Radiation Oncology

RESEARCH

Open Access

Intra- and peri-tumoral radiomics based on dynamic contrast-enhanced MRI for prediction of benign disease in BI-RADS 4 breast lesions: a multicentre study

Yalan Hu[1,2†], Zhenhai Cai[3†], Nijiati Aierken[4†], Yueqi Liu[1], Nan Shao[1], Yawei Shi[1], Mengmeng Zhang[1], Yangling Hu[5], Xiaoling Zhang[5*] and Ying Lin[1*]

Abstract

Background and purpose The study aimed to create a radiomics model based on breast intra- and peri-tumoral regions in dynamic contrast-enhanced (DCE) MRI to distinguish benign from malignant breast lesions of Breast Imaging Reporting and Data System (BI-RADS) 4.

Materials and methods A total of 516 patients from Hospital 1 were assigned to the training cohort. Then, 146 and 52 patients were enrolled from Hospital 2 and 3, respectively, as the internal and external test cohort. Seven classification models were built, using features extracted from the intra- and peri-tumoral regions. Diagnostic performance was evaluated by receiver operating characteristics (ROC) analysis and compared by the DeLong test. Subgroup analysis was performed after stratifying all lesions by enhancement pattern and the subdivision of BI-RADS 4.

Results The Comb2 model, built with features from peri-tumoral 2 mm and intra-tumoral region, demonstrated the best performance with AUCs of 0.828 and 0.844 in the internal and external test cohort, respectively. The Comb2 model was robust in both mass and non-mass enhancement (NME) lesions. At the three exploratory cutoff values on the ROC curve, the model identified 9.1% (sensitivity of C1 ≥ 98%), 27.3% (sensitivity of C2 ≥ 95%) and 36.4% (sensitivity of C3 ≥ 90%) of the benign lesions in the external test cohort. Applying the identified cutoff values in the external test cohort showed the potential to lower the number of unnecessary biopsies of benign lesions.

Conclusion An MRI-based radiomics model built with features extracted from the intra-tumoral region combined with the peri-tumoral 2 mm showed the best potential to reduce false-positive diagnoses and may avoid unnecessary biopsies with a low underestimate risk.

†Yalan Hu, Zhenhai Cai and Nijiati Aierken are co-first authors and contributed equally to this work.

*Correspondence:
Xiaoling Zhang
zhxiaol6@mail.sysu.edu.cn
Ying Lin
linying3@mail.sysu.edu.cn

Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Keywords DCE-MRI, BI-RADS, Breast cancer, Radiomics, Machine learning

Introduction

Breast cancer is one of the most common cancers and the leading cause of cancer deaths in women, accounting for approximately 25% of new cancer cases in women and 16% of cancer deaths [1]. Since the treatment options and prognosis for benign and malignant breast diseases differ, determining whether breast lesions are benign or malignant is critical.

As a noninvasive imaging method with very high sensitivity, dynamic contrast-enhanced MRI (DCE-MRI) has gained a pivotal role in breast cancer management [2, 3]. It allows the detection of breast cancers invisible on mammography in women with dense breast tissue [4–6] and can aid in the management of equivocal findings on mammogram and ultrasound [7]. The diagnostic challenge in DCE-MRI remains to distinguish benign from malignant enhancement. According to the Breast Imaging Reporting & Data System (BI-RADS) [8], breast lesions can be classified into 6 categories based on the probability of malignancy. BI-RADS category 4 (BI-RADS 4) indicates a lesion that does not fulfill the criteria for malignancy but is suspicious enough to warrant biopsy. In women referred to biopsy due to BI-RADS 4 findings, 86% of these lesions yield benign results [9]. Therefore, to avoid false-positive diagnoses, unnecessary biopsies or overtreatment, methods for distinguishing benign from malignant in BI-RADS 4 lesions are warranted.

In recent years, machine learning methods based on medical imaging, known as radiomics, have shown increasing clinical application value [10]. Radiomics can automatically extract large amounts of quantitative features that are beyond human recognition to build clinically useful models for disease diagnosis and prognosis [11, 12]. According to previous studies, radiomics models based on breast MRI have shown great potential to discriminate benign from malignant BI-RADS 4 lesions [13–18]. The majority of these investigations were single-center, had small sample sizes, and lacked external validation. Furthermore, only the features extracted from the tumor area were used in model construction. The interactions between tumor cells and the associated stroma are significantly associated with the disease progression and patient prognosis [19]. Therefore, the texture features derived from the areas surrounding the tumor may also provide valuable information to aid in diagnosis. Niu et al. have shown that features from peri-tumoral regions with 2 mm dilation distances in digital breast tomosynthesis (DBT) can be used in the differentiation of benign and malignant breast lesions [20]. For MRI, prior studies have noted the importance of the peri-tumoral region in the diagnosis [21], prediction of treatment response [22], and assessment of molecular biomarkers of breast cancer [23, 24]. However, no published study reported the predictive value of the peri-tumoral region in distinguishing benign from malignant breast lesions of BI-RADS 4 based on DCE-MRI.

Accordingly, the aim of the study was to build a radiomics model using features extracted from the intra- and peri-tumoral regions to distinguish benign from malignant breast diseases in a multicenter study. Furthermore, we aimed to explore the optimal peri-tumoral size for investigation.

Methods

Patients

This multicenter retrospective study was approved by the Institutional Review Board and exempted from patient informed consent. According to the following criteria, patients who underwent breast DCE-MRI between 01 January 2017 and 31 August 2021 were reviewed in Hospital 1 (The First Affiliated Hospital of Sun Yat-sen University). Each of the patients was classified according to the fifth version of the BI-RADS guidelines [8]. The criteria for inclusion were as follows: (i) Patients were diagnosed with a BI-RADS 4 or 5 breast lesion by breast MRI. (ii) Patients with benign or malignant breast diseases confirmed by histopathology; and (iii) Patients had complete MRI images in the picture archiving and communication system (PACS) with axial DCE, DWI, and T2WI sequences obtained before patients underwent biopsy. The exclusion criteria were as follows: (i) nondiagnostic examinations (examination interrupted by the patient and artifacts); (ii) multi-foci lesions; (iii) received excision biopsy or anticancer therapy before MRI examinations; (iiii) incomplete clinicopathologic data. The flowchart of the whole study in Fig. 1.

Considering the future application of the model, the gain of preventing false-positive diagnosis was expected to be low in the BI-RADS 5 category. Therefore, model performance was evaluated only in BI-RADS 4 lesions. Patients from Hospital 1 were enrolled into the training cohort for model training and validation, and all BI-RADS 5 lesions were excluded from validation. Then, patients were enrolled from Hospital 2 (Jieyang People's Hospital) and Hospital 3 (The Seventh Affiliated Hospital of Sun Yat-sen University) with the same criteria, but only patients with BI-RADS 4 lesions were included. Patients from Hospital 2 and 3 were allocated to the internal and external test cohort, respectively. The recruitment of patients is shown in the Supplementary Figure S1.

Clinical data, such as age, BI-RADS category and pathology results were obtained by reviewing the medical

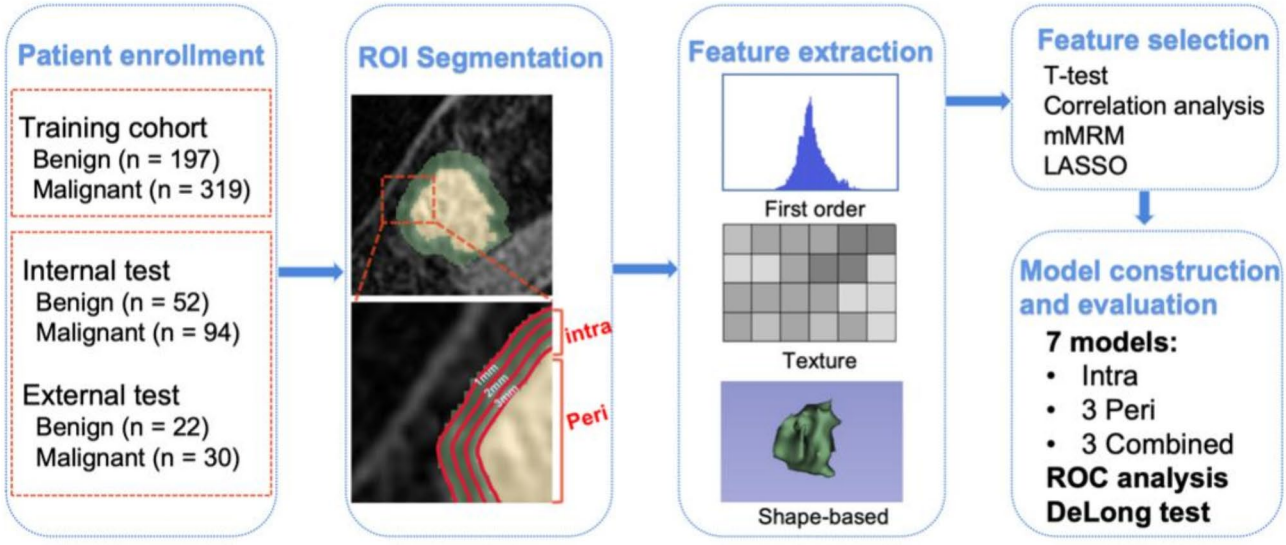

Fig. 1 The design of the whole study

records. All MR scans were reviewed by two radiologists with 5 and 20 years of experience in breast MRI interpretation. They recorded the enhancement pattern, and patients were grouped into mass and NME groups. Besides, they stratified the BIRADS 4 lesions into 4a, 4b, and 4c categories according to the BI-RADS 6th Edition Update presented at the 2023 RSNA annual meeting.

Image acquisition

MRI examinations were performed with a 3.0T MRI scanner equipped with a dedicated eight-channel bilateral breast coil on patients in a prone position. The standard breast MRI examinations included T1W, T2W, DWI, and DCE sequences. After intravenous injection of the contrast agent 0.15 mmol/kg at 4 ml/s and subsequent flushing with an equal volume of saline at the same injection speed, eight phases of post-contrast scans were consecutively acquired. For each patient, eight phases of subtraction images were obtained by subtracting pre-contrast images from eight post-contrast images. The detailed MRI scanning parameters of T1-DCE, FS-T2WI and ADC map in each center were listed in Supplementary Table S1. All images were extracted in DICOM format from the picture archiving and communication system.

Image preprocessing and ROI segmentation

The peak enhanced phase of the enhancement T1 images was selected according to the time intensity curve for subsequent preprocessing and segmentation. To correct the variation in imaging signal intensity caused by the bias field of MRI scanning, the N4 Bias Field Correction algorithm [25] was used before tumor segmentation. To achieve intensity homogeneity, the range of voxel intensity in MR image was scaled to 0-255, to avoid the influence of imaging intensity. Eventually, all images were resampled to 1*1*1 mm isotropic voxels using a trilinear interpolation algorithm.

After imaging prepossessing, the regions of interest (ROIs) of intra-tumoral regions were delineated manually via 3D slicer (Version 4.1; www.slicer.org) on each slice of image by a radiologist with 5 years of experience and confirmed by another radiologist with 20 years of experience. The peri-tumoral regions were then obtained by equidistant 3-dimensional dilation of the tumor border by 1, 2 and 3 mm, respectively, and the skin and chest wall were excluded.

Radiomic feature extraction

Radiomic features were extracted from ROIs using the Pyradiomics toolkit (https://github.com/Radiomics/pyradiomics). In the intra-tumoral region, 19 intensity-based first-order statistical features, 14 shape-based features (3D), 24 Gy level co-occurrence matrix (GLCM) features, 16 Gy level size zone matrix (GLSZM) features, 16 Gy level run length matrix (GLRLM) features, 5 neighborhood gray-tone difference matrix (NGTDM) features, and 14 neighboring gray level dependence matrix (NGLDM) features were extracted from the original images. Moreover, Laplacian of Gaussian(LoG) imaging filters (kernel size: 1, 2, 3, 4, 5, and 6) and wavelet imaging filters were used to deal with all the original images. Eventually, a total of 1409 intensity-based first-order statistical features and texture features were calculated and extracted from the intra-tumoral region. Features from the 1–3 mm peri-tumoral regions of were extracted following the same regimen. To avoid the influence of lesion position, the shape-based features were excluded,

yielding 1395 features from each peri-tumoral region. The calculation formulas of radiomic features are listed in the Supplementary Material1. To ensure a uniform scale of value, all the radiomics features from the training cohort were standardized using z-score normalization. The mean value of all lesions was subtracted by the feature value, and then the standard deviation (SD) was divided by the result. All features were standardized to a mean of zero and SD of one. Then the features of patients in the test cohorts were transformed according to the corresponding feature value in the training cohort.

Feature selection

To choose the feature that was relevant to the classification of benign and malignant lesions, a series of feature selection methods were applied to the following seven feature sets in the training cohort: (1) the intra-tumoral region, (2) the peri-tumoral region of 1 mm, (3) the peri-tumoral region of 2 mm, (4) the peri-tumoral region of 3 mm, (5) the intra-tumoral region combined with the peri-tumoral region of 1 mm (6) the intra-tumoral region combined with the peri-tumoral region of 2 mm, and (7) the intra-tumoral region combined with the peri-tumoral region of 3 mm.

Initially, all features underwent statistical testing using the U-test and T-test, retaining features with a p-value less than 0.05. Next, correlation analysis was performed using Pearson's correlation to identify and exclude highly correlated features, applying a threshold of 0.9. To further refine the feature set, the Minimum Redundancy Maximum Relevance (mRMR) algorithm was employed, balancing feature relevance and redundancy, and reducing the feature pool to 64 candidates. The final feature selection was conducted using the Least Absolute Shrinkage and Selection Operator (LASSO) regression. This approach penalizes regression coefficients to eliminate irrelevant features, ensuring the model's focus on predictive factors. The optimal regularization parameter (λ) was identified through 10-fold cross-validation, which retained only the most critical features for modeling. This comprehensive process, from statistical filtering to LASSO-based refinement, yielded a robust and parsimonious radiomic signature for predictive modeling.

Model development and evaluation

To determine which region of the tumor showed the best classification ability, we built machine learning models using features extracted from each of the above seven feature sets based on Support vector machine (SVM), RandomForest (RF) and Light Gradient Boosting Machine (GBM), respectively. The hyperparameters were tuned by a grid search approach and 10-fold cross-validation, as shown in Fig. 2. The 10-fold cross-validation refers to the random division of the data set into 10 sets, nine of

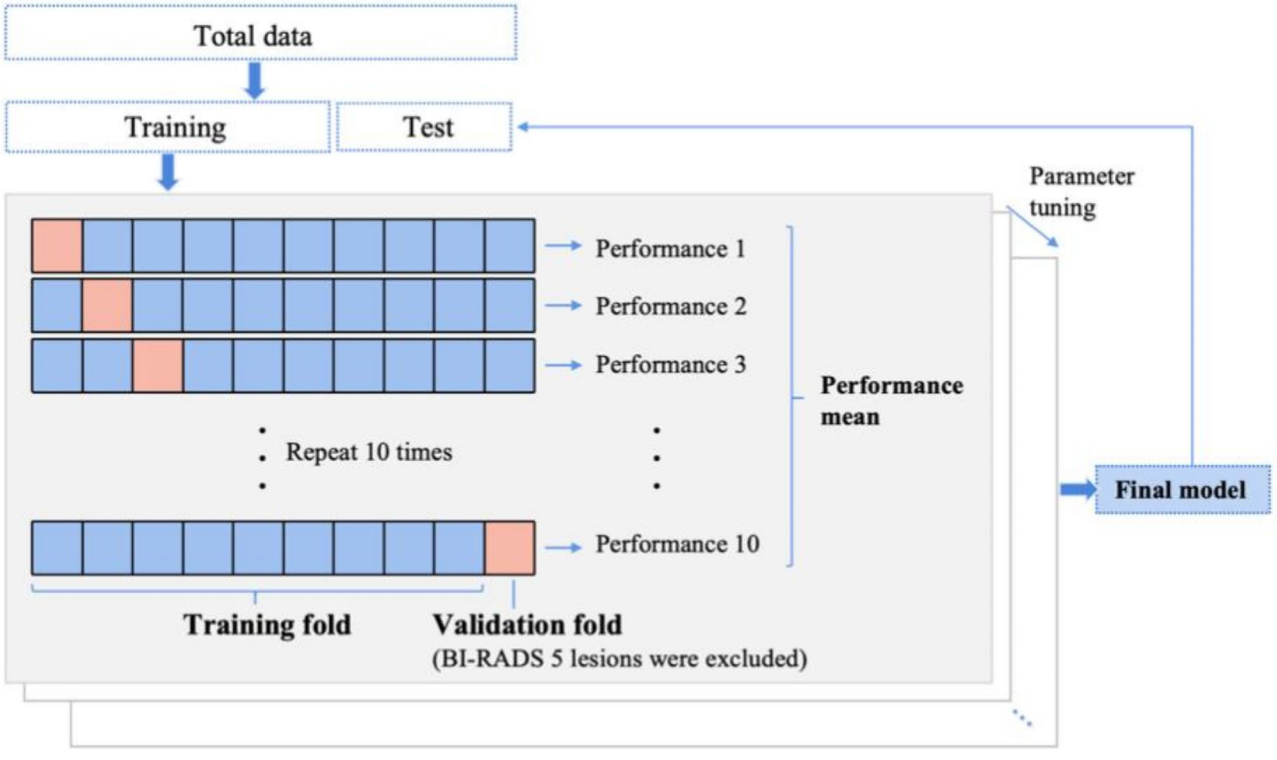

Fig. 2 Visualization of model training

which were used for training and the last of which was used for validation. This process was repeated 10 times, and the validation data differed each time. In each set, all BI-RADS 5 lesions were excluded manually during validation. Subgroup analysis was performed in the optimal model, based on two types of enhancement patterns and three subgroups of BI-RADS 4.

The Shapley (SHAP) method26 was used to facilitate the model interpretation. A total SHAP value for sample was calculated, and a higher value corresponded to a higher likelihood of the target outcome (i.e., malignant lesion). Besides, features were ranked in descending order of their SHAP values to illustrate their contribution to the prediction.

To assess the rate of avoidable biopsies, three exploratory cutoff values on the ROC curve were examined (sensitivity of C1 ≥ 98%, C2 ≥ 95% and C3 ≥ 90%). The sensitivity and specificity of the model in the test cohorts were calculated by applying the three cutoff values achieving from the training cohort.

Besides, to explore whether the discordance between the training and test cohorts can influence the model performance, another 7 models based on 7 feature sets of different ROIs were training only in the patients with BI-RADS 4 lesions from Hospital 1 (Training cohort 2). The test cohorts were the same as the models mentioned above.

Statistical analysis

The data were calculated and analysed with SPSS (software version, 22.0). All numeric data were calculated and expressed as the mean ± SD, while categorical data were expressed as the relative distribution frequency and percentage. The receiver operating characteristic (ROC) curve constructed with statistics packages in Python 3.7.0 was used to assess the classification performance by calculation of the area under the ROC curve (AUC). The corresponding accuracy, sensitivity, and specificity were also determined. The 95% confidence interval (CI) of the AUC was calculated by bootstrapping. The DeLong test was used to compare the difference in the diagnostic performance between the classification models. A *P* value < 0.05 was considered statistically significant.

Results

Patients and lesions

In total, 516 patients from Hospital 1 comprised the training cohort (mean age 46.9 ± 10.0 years), and they included 197 and 319 patients with benign and malignant disease, respectively. The internal test cohort included 146 patients (mean age 50.5 ± 11.2; 52 with benign lesions and 94 with malignant lesions). And the external test cohort included 52 patients (mean age 49.3 ± 11.1; 22 with benign lesions and 30 with malignant lesions). The pathology results are summarized in Table 1. The major malignant lesions included invasive ductal carcinoma (IDC), ductal carcinoma in situ (DCIS) and invasive lobular carcinoma (ILC). In addition, fibrocystic changes, fibroadenoma and papilloma were the three most common benign lesions. All Lesions stratified by BI-RADS subcategory and enhancement pattern are summarized in Table 2 and Supplementary Table S2.

Model construction and performance evaluation

After LASSO regression as the last step of feature selection, seven sets of features were retained as the most

Table 1 Histological diagnosis of all lesions including in the training and test cohorts

Histopathological type	Training cohort		Internal test cohort		External test cohort	
	No. of lesions	%	No. of lesions	%	No. of lesions	%
Benign	**197**		**52**		**22**	
Fibrocystic changes	100	50.8	5	9.6	7	31.8
Fibroadenoma	35	17.8	25	48.1	10	45.4
Papilloma	26	13.2	11	21.2	1	4.5
Usual ductal hyperplasia	15	7.6	2	3.8	2	9.1
Atypical ductal hyperplasia	8	4.1	2	3.8	2	9.1
Phyllodes	6	3.0	6	11.5	0	0.0
Inflammation	4	2.0	0	0.0	0	0.0
Sclerosing adenosis	2	1.0	1	1.9	0	0.0
Lobular hyperplasia	1	0.5	0	0.0	0	0.0
Malignant	**319**		**94**		**30**	
Invasive ductal carcinoma	234	73.4	62	66.0	24	80.0
Ductal carcinoma in situ	52	16.2	19	20.2	3	10.0
Invasive lobular carcinoma	15	4.7	6	6.4	0	0.0
Invasive mucinous carcinoma	11	3.4	1	1.1	2	6.7
Other*	7	2.2	6	6.4	1	3.3

*Other: Lobular carcinoma in situ (LCIS), invasive papillary cancer; malignant phyllodes, medullary carcinoma, metastases

Table 2 Overview of all lesions stratified by enhancement pattern

Shape	Training cohort			Internal test cohort			External test cohort		
	Benign ($n=197$)	Malignant($n=319$)	*p* value	Benign ($n=52$)	Malignant($n=94$)	*p* value	Benign ($n=22$)	Malignant($n=30$)	*p* value
Mass	161	278	0.121	44	69	0.179	17	21	0.789
NME	36	41		8	25		5	9	

Abbreviations: NME, non-mass enhancement

Table 3 Performance of the best models established by using the features extracted from seven different ROIs

Models	Algorithm	Cohort	AUC	95% CI	Sensitivity	Specificity	Accuracy	PPV	NPV
Intra	SVM	Training	0.898	0.867–0.928	0.793	0.821	0.732	0.798	0.817
		Internal test	0.724	0.633–0.816	0.723	0.712	0.719	0.819	0.587
		External test	0.797	0.667–0.927	0.767	0.682	0.731	0.767	0.682
Peri1	SVM	Training	0.801	0.755–0.846	0.948	0.291	0.600	0.543	0.864
		Internal test	0.793	0.719–0.867	0.968	0.250	0.712	0.700	0.812
		External test	0.815	0.696–0.934	0.933	0.227	0.635	0.622	0.714
Peri2	SVM	Training	0.910	0.881–0.939	0.885	0.760	0.819	0.766	0.882
		Internal test	0.783	0.708–0.858	0.851	0.519	0.733	0.762	0.659
		External test	0.823	0.704–0.942	0.867	0.773	0.827	0.839	0.810
Peri3	LightGBM	Training	0.817	0.775–0.860	0.649	0.791	0.724	0.734	0.718
		Internal test	0.748	0.663–0.833	0.628	0.769	0.678	0.831	0.533
		External test	0.788	0.659–0.917	0.700	0.773	0.731	0.808	0.654
Comb1	SVM	Training	0.907	0.879–0.936	0.764	0.862	0.816	0.831	0.805
		Internal test	0.737	0.651–0.822	0.670	0.712	0.685	0.808	0.544
		External test	0.835	0.715–0.955	0.967	0.545	0.788	0.744	0.923
Comb2	SVM	Training	0.916	0.886–0.946	0.862	0.801	0.830	0.794	0.867
		Internal test	0.828	0.763–0.893	0.830	0.615	0.753	0.796	0.667
		External test	0.844	0.713–0.975	0.967	0.682	0.846	0.806	0.937
Comb3	SVM	Training	0.856	0.817–0.894	0.960	0.332	0.627	0.560	0.903
		Internal test	0.788	0.712–0.865	0.968	0.173	0.685	0.679	0.750
		External test	0.812	0.696–0.928	0.967	0.227	0.654	0.630	0.833

* Intra represents the model based on features extracted from the intra-tumoral region. Peri1, Peri2 and Peri3 represent the models based on features extracted from the peri-tumoral regions of 1 mm, 2 mm and 3 mm, respectively. Comb1, Comb2 and Comb3 represent the models based on features extracted from the intra-tumoral region combined with the peri-tumoral regions of 1 mm, 2 mm and 3 mm, respectively. Abbreviations: AUC, the area under curve; CI, confidence interval; PPV, Positive predictive value; NPV, Negative predictive value; SVM, Support vector machine; GBM, Gradient Boosting Machine

relevant to the prediction of malignancy. The selected radiomic features for the model input and their regression coefficients are revealed in the Supplementary Figure S2-8.

Then, we evaluated the model performance using three different machine learning algorithms for each of the seven feature sets, as shown in Supplementary Table S3, and the algorithm with the highest AUC was chosen for the following comparison. Table 3 provides an overview of the seven classification models based on the best algorithm for the prediction of malignant lesion in each cohort, and the corresponding ROC curves are presented in Fig. 3. The models established from different ROIs were compared by DeLong test, as shown in the Supplementary Figure S9.

In the training cohort, the model built with features from the intra-tumoral region yielded an AUC of 0.898 (95% CI, 0.867–0.928), while combining intra and peri-tumoral features improved the AUCs. Among the fusion models, the Comb2 (intra + peri2mm) model exhibited the best performance, achieving an AUC of 0.916 (95% CI: 0.886–0.946), which surpassed the Comb3 model (AUC = 0.856, $p = 0.001$). There was no significant difference in performance between the Comb2 model and either the Intra model ($p = 0.127$) or the Comb1 model ($p = 0.454$). In the internal test cohort, the Comb2 model achieved an AUC of 0.828 (95% CI: 0.763–0.893), significantly outperforming the Intra model (AUC = 0.724, $p = 0.006$). In the external test cohort, the intra-only model achieved an AUC of 0.797 (95% CI, 0.667–0.927). Among the fusion models, Comb2 again showed the best performance, achieving an AUC of 0.844 (95% CI: 0.713–0.975), outperforming Comb1 (AUC: 0.835) and Comb3 (AUC: 0.812). Integrating the intra- and peri-tumoral features consistently improved the model's predictive performance across all cohorts. The fusion of features from intra-trumoral region and peri-tumoral region of 2 mm demonstrated superior AUC values compared to other fusion models, indicating that the 2 mm peri-tumoral

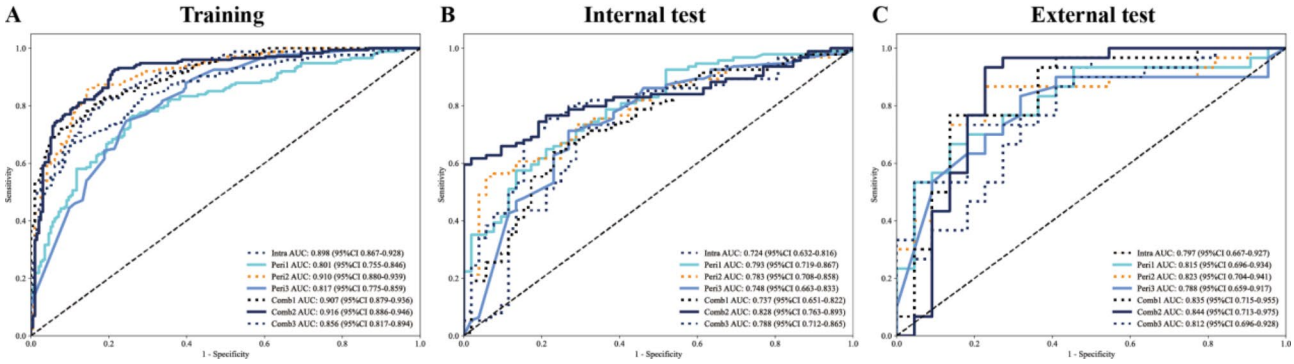

Fig. 3 ROC curves of the optimal models established by the features extracted from seven different ROIs in the training cohort (**A**), internal test cohort (**B**) and external test cohort (**C**)

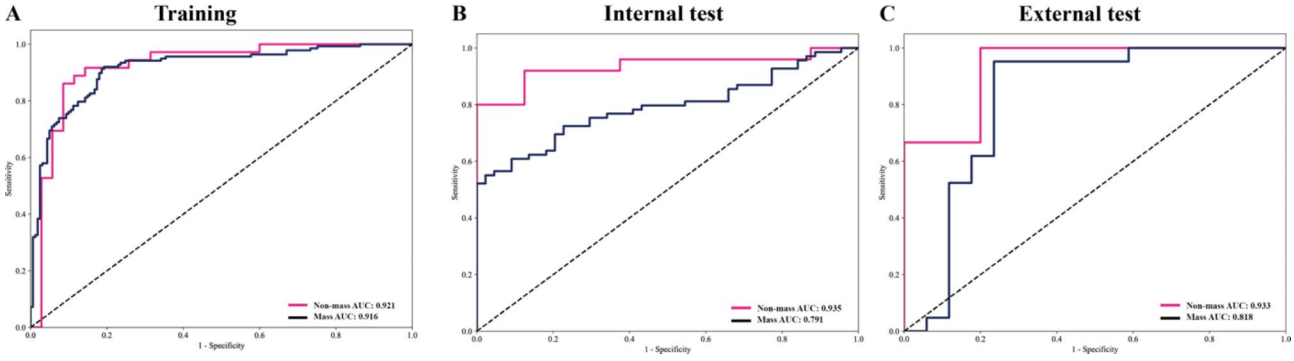

Fig. 4 ROC Curves of the optimal model for lesion discrimination in mass and NME lesions in the training cohort (**A**), internal test cohort (**B**) and external test cohort (**C**)

region provides the optimal spatial context for the prediction of malignant lesion.

Only patients with BI-RADS 4 lesions (196 with benign lesions and 174 with malignant lesions) from Hospital 1 were used as a new training cohort (Training cohort 2). The test cohorts were the same as the origin work. The performance of 7 models based on different ROIs were shown in Supplementary Table S4, Figure S10. Those new models trained only in BI-RADS 4 lesions did not demonstrate superiority in all the cohorts, compared to the previous models. Therefore, the Comb2 model was regarded as the optimal model and used for further analysis.

ROC analysis was applied for only mass lesions and only NME lesions separately using the Comb2 model as shown in Fig. 4. The results of only NME lesions (AUCs = 0.921, 0.935 and 0.933 in the training, internal and external test cohort, respectively) were no worse than that of only mass lesions (AUCs = 0.916, 0.791 and 0.818 in the training, internal and external test cohort, respectively) for distinguishing malignant and benign breast lesions. Besides, the performance of Comb2 model in three subcategories of BI-RADS 4 lesions was also evaluated (Supplementary Figure S11). The Comb2 model exhibited stable performances, especially in BI-RADS 4a.

Analysis of the most predictive features

SHAP analysis was used to identify the predominant factors influencing the final output of the Comb2 model, as shown in Fig. 5. The SHAP bar plot (Fig. 5A) was generated by calculating the SHAP values for each feature, thereby illustrating the impact of feature on the prediction outcome of a specific sample. The SHAP bees-warm plot (Fig. 5B) displays the impact of top features on the model's output. For each feature, dots to the right are related to a higher risk of malignancy, while they have a negative impact if they are positioned on the right. The color of the dots indicates the magnitude of the feature value, which transitions from blue (low values) to red (high values). Among the five most important features, the firstorder_Skewness, gldm_DependenceVariance, glszm_SmallAreaEmphasis of the intra-tumoral area and the glszm_ LowGrayLevelZoneEmphasis of the peritumoral region of 2 mm were related to malignancy, for the red dots are predominantly on the right side. Conversely, the high value of glszm_Zone% in the intratumoral region had the opposite effect.

Estimation of the potential to avoid unnecessary biopsies

The summary of sensitivity and specificity values at different cutoff values in the training cohort and test cohort

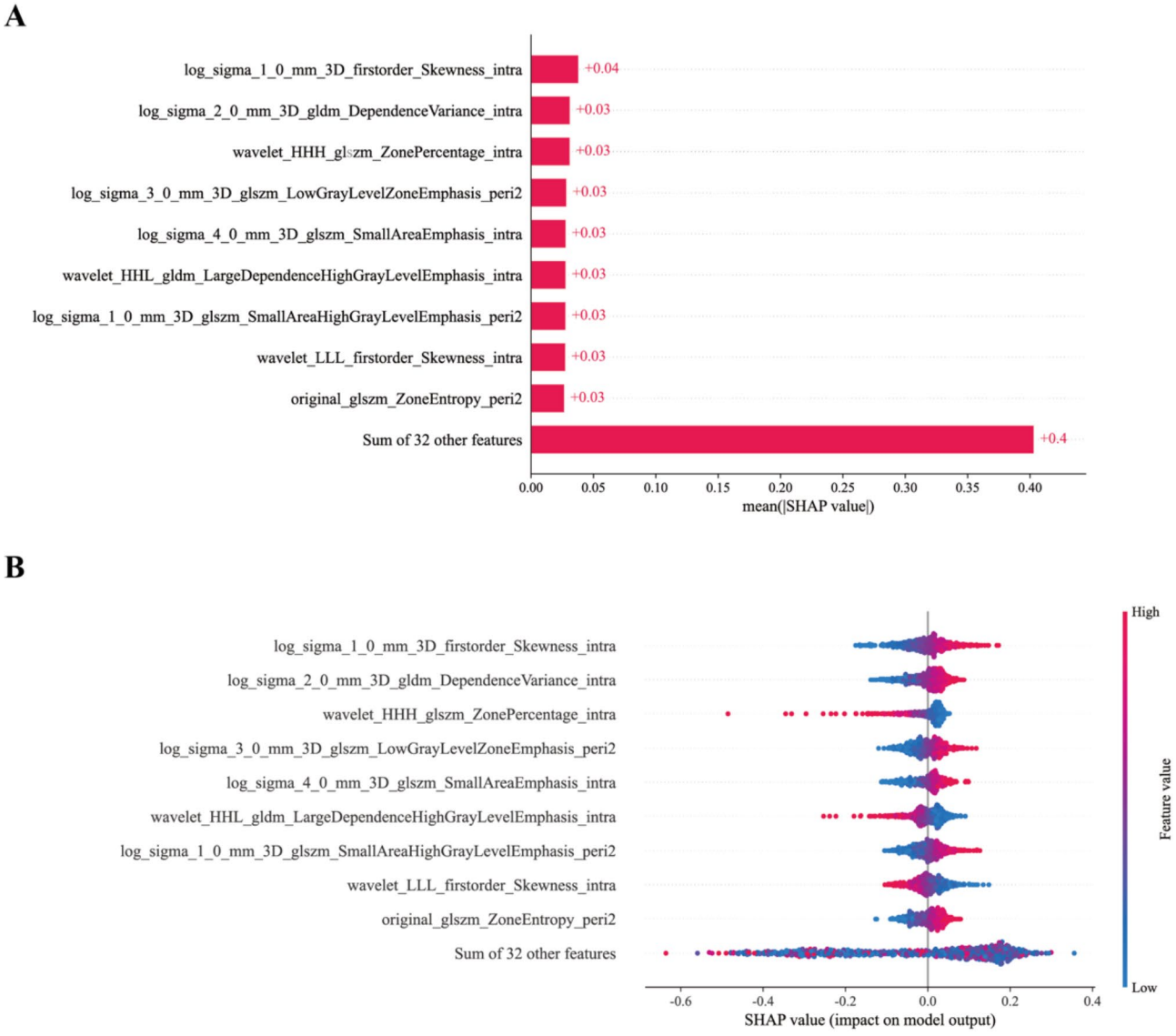

Fig. 5 The Comb2 model utilizing the SHAP algorithm. (**A**) The ranking of feature importance for all samples. (**B**) The SHAP value of each sample. Each dot represented a sample. As the color becomes redder, the feature's value increases, while a bluer color indicates a lower value

Table 4 Performance of the Comb2 model at 3 operating points

Target Sensitivity	Training cohort		Internal test cohort		External test cohort	
	Actual sensitivity	Specificity	Actual sensitivity	Specificity	Actual sensitivity	Specificity
98%	98.3%	29.6% (58/196)	96.8%	1.9% (1/52)	100.0%	9.1% (2/22)
95%	95.4%	32.1% (63/196)	93.6%	9.6% (5/52)	90.0%	27.3% (6/22)
90%	90.2%	79.6% (156/196)	70.2%	23.1% (12/52)	80.0%	36.4% (8/22)

are displayed in Table 4. At the cutoff values of C1, C2 and C3, the model identified 29.6% (58/196), 32.1% (63/196) and 79.6% (156/196) of the benign lesions, respectively, in the training cohort, showed the potential to lower the number of unnecessary biopsies of benign lesions. At the cutoff of C1, 58 patients in the training cohort would be exempted from biopsy, with only 3 malignant lesions were missed. Applying the identified cutoff values in the external test cohort, the model identified 9.1% (sensitivity of C1 ≥ 98%), 27.3% (sensitivity of C2 ≥ 95%) and 36.4% (sensitivity of C3 ≥ 90%) of the benign lesions, showed the potential to lower the number of unnecessary biopsies of benign lesions.

Discussion

In this multicenter study, we investigated the ability of DCE-MRI based radiomics models to distinguish benign from malignant breast lesions stratified as BI-RADS 4, including features extracted from intra- and peri-tumoral regions. Our findings showed that the peri-tumoral regions have additional predictive values relative to the intra-tumoral regions, and the model built with features extracted from the intra- tumoral region combined with peri-tumoral region showed the best performance to reduce false-positive diagnoses with a low risk of missing cancer.

There have been several radiomics models based on breast MRI for distinguishing between benign and malignant disease, and the performance of those models was moderate (AUC ranged between 0.790 and 0.960) [13–18]. However, these results might not be directly comparable to our results because they were mainly based on single institution data. In addition, to make our model closer to clinical reality, the BI-RADS 5 lesions were excluded in the performance evaluation. The rationale for this omission is that the problem of false-positives seldom occurs in BI-RADS 5, for BI-RADS 5 lesions have typical signs of malignancy, and more than 95% of them are malignant [8]. Compared to previous work, our model is tested on the most difficult and clinically relevant cases

This study represents the first attempt to systematically explore the optimal size of peri-tumoral region in the diagnosis of BI-RADS 4 lesions and was validated in an external test cohort. Previous studies have primarily focused on the parenchyma of lesions, and only features extracted from the tumor region were utilized in radiomics analysis. Peri-tumoral region was the tissue surrounding the tumor and different from normal tissue due to the tumor invasion, which provides additional information about tumor pathophysiology as reported [27–29]. Niu et al. found that the features from the peri-tumoral regions at 2 mm dilation distances in the DBT image showed the best discriminative performance in diagnosing breast cancer [20]. In our study, the models built with features from the combined regions achieved better performance than models built with features from the tumor area alone, and peri-tumoral range of 2 mm is the optimal, consistent with previous work. Radiomics features extracted from the proximal peri-tumoral stroma have higher predictive power than the other regions. In the Comb2 model, log_sigma_1_0_mm_3D_firstorder_Skewness, log_sigma_2_0_mm_3D_gldm_DependenceVariance, log_sigma_4_0_mm_3D_glszm_SmallAreaEmphasis and wavelet_HHH_glszm_Zone% extracted from the intra-tumoral region and log_sigma_3_0_mm_3D_glszm_ LowGrayLevelZoneEmphasis extracted from the peri-tumoral 2 mm significantly contributed to the diagnosis of malignancy. Those transform-based features were extracted from filtered images using wavelet filters and LoG filters. Considering the first order feature reflects the gray value distribution within the ROI, the high value of firstorder_Skewness revealed that malignant lesions exhibited greater asymmetry with respect to gray value distributions about the mean as compared with benign lesions. Besides, the high-order features of GLDM and GLSZM could quantify the spatial relationships and interactions between pixel intensities, which may help to capture the distinctive heterogeneity of intra- and peri-tumoural regions.

BI-RADS 4a is generally regarded as a low-risk disease, with a large part of the pathology result after biopsy is confirmed to be benign, leading to an unnecessary invasive examination [30]. By further analysis of BI-RADS 4a lesion, clinicians can define whether the patients are suitable for short-term follow-up or are recommended to biopsy. A multicenter study by Niu et al. incorporated clinical factors and imaging characteristics of ultrasound for the classification of BI-RADS 4a, yielding the AUC of 0.782-0.747 [31]. Rong et al. utilized the BI-RADS lexicon from contrast-enhanced mammography in the lesions previously DBT classified as BI-RADS 4a, achieving AUCs between 0.880 and 0.906 [32]. Specially, after stratified by the subcategories of BI-RADS 4, the Comb2 model in our study showed stable distinguish ability in the subgroup of BI-RADS 4a, with the AUCs of 0.824 and 0.917 in the internal and external test cohort, respectively.

Breast lesions on MRI can be divided into two categories according to enhancement pattern, i.e., mass, and NME. Compared to mass lesions, detection and diagnosis of NME have been known as a more challenging problem [33]. The common histopathology that may manifest as NME includes benign conditions such as fibrocystic or proliferative changes, and malignant lesions such as DCIS and ILC [34]. Most prior studies have been focused on the diagnosis of mass lesions, primarily because of the challenges in determining the tumor border of NME. To mimic clinical practice, both mass and NME lesions were included in our study. Comparable performance was observed between mass and NME lesions, indicating that the model was robust.

In our study, 53.0% (196/370) in the training cohort, 35.6% (52/146) in the internal test cohort and 42.3% (22/52) in the external test cohort of BI-RADS 4 lesions were pathologically confirmed to be benign, indicating that more than half of the patients received unnecessary biopsies. Although AUC is a routine statistic to evaluate model performance, the specificity on high sensitivity is also important in clinical management, which represents the number of biopsies that would be avoided with a low

risk of missing cancer. At the cutoff of C1, 58 patients with benign lesions would have been not recommended for biopsy in the training cohort, with a satisfactory sensitivity of 98.3%. The three missed cases were all DCIS. Applied the cutoff of C1 to the internal and external test cohort, the sensitivity was 96.8% and 100%, respectively. Since this was a retrospective and multicenter study, the slight discordance of sensitivity at the same cutoff value may have been caused by the differences in scanning equipment and indication of MRI examination. The false negative cases in the internal test cohort were DCIS (2/3) and ILC (1/3). Growing evidence has suggested that up to 80% of DCIS cases may be indolent, which rarely progress to invasive cancer [35, 36]. MRI was reported to be less sensitive in the diagnosis of DCIS than invasive cancer, partly because a substantial proportion of DCIS cases do not exhibit the typical enhancement observed in invasive cancer [37, 38]. Radiomics features from MRI do provide additional information for the non-invasive diagnosis of breast cancer.

There still exist some limitations in our study. First, the retrospective nature and comparatively low proportion of benign lesions limited the ability to perform stratified analyses. Second, to ensure a confirmed pathological result, this study only included patients who received breast surgery or core needle biopsy. The patients who were unwilling to undergo the operation or those who were in a follow-up process were ruled out, creating a possible source of selection bias. Third, we only explored the radiomics features extracted from DCE-MRI, and more information might be obtained from other sequences and other image modalities to explore the diagnosis ability of multi-omics model.

Conclusion

The study demonstrated that the MRI-based radiomics model built with features extracted from the combined intra-tumoral and peri-tumoral 2 mm regions showed the best potential to reduce false-positive diagnoses and, consequently, avoid unnecessary biopsy with a low risk of missing cancer.

Abbreviations

DCE-MRI	Dynamic contrast-enhancement magnetic resonance imaging
BI-RADS	Breast imaging reporting and data system
ROC	Receiver operation characteristic curve
AUC	Area under the receiver operating characteristic curve
NME	Non-mass enhancement
DBT	Digital breast tomosynthesis
ROI	Region of interest
GLCM	Gray level co-occurrence matrix
GLSZM	Gray level size zone matrix
GLRLM	Gray level run length matrix
NGTDM	Neighborhood gray-tone difference matrix
NGLDM	Neighboring gray level dependence matrix
LASSO	Least absolute shrinkage and selection operator
RF	Random forest
mRMR	Minimum redundancy maximum relevance
SVM	Support vector machine
GBM	Gradient boosting machine
SHAP	Shapley
CI	Confidence interval
SD	Standard deviation
IDC	Invasive ductal carcinoma
DCIS	Ductal carcinoma in situ
ILC	Invasive lobular carcinoma

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13014-025-02605-y.

Supplementary Material-1

Author contribution

YH, ZC, and NA designed and implemented the study. YL and XZ supervised the study. YH, ZC and NA completed the data collection and management. XZ and YH delineated the region of interest of MR images. YL, NS, YS and MZ performed the statistical analyses. YH drafted the preliminary manuscript and all co-authors revised the final manuscript.

Funding

This project was funded by the Sun Yat-Sen University Clinical Research 5010 Program (2016007) and Guangdong Health Management Association (2023001).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethical approval and consent to participate

This study followed all dictates of the Declaration of Helsinki and the Ethics Review Board of the First Affiliated Hospital of Sun Yat-sen University. The written consent to participate was waived due to the retrospective nature of this study.

Consent for publication

All authors approved the final manuscript and the submission to this journal.

Competing interests

The authors declare no competing interests.

Author details

[1]Breast Disease Center, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China
[2]Department of Medical Ultrasonics, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China
[3]Department of Breast Surgery, Jieyang People's Hospital, Jieyang, China
[4]Department of Breast and Thyroid Surgery, The Seventh Affiliated Hospital, Sun Yat-sen University, ShenZhen, China
[5]Department of Radiology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

Received: 14 July 2024 / Accepted: 17 February 2025
Published online: 28 February 2025

References

1. Sung H, Ferlay J, Siegel RL et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians 2021.
2. Amornsiripanitch N, Bickelhaupt S, Shin HJ, et al. Diffusion-weighted MRI for unenhanced breast Cancer screening. Radiology. 2019;293(3):504–20.
3. Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. JAMA. 2012;307(13):1394–404.

4. Bakker MF, de Lange SV, Pijnappel RM, et al. Supplemental MRI screening for women with extremely dense breast tissue. N Engl J Med. 2019;381(22):2091–102.
5. Comstock CE, Gatsonis C, Newstead GM, et al. Comparison of abbreviated breast MRI vs digital breast tomosynthesis for breast Cancer detection among women with dense breasts undergoing screening. JAMA. 2020;323(8):746–56.
6. Sardanelli F, Boetes C, Borisch B, et al. Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group. Eur J Cancer. 2010;46(8):1296–316.
7. Strobel K, Schrading S, Hansen NL, Barabasch A, Kuhl CK. Assessment of BI-RADS category 4 lesions detected with screening mammography and screening US: utility of MR imaging. Radiology. 2015;274(2):343–51.
8. D'Orsi C, Morris E, Mendelson E. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. 2013.
9. Gity M, Arabkheradmand A, Taheri E, Shakiba M. Diagnostic investigation of breast magnetic resonance imaging in malignant and benign mass lesions. Archives Med Science: AMS. 2018;14(5):1061–9.
10. Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. Radiology. 2021;299(2):E256.
11. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the Bridge between medical imaging and personalized medicine. Nat Reviews Clin Oncol. 2017;14(12):749–62.
12. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. Cancer J Clin. 2019;69(2):127–57.
13. Zhang R, Wei W, Li R, et al. An MRI-Based radiomics model for predicting the benignity and malignancy of BI-RADS 4 breast lesions. Front Oncol. 2021;11:733260.
14. Zhang B, Song L, Yin J. Texture analysis of DCE-MRI intratumoral subregions to identify benign and malignant breast tumors. Front Oncol. 2021;11:688182.
15. Perre SV, Duron L, Milon A, et al. Radiomic analysis of HTR-DCE MR sequences improves diagnostic performance compared to BI-RADS analysis of breast MR lesions. Eur Radiol. 2021;31(7):4848–59.
16. Pötsch N, Dietzel M, Kapetas P, et al. An A.I. Classifier derived from 4D radiomics of dynamic contrast-enhanced breast MRI data: potential to avoid unnecessary breast biopsies. Eur Radiol. 2021;31(8):5866–76.
17. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus convolutional neural networks analysis for classification of Contrast-enhancing lesions at multiparametric breast MRI. Radiology. 2019;290(2):290–7.
18. Cui Q, Sun L, Zhang Y, et al. Value of breast MRI omics features and clinical characteristics in breast imaging reporting and data system (BI-RADS) category 4 breast lesions: an analysis of radiomics-based diagnosis. Ann Transl Med. 2021;9(22):1677.
19. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19(11):1423–37.
20. 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 (Ankara Turkey). 2022;28(3):217–25.
21. Niu S, Jiang W, Zhao N, et al. Intra- and peritumoral radiomics on assessment of breast cancer molecular subtypes based on mammography and MRI. J Cancer Res Clin Oncol. 2022;148(1):97–106.
22. Braman NM, Etesami M, Prasanna P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res. 2017;19(1):57.
23. Li C, Song L, Yin J. Intratumoral and peritumoral radiomics based on functional parametric maps from breast DCE-MRI for prediction of HER-2 and Ki-67 status. J Magn Reson Imaging: JMRI. 2021;54(3):703–14.
24. Ming W, Li F, Zhu Y, et al. Predicting hormone receptors and PAM50 subtypes of breast cancer from multi-scale lesion images of DCE-MRI with transfer learning technique. Comput Biol Med. 2022;150:106147.
25. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29(6):1310–20.
26. Lundberg SM, Lee S-IJA. A Unified Approach to Interpreting Model Predictions. 2017; abs/1705.07874.
27. Mohammed ZM, McMillan DC, Edwards J, et al. The relationship between lymphovascular invasion and angiogenesis, hormone receptors, cell proliferation and survival in patients with primary operable invasive ductal breast cancer. BMC Clin Pathol. 2013;13(1):31.
28. Freed M, Storey P, Lewin AA, et al. Evaluation of breast lipid composition in patients with benign tissue and Cancer by using multiple Gradient-Echo MR imaging. Radiology. 2016;281(1):43–53.
29. Cheon H, Kim HJ, Kim TH, et al. Invasive breast cancer: prognostic value of peritumoral edema identified at preoperative MR imaging. Radiology. 2018;287(1):68–75.
30. Flowers CI, O'Donoghue C, Moore D, et al. Reducing false-positive biopsies: a pilot study to reduce benign biopsy rates for BI-RADS 4A/B assessments through testing risk stratification and new thresholds for intervention. Breast Cancer Res Treat. 2013;139(3):769–77.
31. Niu Z, Tian JW, Ran HT, et al. Risk-predicted dual nomograms consisting of clinical and ultrasound factors for downgrading BI-RADS category 4a breast lesions - A multiple centre study. J Cancer. 2021;12(1):292–304.
32. Rong X, Kang Y, Xue J, et al. Value of contrast-enhanced mammography combined with the Kaiser score for clinical decision-making regarding tomosynthesis BI-RADS 4A lesions. Eur Radiol. 2022;32(11):7439–47.
33. Meyer-Base A, Morra L, Tahmassebi A, Lobbes M, Meyer-Base U, Pinker K. AI-Enhanced diagnosis of challenging lesions in breast MRI: A methodology and application primer. J Magn Reson Imaging: JMRI. 2021;54(3):686–702.
34. Liu C, Liang C, Liu Z, Zhang S, Huang B. Intravoxel incoherent motion (IVIM) in evaluation of breast lesions: comparison with conventional DWI. Eur J Radiol. 2013;82(12):e782–9.
35. Ryser MD, Weaver DL, Zhao F, et al. Cancer outcomes in DCIS patients without locoregional treatment. J Natl Cancer Inst. 2019;111(9):952–60.
36. van Seijen M, Lips EH, Thompson AM, et al. Ductal carcinoma in situ: to treat or not to treat, that is the question. Br J Cancer. 2019;121(4):285–92.
37. Schnall MD, Blume J, Bluemke DA, et al. Diagnostic architectural and dynamic features at breast MR imaging: multicenter study. Radiology. 2006;238(1):42–53.
38. Kuhl CK, Schrading S, Bieling HB, et al. MRI for diagnosis of pure ductal carcinoma in situ: a prospective observational study. Lancet. 2007;370(9586):485–92.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.