.78–0.89), with corresponding values of 0.773, 0.670, 0.724, 0.721, and 0.746 (Supplementary Table 2 and Supplementary Figure 1). The DeLong test confirmed that there were significant differences between the combination model and the clinical model ($Z=-3.974$, $P<0.001$), ultrasound model ($Z=3.338$, $P=0.001$), and radiomics model ($Z=-3.468$, $P<0.001$) in the validation set. The calibration curve shows that compared with the other three models, the combined model has better consistency between predicted probability and actual pathological results (Figure 3). DCA indicates that the combined model achieved the highest net benefit in both the training set (DCA curve area=0.12) and the validation set (DCA curve area=0.22), demonstrating its superior clinical practicality (Figure 4). The comprehensive model is obviously superior to the single model in terms of diagnostic accuracy, consistency and clinical practicability, making it a promising clinical application tool.

Discussion

The diagnosis of NML in dense mammary glands has always been a clinical challenge due to the overlapping imaging features and the blurring effect of dense glandular tissue. In this study, we constructed a machine learning based composite model that combines ultrasound radiological features with clinical and ultrasound features, and demonstrated its superior diagnostic performance in distinguishing benign and malignant NML. The main findings and their clinical significance are as follows. It should be emphasized that radiomics features are complementary to composite models rather than dominant components. In the validation set, the AUC (0.724) of the radiomics model alone was lower than that of the ultrasound model alone (0.754), but the integration of radiomics features with clinical and ultrasound features significantly improved the diagnostic performance of the model (AUC increased from 0.754 to 0.83), the correction rate of misclassified cases by the combined model is shown in Supplementary Table 3. This is because radiomics captures subtle quantitative texture and structural changes that are not visible in visual ultrasound assessment, providing unique supplementary information for traditional clinical and ultrasound features. Among all test models, the combination model had the highest AUC (0.89 for the training set and 0.83 for the validation set), indicating a significant improvement in specificity compared to a single radiomics model. Encouragingly, this advantage can be attributed to the complementary information provided by clinical, ultrasound, and radiological features.

Previous studies have shown that clinical factors such as age reflect the inherent risk of cancer, as aging is associated with metabolic and immune changes, accumulation of primary genes, and insufficient DNA methylation, all of which increase susceptibility to cancer.[16–18] Ultrasound features include microcalcifications, deformation of surrounding structures, and blood flow status, which directly reflect the biological characteristics of the lesion. Microcalcification is associated with high metabolic activity and local ischemia or necrosis of tumor cells. The deformation of the surrounding structure indicates invasive growth. The increase in blood flow reflects tumor induced angiogenesis to support rapid proliferation of tumor cells. On the other hand, radiological features quantify the heterogeneity of lesions at the microscopic level, capturing subtle texture and echo changes that are not visible to the naked eye. However, many previously published predictive models have performed poorly in clinical validation, failing to provide satisfactory

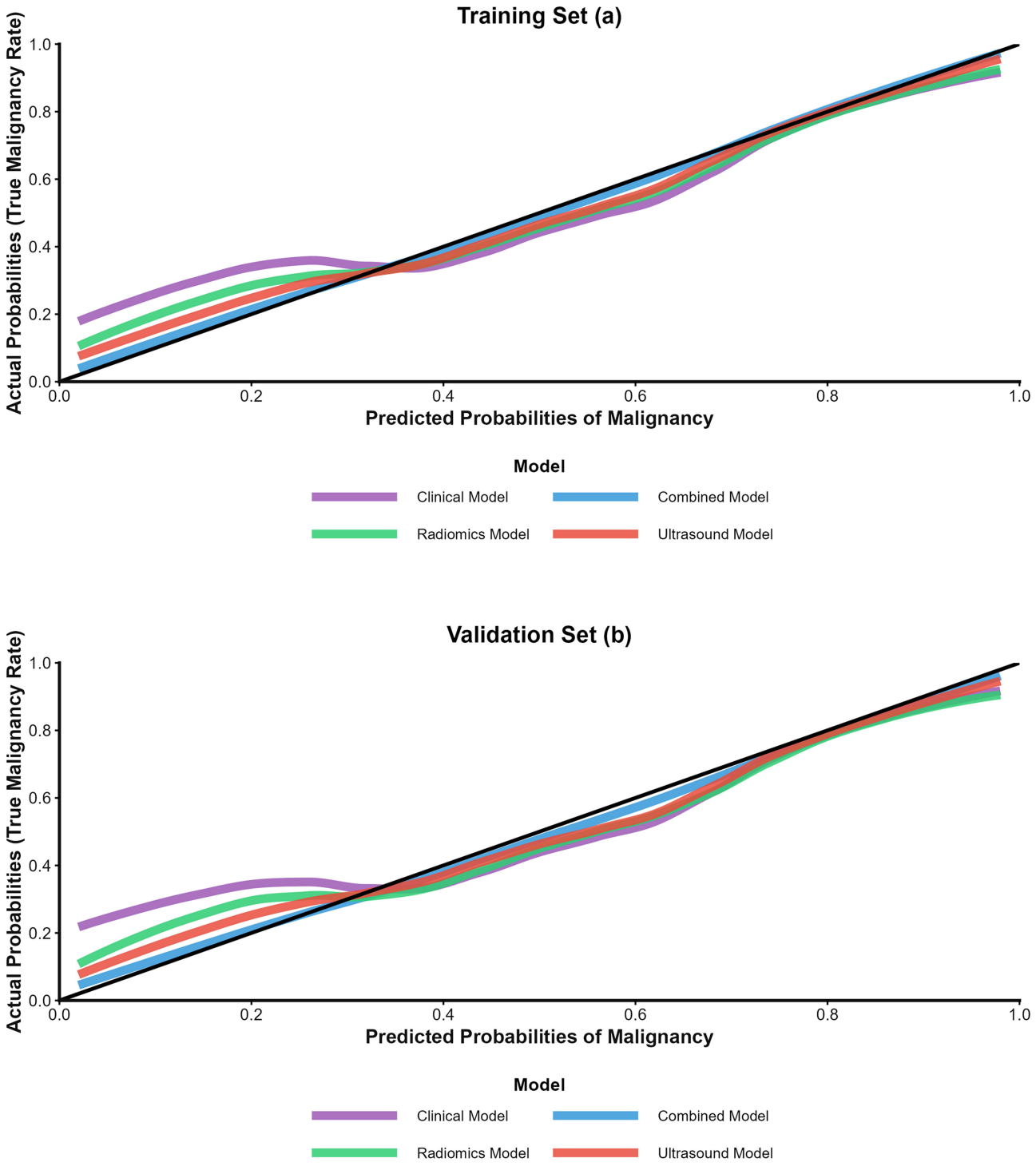

Figure 3 Calibration curves of the four models in the training set (**a**) and validation set (**b**). The x-axis represents predicted probabilities of malignancy, and the y-axis represents actual probabilities (true malignancy rate). The diagonal black line indicates perfect calibration. The combined model (red curve) is closest to the perfect calibration line, indicating better consistency between predicted and actual outcomes compared with the other models.

predictive performance for lesion features.[19–21] By integrating these multidimensional data, the composite model overcomes the limitations of single modal models, which rely on incomplete information and therefore have suboptimal diagnostic performance.

https://doi.org/10.2147/IJGM.S596385

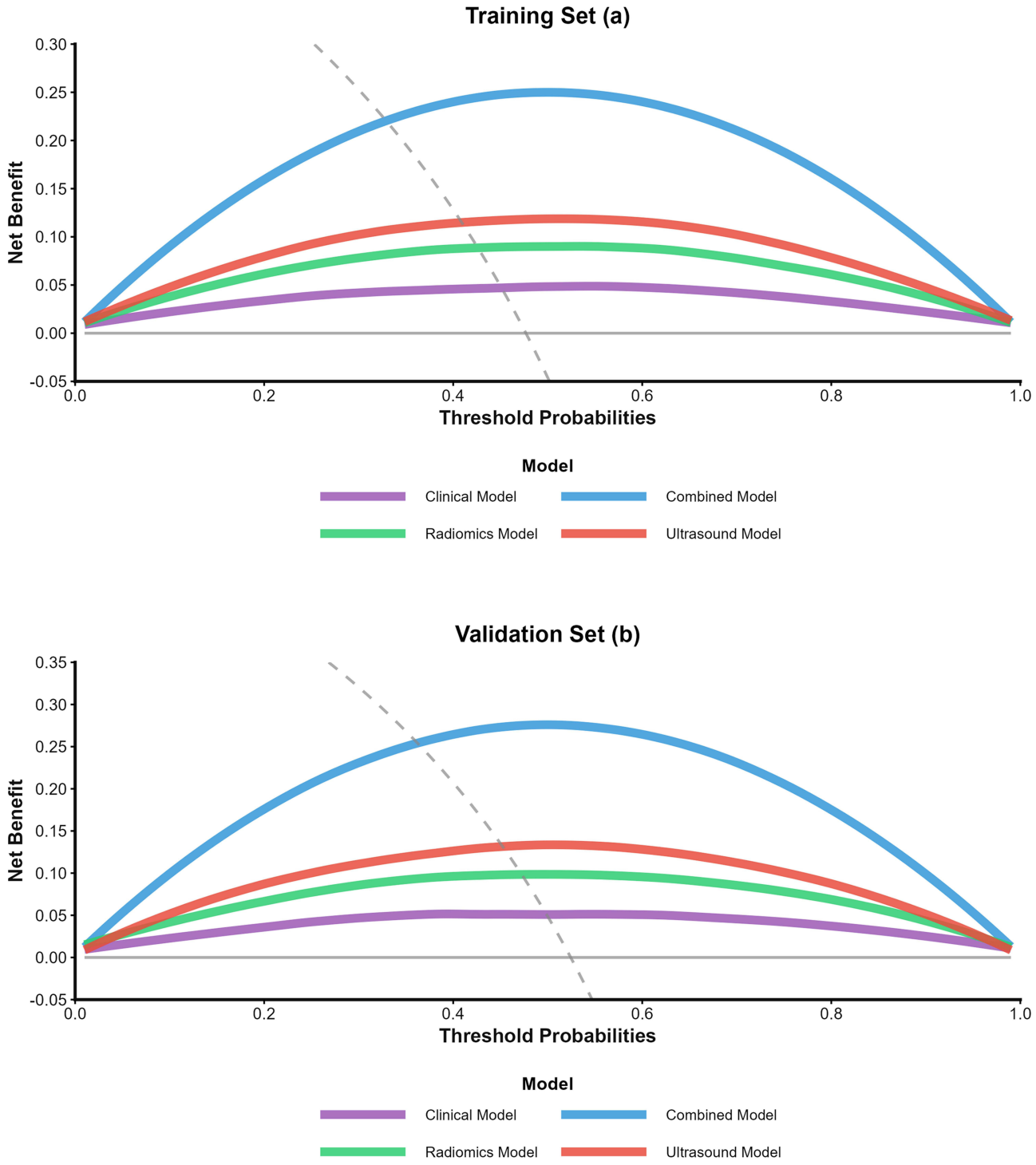

Figure 4 Decision curve analysis (DCA) of the four models in the training set (**a**) and validation set (**b**). The x-axis represents threshold probabilities, and the y-axis represents net benefit. The gray curves represent the "treat none" and "treat all" strategies for reference.

The identification of independent predictive factors in this study is consistent with previous research results in some aspects, but also shows certain differences. Consistent with previous studies, microcalcifications were found to be the strongest predictor of malignant tumors, highlighting their critical role in the diagnosis of NML.[22,23] Malignant lesions are also associated with older age, which is consistent with the epidemiological characteristics of cancer. However, this

study found a negative correlation between lesion length and malignancy, which contradicts previous reports suggesting that the larger the lesion size, the higher the risk of malignancy. This difference may be due to the unique pathological composition of NML in dense mammary glands: benign lesions such as fibrocystic changes may manifest as diffuse large-area lesions, while malignant NML in dense mammary glands may be more localized in the early stages due to the restrictive effect of dense glandular tissue.

In addition, the frequency of structural distortions around malignant lesions observed in this study is relatively high, which is consistent with the invasiveness of cancer. However, due to differences in the study population and pathological types, the degree of association may vary. These findings emphasize the need to develop tailored diagnostic criteria for NML in dense breasts, taking into account the specific characteristics of this type of lesion. The potential risk of overfitting in radiomics models has been effectively addressed through various robust validation methods, such as using 5-fold nested cross validation instead of single random segmentation validation, adding 1000 guided validations to quantify the variability of performance indicators, and training early cohorts and testing late cohorts for time validation. The results show that the model performance of all validation methods is consistent, indicating no significant overfitting and reliable model performance.

In this study, we provided valuable insights into the pathological basis of NML by examining 12 selected radiological features, including histograms, textures, and wavelet features. Among them, histogram features (such as interquartile range) reflect the distribution of grayscale values in lesions, which is related to the uniformity of tissue composition. Texture features, such as the joint mean and mean of the gray level co-occurrence matrix, quantify the spatial relationship of gray values and reflect the structural heterogeneity of lesions. Wavelet features originate from multi-scale decomposition of images, capturing subtle changes in tissue texture at different frequencies, which are related to the microstructural differences between benign and malignant tissues. Integrating these features into radiomics can objectively and quantitatively evaluate the heterogeneity of lesions, reducing the subjectivity of traditional ultrasound diagnosis. This is particularly important for NML, as visual assessment is often challenging due to boundary blurring and feature overlap.

Compared with previous radiomics studies on breast NML, our study has several significant advantages. Most previous studies have focused on DCE-MRI, which is an expensive, time-consuming, and not widely used method in primary hospitals.[24,25] In contrast, ultrasound is a routine clinical examination with high accessibility and low cost, making the proposed model more feasible in clinical practice. In addition, previous ultrasound based NML studies often relied on subjective visual features or a limited number of quantitative indicators, while this study extracted 848 high-dimensional radiological features and used machine learning to select the best features, ensuring the comprehensiveness and accuracy of the features.[26,27] In addition, integrating clinical and ultrasound features into radiological models can improve their interpretability, as independent predictive factors such as microcalcifications and age have clinical significance and are familiar to doctors, which increases the model's acceptance in clinical practice. It is worth noting that a slight increase in AUC of the combined model compared to the individual ultrasound model does not necessarily mean that its clinical value is small. DCA reflects the actual clinical net benefit of the model, which is a more clinically relevant indicator than AUC. This combination model has significantly higher net benefits within a wide range of clinically relevant threshold probabilities, indicating that compared to traditional clinical combined ultrasound models, it can identify more malignant lesions, reduce more unnecessary biopsies, and has substantial clinical value.

The clinical practicality of this combination model is further supported by its calibration and DCA performance. The calibration curve shows good consistency between the predicted and actual results, indicating that the model can provide reliable probability estimates for individual patients. DCA indicates that within a wide range of threshold probabilities, the combined model achieved higher net benefits than other models, which means its application can help clinicians make wiser decisions, avoid unnecessary biopsies of benign lesions, and ensure timely intervention for malignant lesions. This is particularly valuable in clinical practice, as unnecessary biopsy not only increases the physical and mental burden on patients, but also consumes medical resources. As an auxiliary tool for BI-RADS assessment, this combination model can improve the accuracy of risk stratification for dense breast NML and optimize management strategies for these patients.

Our research inevitably has certain limitations. First, this is a single center retrospective study. The sample size is relatively small, because non breast cancer is rare, which may introduce selection bias and limit the generalization of the results. Future multicenter studies with larger sample sizes are needed to validate this model. Secondly, ROI

segmentation is manually performed by a doctor, which may be influenced by subjective differences; In future research, automatic segmentation methods should be explored to improve reproducibility. Thirdly, the model does not include other potential predictive factors such as genetic markers or molecular subtypes, which may further improve its diagnostic efficiency. Despite these limitations, this study provides a promising approach for the diagnosis of dense breast NML and has significant clinical implications.

Conclusion

In summary, our machine learning based combination model combines ultrasound radiological features with clinical and ultrasound features, and has excellent diagnostic performance in distinguishing benign and malignant non mass lesions in dense breasts. It outperforms single modal models in accuracy, consistency, and clinical practicality, providing a reliable and feasible tool for supporting clinical decision-making, optimizing patient management, and improving cancer patient prognosis.

Funding

This study is supported by the National Key Clinical Specialty Construction Discipline Project (HBCHBCC-D09) and Knowledge innovation project of Wuhan (2022020801020514).

Disclosure

Wei Chen and Yudi Xiong are co-first authors for this study. All authors declare no competing interests related to this study.

References

1. Yamaguchi R, Watanabe H, Mihara Y, Yamaguchi M, Tanaka M. Histopathology of non-mass-like breast lesions on ultrasound. *J. Med. Ultrasound*. 2023;50(3):375–380. doi:10.1007/s10396-023-01286-y
2. Uematsu T. Non-mass lesions on breast ultrasound: why does not the ACR BI-RADS breast ultrasound lexicon add the terminology? *J. Med. Ultrasound*. 2023;50(3):341–346. doi:10.1007/s10396-023-01291-1
3. Arian A, Teymouri Athar MM, Nouri S, et al. Role of breast MRI BI-RADS descriptors in discrimination of non-mass enhancement lesion: a systematic review & meta-analysis. *Eur. J. Radiol*. 2025;185:111996. doi:10.1016/j.ejrad.2025.111996
4. Kim SH, Kim HH, Moon WK. Automated Breast Ultrasound Screening for Dense Breasts. *Korean j Radiol*. 2020;21(1):15–24. doi:10.3348/kjr.2019.0176
5. Li H, Cong P, Yu YN, Zhang YF. Value of sonoelastography for diagnosis of breast non-mass lesions and comparison with BI-RADS: a systematic review and meta-analysis. *Medicine*. 2024;103(23):e38425. doi:10.1097/MD.0000000000038425
6. Berg WA, Rafferty EA, Friedewald SM, Hruska CB, Rahbar H. Screening algorithms in dense breasts: ajr expert panel narrative review. *AJR*. 2021;216(2):275–294. doi:10.2214/AJR.20.24436
7. Nissan N, Ochoa Albiztegui RE, Fruchtman-Brot H, et al. Extremely dense breasts: a comprehensive review of increased cancer risk and supplementary screening methods. *Eur. J. Radiol*. 2025;182:111837. doi:10.1016/j.ejrad.2024.111837
8. Ren W, Chen M, Qiao Y, Zhao F. Global guidelines for breast cancer screening: a systematic review. *Breast*. 2022;64:85–99. doi:10.1016/j.breast.2022.04.003
9. Supplemental screening as an adjunct to mammography for breast cancer screening in people with dense breasts: a health technology assessment. *Ont Health Technol Assess Ser*. 2023;23(9):1–293.
10. Thigpen D, Kappler A, Brem R. The role of ultrasound in screening dense breasts-a review of the literature and practical solutions for implementation. *Diagnostics*. 2018;8(1):20. doi:10.3390/diagnostics8010020
11. Lee SE, Yoon JH, Son NH, Han K, Moon HJ. Screening in patients with dense breasts: comparison of mammography, artificial intelligence, and supplementary ultrasound. *AJR*. 2024;222(1):e2329655. doi:10.2214/AJR.23.29655
12. Hadadi I, Clarke J, Rae W, McEntee M, Vincent W, Ekpo E. Reducing unnecessary biopsies using digital breast tomosynthesis and ultrasound in dense and nondense breasts. *Curr oncol*. 2022;29(8):5508–5516. doi:10.3390/curroncol29080435
13. Hadadi I, Clarke J, Rae W, McEntee M, Vincent W, Ekpo E. Diagnostic efficacy across dense and non-dense breasts during digital breast tomosynthesis and ultrasound assessment for recalled women. *Diagnostics*. 2022;12(6):1477. doi:10.3390/diagnostics12061477
14. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics*. 2017;37(2):505–515. doi:10.1148/rg.2017160130
15. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Internal Med*. 2018;284(6):603–619. doi:10.1111/joim.12822
16. Katsura C, Ogunmwonyi I, Kankam HK, Saha S. Breast cancer: presentation, investigation and management. *Br J Hosp Med*. 2022;83(2):1–7. doi:10.12968/hmed.2021.0459
17. Barzaman K, Karami J, Zarei Z, et al. Breast cancer: biology, biomarkers, and treatments. *Int Immunopharmacol*. 2020;84:106535. doi:10.1016/j.intimp.2020.106535
18. Tsang JYS, Tse GM. Molecular Classification of Breast Cancer. *Adv Anat Pathol*. 2020;27(1):27–35. doi:10.1097/PAP.0000000000000232

19. Perić I, Brkljačić B, Tadić T, et al. DWI in the differentiation of malignant and benign breast lesions presenting with non-mass enhancement on CE-MRI. *Cancers*. 2024;17(1):31. doi:10.3390/cancers17010031
20. Liu J, Chen J, Qiu L, et al. The value of intratumoral and peritumoral ultrasound radiomics model constructed using multiple machine learning algorithms for non-mass breast cancer. *Sci Rep*. 2025;15(1):19953. doi:10.1038/s41598-025-03704-2
21. Bumberger A, Clauser P, Kolta M, et al. Can we predict lesion detection rates in second-look ultrasound of MRI-detected breast lesions? A systematic analysis. *Eur. J. Radiol*. 2019;113:96–100. doi:10.1016/j.ejrad.2019.02.008
22. Cerimele F, Tagliati C, Salvatori F, et al. Invasive ductal carcinoma mammographic findings: correlation with age, breast composition and tumor size. *Breast Disease*. 2022;41(1):45–49. doi:10.3233/BD-201072
23. Shakki Katouli F, Salehi N, Soveyzi F, et al. A systematic review and meta-analysis of imaging characteristics and upgrade rates in noninvasive lobular neoplasia of the breast. *Eur j Radiol Open*. 2025;15:100691. doi:10.1016/j.ejro.2025.100691
24. Spear GG, Mendelson EB. Automated breast ultrasound: supplemental screening for average-risk women with dense breasts. *Clin. Imaging*. 2021;76:15–25. doi:10.1016/j.clinimag.2020.12.007
25. Lee SH, Moon WK. Glandular tissue component on breast ultrasound in dense breasts: a new imaging biomarker for breast cancer risk. *Korean j Radiol*. 2022;23(6):574–580. doi:10.3348/kjr.2022.0099
26. Penn A, Medved M, Abe H, Dialani V, Karczmar GS, Brousseau D. Safely reducing unnecessary benign breast biopsies by applying non-mass and DWI directional variance filters to ADC thresholding. *BMC Med. Imaging*. 2022;22(1):171.
27. Ziada K, Siu M, Gatt R, et al. High-risk breast cancer screening with MRI: diagnostic performance of a local scoring system. *Clin Radiol*. 2025;87:106968. doi:10.1016/j.crad.2025.106968