International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

A Multimodal Ultrasound Radiomics Study for Discriminating Benign and Malignant Breast Lesions Classified as BI-RADS Category 4

Dongxu Li[1], Shanshan Chang[1], Chongyang Sun[2], Feng Lin[3], Ze Zhang[3], Qi Chen[1]

[1]Department of Ultrasound, First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, 150001, People's Republic of China; [2]Department of CT, First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, 150001, People's Republic of China; [3]Department of Surgery, First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, 150001, People's Republic of China

Correspondence: Qi Chen, Department of Ultrasound, First Affiliated Hospital of Harbin Medical University, 23 You Zheng Street, Harbin, Heilongjiang, 150001, People's Republic of China, Email 814440@hrbmu.edu.cn

Purpose: This study evaluates multimodal radiomics feature fusion integrating intratumoral and peritumoral radiomics derived from B-mode ultrasound (US) and Virtual Touch Tissue Imaging and Quantification (VTIQ) to differentiate benign from malignant BI-RADS category 4 breast lesions.

Patients and Methods: A retrospective analysis was conducted on 264 patients. Radiomic features were extracted from intratumoral and peritumoral regions (1 mm, 3 mm, 5 mm) and VTIQ elastography. Three fusion strategies were compared: (i) B-mode US + peritumoral features, (ii) B-mode US + VTIQ, and (iii) triple fusion (B-mode US + peritumoral + VTIQ). Features were selected via LASSO regression, and models (LR, SVM, RF) were developed and validated on separate training (n = 184) and test (n = 80) cohorts.

Results: The B-mode US model achieved area under the curve (AUC) values of 0.890 (training) and 0.844 (test), surpassing the VTIQ model (training AUC = 0.850, test AUC = 0.812). Fusion of B-mode US and VTIQ features improved performance to AUCs of 0.913 and 0.874, respectively. Among peritumoral margins, the 3 mm region provided the best discrimination, with AUCs of 0.893 (training) and 0.871 (test). Integration of B-mode US with peritumoral features further improved performance (AUC = 0.908 training, 0.859 test). The triple-fusion model (B-mode US + peritumoral + VTIQ) achieved the highest diagnostic accuracy, with AUCs of 0.933 in the training cohort and 0.903 in the test cohort.

Conclusion: Multimodal feature fusion significantly enhances the differentiation of benign and malignant breast lesions, supporting its potential for clinical decision-making.

Keywords: breast lesions, ultrasound, VTIQ, radiomics

Introduction

Breast cancer remains the most frequently diagnosed malignancy and a leading cause of cancer-related mortality among women globally, with an estimated 2.3 million new cases reported in 2022.[1] The mortality rate accounts for nearly one-quarter of its incidence, emphasizing the urgent need for early detection, accurate diagnosis, and timely treatment in reducing breast cancer burden. Ultrasound imaging is a key modality in breast cancer screening and diagnosis.[2] The Breast Imaging Reporting and Data System (BI-RADS) is widely applied for classifying breast lesions. Although B-mode ultrasound morphological features achieve high sensitivity (95–97%) in distinguishing benign from malignant lesions, the specificity remains relatively low (55–68%),[3] resulting in a considerable number of unnecessary biopsies. Notably, BI-RADS category 4 lesions encompass a broad malignancy risk range (2–95%), making precise differentiation within this category clinically critical to prevent avoidable invasive procedures in benign cases.

Malignant tumors are characterized by pathological features such as high cellular density, irregular cellular arrangement, and stromal fibrosis, which contribute to increased tissue stiffness. Shear wave elastography (SWE) enables quantitative assessment of tissue stiffness and has demonstrated substantial diagnostic value in differentiating benign

https://doi.org/10.2147/IJGM.S579165
Received: 27 November 2025
Accepted: 16 March 2026
Published: 23 March 2026

from malignant breast lesions.[4] As a specialized form of SWE, Virtual Touch Tissue Imaging and Quantification (VTIQ) enhances conventional ultrasound by improving the specificity of breast mass evaluation.[5]

Radiomics allows high-throughput extraction of quantitative features from medical images, capturing subtle information beyond human visual perception.[6] These features are highly valuable for distinguishing benign from malignant tumors. However, most existing studies emphasize single-modality feature analysis and overlook the integration of multi-modal imaging data. Thus, comprehensive exploration of radiomic characteristics through multi-modal feature integration may offer novel strategies and insights for improving clinical breast cancer diagnosis.

Integrating multi-parametric ultrasound data (B-mode US and VTIQ) with machine learning algorithms provides a novel technical strategy for enhancing diagnostic accuracy. Traditional radiomics studies have primarily focused on intratumoral features, often neglecting the diagnostic value of the peritumoral region.[7,8] Notably, the peritumoral area contains key biological information associated with tumor progression and metastasis, including stromal reactions, angiogenesis, and lymphatic or vascular invasion.[9] Previous studies have demonstrated that peritumoral features play a significant role in distinguishing benign from malignant tumors.[10]

This retrospective study analyzed imaging data from 264 female patients with surgically confirmed breast tumors. Radiomics models were developed using machine learning algorithms to extract multidimensional tumor features. The study systematically assessed the diagnostic performance of different imaging features in differentiating benign from malignant breast lesions, with particular emphasis on the role of multi-modal image fusion in enhancing the accuracy of BI-RADS category 4 classifications. The results are anticipated to provide new theoretical insights and technical support for advancing early precision diagnosis and personalized management of breast cancer.

Materials and Methods

Study Subject

This retrospective study included female patients with surgically resected breast tumors treated at the First Affiliated Hospital of Harbin Medical University between March 2023 and February 2025. Inclusion criteria were: (1) histopathological confirmation of a breast lesion; (2) completion of both B-mode ultrasound and VTIQ examinations within two weeks prior to surgery; (3) BI-RADS category 4 classification; and (4) age $\geq$ 18 years. Exclusion criteria included: (1) pregnancy or lactation; (2) psychiatric or cognitive disorders affecting informed consent or protocol compliance; (3) missing pathological information or prior radiotherapy/chemotherapy; and (4) poor-quality ultrasound images or lesions > 40 mm, which prevent accurate ROI delineation. For patients with multiple lesions, only the largest pathologically confirmed lesion was analyzed. The final dataset consisted of 264 patients with corresponding tumors. This study is a retrospective analysis. All patient imaging and clinical data used were derived from data collected during previous routine diagnostic and therapeutic procedures, and no additional interventions were implemented for the patients during the study period. The Ethics Committee of the First Affiliated Hospital of Harbin Medical University has reviewed and approved the exemption of written informed consent for this study (Ethics Approval Number: 2025JS78). Throughout the study, all patient data were anonymized, used solely for scientific research, strictly adhered to the principles of patient privacy protection, and complied with the relevant requirements of the Declaration of Helsinki. All subjects enrolled in this study were breast lesions that ultimately underwent surgical resection with complete postoperative pathological confirmation. This design was intended to adopt surgical pathology as a unified and reliable endpoint standard, thereby minimizing the potential confounding effect of outcome uncertainty on imaging feature analysis and model development. Both ultrasound and VTIQ examinations employed in the study were derived from routine clinical workflows and utilized for further qualitative assessment of suspicious lesions categorized as BI-RADS 4 during initial imaging evaluation. The timing of these examinations aligns with real-world clinical imaging practice.

Image Acquisition

All examinations were performed using an ACUSON S3000 ultrasound system (Siemens Medical Solutions, Mountain View, CA, USA) equipped with 4–9 MHz and 6–18 MHz linear array transducers. The lower-frequency probe was used for integrated real-time elastography. Patients were positioned supine with arms abducted 90° to ensure optimal visualization of the breast and axillary regions. A systematic scan was conducted from the nipple outward, and once a lesion was identified,

a detailed assessment of the region of interest was performed. The image displaying the largest tumor cross-section was selected to record lesion size and location, and BI-RADS classification was assigned according to standard criteria. After conventional ultrasound, VTIQ elastography was carried out: patients were instructed to breathe quietly, the VTIQ function was activated, the lesion was centered within the sampling box, and box dimensions were adjusted to enhance image stability. A velocity-mode image was then obtained using the "Update" command.

Image Processing and Analysis

Segmentation of all breast lesion images was performed manually along the ROI on the largest cross-sectional view using ITK-SNAP (v3.8.0, http://www.itksnap.org), consistently adhering to the lesion margins. To evaluate inter-observer variability, 30 lesions randomly selected from the training set were independently segmented by two experienced breast radiologists, each with more than five years of expertise in breast imaging. Inter-rater reliability was quantified using the intraclass correlation coefficient (ICC), with values >0.9 considered indicative of excellent agreement. All images were uniformly resampled to a resolution of 1×1 mm. Intratumoral ROIs were traced along the tumor margin. Peritumoral regions extending 0–1 mm, 0–3 mm, and 0–5 mm from the tumor boundary were automatically generated by the software, following the methodology described by Wang et al[11] This fixed-distance expansion strategy is consistent with commonly used methods in previous radiomics studies, demonstrating biological rationality and methodological consistency. The delineation process is illustrated in Figures 1 and 2.

Radiomics Feature Extraction

Radiomic features were extracted using the open-source PyRadiomics package (v3.7.12) integrated into the OnekeyAI platform. Features were derived from both original and filtered images, encompassing first-order statistics, shape descriptors, and texture-based matrices. Feature stability was first evaluated through variance analysis, and only those with an ICC >0.9 and statistical significance were retained. Further dimensionality reduction and selection were conducted using least absolute shrinkage and selection operator (LASSO) regression with ten-fold cross-validation, ensuring that only features with non-zero coefficients were incorporated into the development of the radiomic signature.

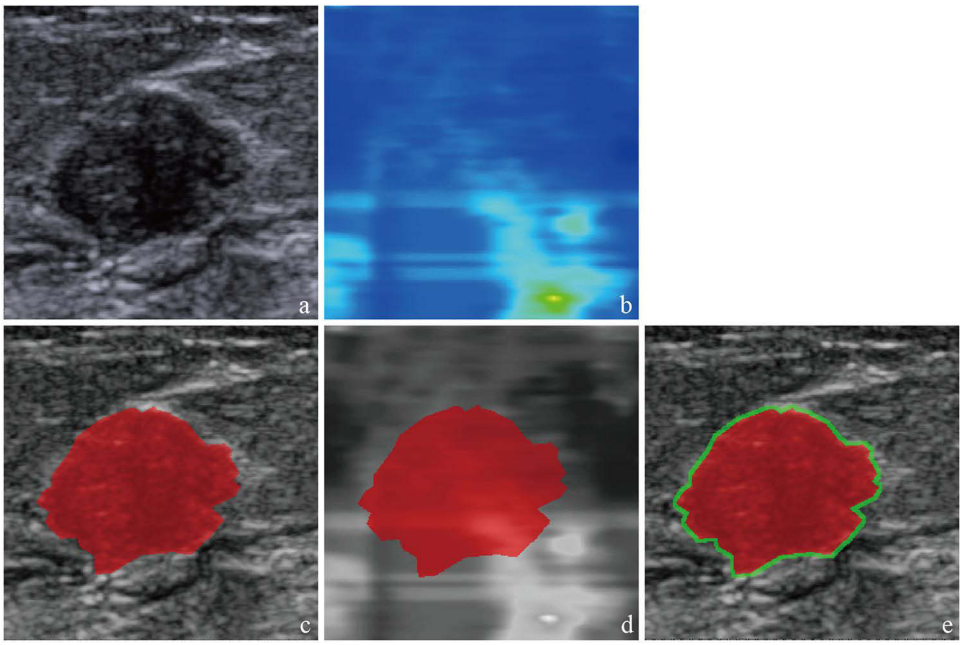

Figure 1 Representative example of manual segmentation of a benign breast lesion on B-mode images (**a** and **c**) and VTIQ images (**b** and **d**), with delineation of the peritumoral ROI (**e**).
Note: red region: tumor ROI; green borders: 3mm peritumoral region.

https://doi.org/10.2147/IJGM.S579165

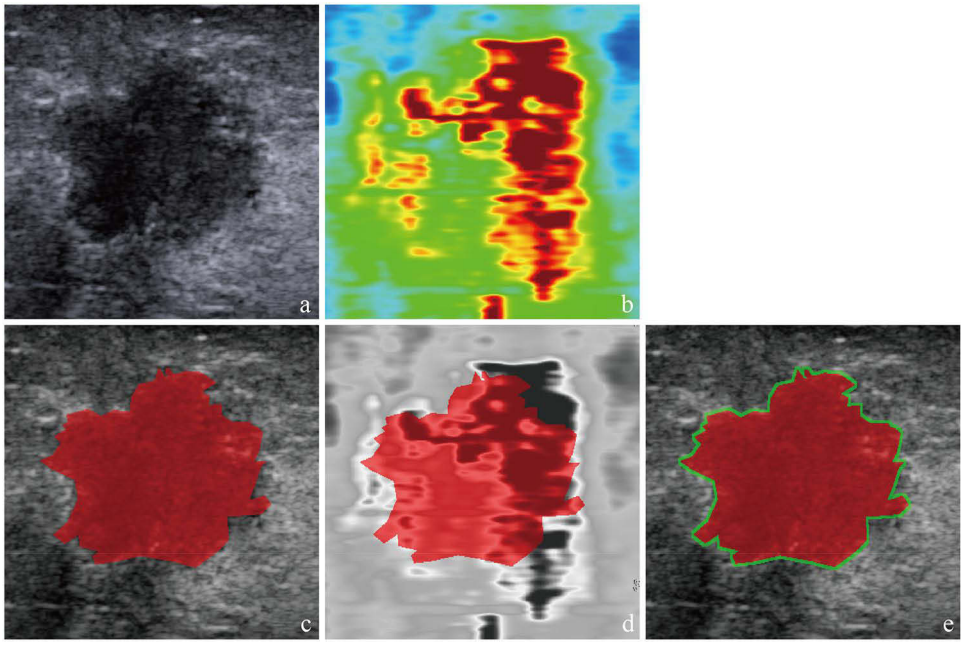

Figure 2 Representative example of manual segmentation of a malignant breast lesion on B-mode images (**a** and **c**) and VTIQ images (**b** and **d**), with delineation of peritumoral ROI (**e**).
Note: red region: tumor ROI; green borders: 3mm peritumoral region.

Model Development

After LASSO-based feature selection, predictive models were developed using several machine learning algorithms, including logistic regression, support vector machines, and random forest classifiers. Model performance was assessed by calculating the area under the receiver operating characteristic curve (AUC), along with accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). We reported the AUC with 95% confidence intervals. Furthermore, multi-modal feature fusion strategies were investigated to enhance predictive performance by integrating complementary imaging information.

Statistical Methods

Normality of continuous variables was assessed using the Shapiro–Wilk test. For group comparisons, normally distributed variables were analyzed with independent *t*-tests, while non-normally distributed variables were evaluated using the Mann–Whitney *U*-test. Categorical variables were compared with chi-square (χ^2) or Fisher's exact tests, as appropriate. Statistical significance was set at $p < 0.05$ to minimize potential bias in group comparisons. All analyses were performed using Python 3.7.12 on the OnekeyAI platform (v4.9.1). Statistical tests were conducted with Statsmodels v0.13.2, while radiomic feature extraction was accomplished with PyRadiomics v3.0.1. Machine learning models were implemented using Scikit-learn v1.0.2. Model discrimination was assessed using the AUC, with 95% confidence intervals estimated via the DeLong method. Pairwise comparisons of AUCs were subsequently conducted using the DeLong test.

Results

Patient Enrollment and Pathological Characteristics

The patient selection process is illustrated in Figure 3. A total of 407 breast lesions from 407 patients initially met the inclusion criteria. Following application of exclusion criteria—lactation or pregnancy (n = 6), unclear pathological diagnosis (n = 19), prior radiotherapy or chemotherapy (n = 12), poor B-mode image quality (n = 27), inadequate VTIQ elastography (n = 43), and lesion size > 40 mm (n = 36)—264 lesions from 264 patients were ultimately included in the analysis. Pathological evaluation identified 111 benign lesions, comprising fibroadenoma (n = 56), adenosis (n = 21), papilloma (n = 13), phyllodes tumor (n = 15), and granulomatous inflammation (n = 6). The remaining 153 lesions were malignant, including invasive ductal carcinoma (n = 132), encapsulated papillary carcinoma (n = 6), mucinous carcinoma

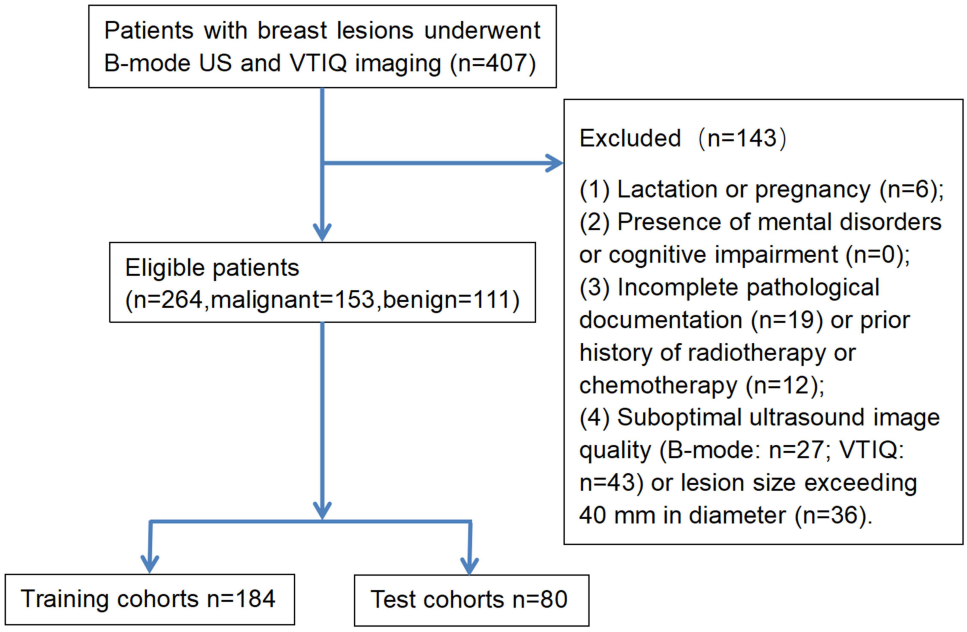

Figure 3 The flowchart showing the enrollment pathway in this study and the distribution of patients in the training and test cohorts.

(n = 9), and ductal carcinoma in situ (n = 6). Patient ages ranged from 16 to 85 years (mean: 49 years), and nodule sizes varied from 0.43 to 4 cm (mean diameter: 2.12 cm).

All cases were randomly divided into a training set (184 nodules, 70%) and a validation set (80 nodules, 30%). Baseline clinical and pathological characteristics were comparable between the two groups, with no significant differences observed according to chi-square or Fisher's exact tests (Table 1).

Radiomic Feature Extraction From B-Mode US and VTIQ Images

A total of 1561 radiomic features were extracted from B-mode US and VTIQ images, comprising 306 first-order features, 374 gray-level co-occurrence matrix (GLCM) features, 238 gray-level dependence matrix (GLDM) features, 272 gray-level run-length matrix (GLRLM) features, 272 gray-level size zone matrix (GLSZM) features, 85 neighborhood gray-tone difference matrix (NGTDM) features, and 14 shape-based features. In the training cohort, feature selection was performed in two steps: first, Spearman correlation analysis was applied to assess inter-feature redundancy, and subsequently, least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation was used to determine the optimal regularization parameter (λ). This procedure yielded 28 selected features from B-mode US data and 15 from VTIQ data.

Performance of B-Mode US and VTIQ Radiomics Models

Predictive models were constructed using six classifiers: logistic regression (LR), support vector machine (SVM), random forest (RF), extremely randomized trees (ExtraTrees), XGBoost, and LightGBM. The areas under the curve (AUCs) for the B-mode US and VTIQ models in both the training and test sets are detailed in Table 2. Considering

Table 1 Clinical Features of Training and Test Cohorts

	Training Set (n = 184)	**Test Set (n = 80)**	***p* value**
Benign lesions	87	24	
Malignant lesions	97	56	
Mean Age (years)	48.53	48.25	0.12
Maximum lesion diameter (mm)	19.5	14	0.36

https://doi.org/10.2147/IJGM.S579165

Table 2 AUC Performance of Classifiers Built on Diverse Features

		LR	SVM	RF	ExtraTrees	XGBoost	LightGBM
B-mode US	Train	0.890(95% CI:0.843–0.938)	0.815(95% CI:0.753–0.877)	0.772(95% CI:0.708–0.835)	0.699(95% CI:0.635–0.763)	0.873(95% CI:0.822–0.923)	0.811(95% CI:0.751–0.871)
	Test	0.844(95% CI:0.753–0.934)	0.762(95% CI:0.656–0.868)	0.677(95% CI:0.564–0.790)	0.726(95% CI:0.617–0.835)	0.791(95% CI:0.692–0.890)	0.759(95% CI:0.664–0.854)
VTIQ model	Train	0.850(95% CI:0.795–0.905)	0.713(95% CI:0.640–0.787)	0.793(95% CI:0.732–0.855)	0.789(95% CI:0.726–0.853)	0.851(95% CI:0.798–0.904)	0.747(95% CI:0.687–0.806)
	Test	0.812(95% CI:0.698–0.927)	0.771(95% CI:0.648–0.894)	0.820(95% CI:0.721–0.920)	0.851(95% CI:0.763–0.940)	0.775(95% CI:0.684–0.865)	0.661(95% CI:0.557–0.755)
1mm peritumoral region	Train	0.872(95% CI:0.826–0.922)	0.774(95% CI:0.708–0,840)	0.821(95% CI:0.760–0.882)	0.827(95% CI:0.771–0.883)	0.768(95% CI:0.706–0.830)	0.736(95% CI:0.672–0.799)
	Test	0.800(95% CI:0.696–0.903)	0.834(95% CI:0.730–0.938)	0.811(95% CI:0.716–0.907)	0.736(95% CI:0.613–0.859)	0.726(95% CI:0.626–0.826)	0.720(95% CI:0.620–0.820)
3mm peritumoral region	Train	0.893(95% CI:0.849–0.938)	0.806(95% CI:0.743–0.869)	0.837(95% CI:0.780–0.894)	0.724(95% CI:0.658–0.790)	0.791(95% CI:0.729–0.853)	0.793(95% CI:0.732–0.855)
	Test	0.871(95% CI:0.795–0.948)	0.729(95% CI:0.610–0.849)	0.785(95% CI:0.679–0.891)	0.692(95% CI:0.575–0.810)	0.793(95% CI:0.692–0.895)	0.760(95% CI:0.650–0.870)
5mm peritumoral region	Train	0.883(95% CI:0.835–0.931)	0.731(95% CI:0.659–0.802)	0.833(95% CI:0.776–0.891)	0.759(95% CI:0.689–0.828)	0.728(95% CI:0.663–0.793)	0.807(95% CI:0.749–0.865)
	Test	0.842(95% CI:0.751–0.932)	0.731(95% CI:0.599–0.863)	0.693(95% CI:0.575–0.811)	0.850(95% CI:0.768–0.933)	0.705(95% CI:0.597–0.814)	0.747(95% CI:0.628–0.867)

overall performance, LR was selected as the optimal classifier for both imaging modalities. Comparative evaluation showed that the B-mode US model outperformed the VTIQ model in diagnostic accuracy. Furthermore, combining features from both modalities further improved predictive capability, with the fused model attaining AUCs of 0.913 (95% CI:0.872–0.953) in the training set.

Extraction of Peritumoral Radiomic Features

Uilding on the superior performance of the B-mode US LR classifier, peritumoral analysis was carried out using B-mode images. Peritumoral regions of interest were automatically delineated by expanding outward 0–1 mm, 0–3 mm, and 0–5 mm from the tumor boundary. From each region, 1561 radiomic features were extracted, consistent with the feature categories used in the intratumoral analysis. Dimensionality reduction with LASSO regression identified 20 discriminative features from the 0–1 mm region, 32 from the 0–3 mm region, and 27 from the 0–5 mm region.

Diagnostic Performance of Peritumoral Features

Machine learning models were constructed using the selected peritumoral features with the same six classifiers. The corresponding AUC values for each configuration are summarized in Table 2. Among them, the logistic regression model based on features from the 0–3 mm peritumoral region demonstrated the highest diagnostic performance.

The Diagnostic Efficacy of Feature Integration

The logistic regression (LR) classifier was selected for fusion model development. Integration of B-mode US with VTIQ features achieved AUCs of 0.913 (95% CI:0.872–0.953). Combining B-mode US with features from the 0–3 mm peritumoral region resulted in AUCs of training=0.908 (95% CI:0.865–0.951). When all three modalities were incorporated, diagnostic performance improved further, with AUCs of 0.933 (95% CI:0.899–0.966) in the training set.

Table 3 Summary of Sensitivity, Specificity, NPV, PPV and Accuracy of Feature Integration

		AUC	Accuracy	95% CI	Sensitivity	Specificity	PPV	NPV
B-mode US+VTIQ	Train	0.913	0.848	0.872–0.953	0.773	0.931	0.926	0.786
	Test	0.874	0.838	0.778–0.970	0.857	0.792	0.906	0.704
B-mode US+Peri	Train	0.908	0.837	0.865–0.951	0.825	0.851	0.860	0.813
	Test	0.859	0.850	0.772–0.947	0.893	0.750	0.893	0.750
B-mode US+VTIQ+Peri	Train	0.933	0.864	0.899–0.966	0.876	0.851	0.867	0.860
	Test	0.903	0.875	0.819–0.986	0.893	0.833	0.926	0.769

Note: Data are reported as percentages (%) with 95% confidence intervals in brackets.
Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

Comprehensive diagnostic metrics—including AUC, 95% confidence interval (CI), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)—are summarized in Table 3, and the corresponding ROC curves are presented in Figure 4. The diagnostic performance of different feature integration methods was compared using the DeLong test, and the results are presented in Figure 5.

Discussion

Breast cancer remains the most prevalent and deadliest malignancy, with an increasing incidence of early-onset cases among younger patients.[12] Although early screening and precise diagnosis are of paramount importance, current methodologies present certain limitations. For instance, while the BI-RADS classification system offers high sensitivity, its accuracy is often affected by inter-observer variability and subjective judgment by radiologists. In addition, benign and malignant breast tumors frequently share overlapping ultrasound features, making differential diagnosis more challenging.[13] Moreover, prior studies indicate that inadequate ultrasonographer expertise or poor image quality can result in false positives in nearly 50% of cases, leading to unnecessary biopsies.[14]

Therefore, there is an urgent need to develop more efficient and objective imaging techniques to enhance diagnostic accuracy and timeliness. Precise early classification of breast tumors plays a critical role in guiding clinical decisions, ultimately contributing to better survival outcomes and quality of life for patients. Radiomics, by enabling the extraction of quantitative features from medical imaging, offers an objective framework to support tumor detection and treatment planning. Prior research has demonstrated the utility of B-mode US models in breast nodule diagnosis, reporting AUC

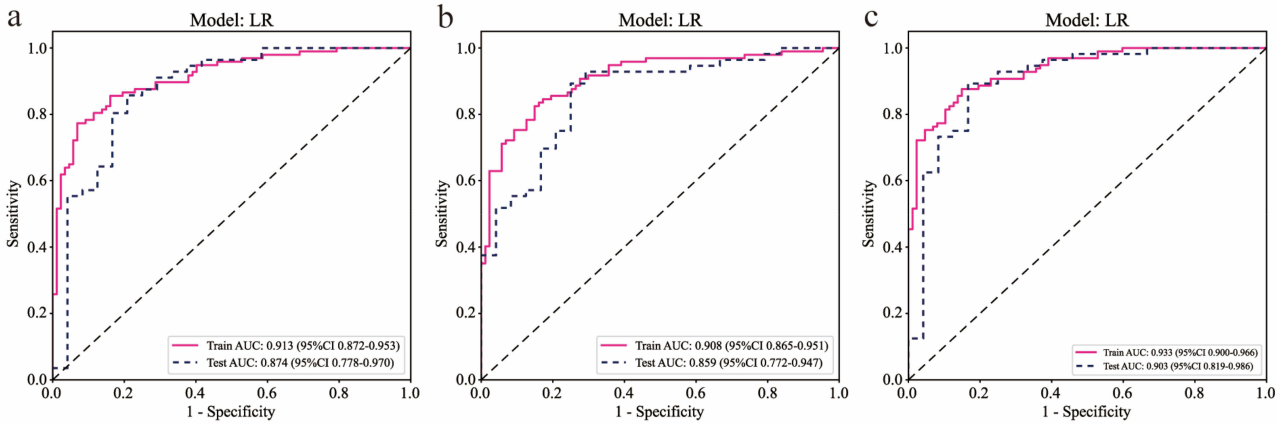

Figure 4 (**a**) ROC curve of radiologist evaluation on US B-mode and VTIQ. (**b**) ROC curve of radiologist evaluation on US B-mode and peritumoral. (**c**) ROC curve of radiologist evaluation on US B-mode and VTIQ and peritumoral.

https://doi.org/10.2147/IJGM.S579165

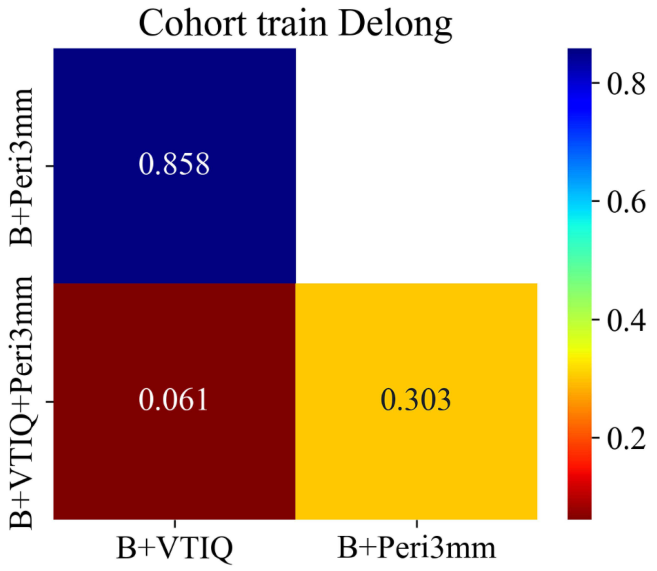

Figure 5 Heatmap of Delong test P-values between different models in the training cohort.
Notes: The heatmap presents the results of Delong test for pairwise comparisons between the "B+VTIQ" model and the "B+Peri3mm" model in the training cohort. The horizontal and vertical axes represent the two model combinations to be compared, respectively. The color of the blocks and the labeled values indicate the corresponding P-values; colors closer to red indicate smaller P-values and more significant statistical differences, while colors closer to blue indicate larger P-values and less significant statistical differences.

values between 0.790 and 0.943.[15–17] In the present study, our B-mode US-based model distinguished benign from malignant breast tumors with an AUC of 0.890 and an overall accuracy of 0.842.

In B-mode US imaging, peritumoral features such as spiculations and hyperechoic halos are commonly seen in breast cancers, reflecting macroscopic infiltration of malignant cells into the surrounding adipose tissue. Microscopically, peritumoral lymphatic vessels serve as key routes for metastatic spread, characterized by elevated lymphatic density and frequent lymphovascular invasion. As interest in the tumor microenvironment grows, radiomic studies have increasingly shifted focus from purely intratumoral regions to also include peritumoral characteristics. Mayerhoefer et al[6] reported that radiomic features capture not only tumor heterogeneity but also properties of the tumor microenvironment, which play a critical role in shaping treatment response and clinical outcomes.[18] Furthermore, MRI-based peritumoral radiomic features have been shown to predict breast cancer subtype, molecular classification, and treatment response.[19]

In this study, radiomic features were extracted from peritumoral regions extending 1 mm, 3 mm, and 5 mm beyond the tumor boundary. The best diagnostic performance was observed at the 3 mm margin, yielding an AUC of 0.893. A 1mm margin may be inadequate to comprehensively capture information pertaining to peritumoral stromal reactions and microinvasion. Conversely, a 5 mm margin is prone to introducing superfluous signals from normal glandular or adipose tissue, which could compromise the discriminative capability of the model to some degree. In comparison, a 3 mm margin strikes a relative balance between adequate information coverage and effective noise suppression.

This result aligns with the findings of Wang et al[11] who reported optimal performance in the 3 mm peritumoral region on mammographic images (AUC = 0.93). When intratumoral and peritumoral features were combined, the diagnostic accuracy further improved, with the AUC increasing to 0.908. Consistently, Guo et al[20] in a study of 591 breast nodules, showed that integrating intratumoral (AUC 0.892) and peritumoral (AUC 0.866) features into a single model enhanced the AUC to 0.929. Similarly, Zhong et al demonstrated that a combined model achieved an AUC of 0.910, markedly outperforming an intratumoral-only model (AUC 0.797).[15]

Deep learning–based radiomics incorporating shear wave elastography (SWE) has demonstrated potential for enhancing breast cancer diagnosis.[21,22] However, the diagnostic performance of SWE may be affected by the operator's level of expertise, differences in ultrasound equipment, and the inherent subjectivity involved in defining the region of interest (ROI) within the lesion as well as in the adjacent tissues.[23] Despite these limitations, the addition of SWE-derived radiomic features has been shown to improve the diagnostic capability of B-mode US-based radiomics.[24] In our study,

VTIQ alone achieved an AUC of 0.850 and an accuracy of 0.799 for differentiating breast tumors, which was slightly inferior to models relying solely on intratumoral or peritumoral features. However, combining B-mode US with VTIQ features increased the AUC to 0.913 and accuracy to 0.848. Notably, the integration of B-mode US, VTIQ, and peritumoral features into a tri-modal model produced the best overall performance, achieving an AUC of 0.933. Although the diagnostic area under the curve (AUC) of the triple-modal fusion model was numerically higher than that of the dual-modal fusion model in the present study, the Delong test failed to demonstrate a statistically significant difference between the two groups ($P > 0.05$). Nevertheless, the triple-modal model consistently exhibited the highest AUC, accuracy among all models evaluated in this study. Notably, such numerical improvements still carry practical reference significance for reducing unnecessary invasive interventions in BI-RADS category 4 lesions and may offer a more objective reference for the differentiation of clinically recalcitrant lesions. From a statistical perspective, this observation is primarily ascribed to the inherent limitations of the single-center retrospective design and the relatively modest sample size of the current study; specifically, the small sample size resulted in inadequate statistical power, thereby precluding the detection of subtle inter-model differences in diagnostic efficacy.

The novelty of this study lies in its multi-dimensional approach to feature extraction, integrating complementary information from B-mode US and elastographic imaging while systematically comparing traditional intratumoral radiomics with multiple peritumoral regions. These components were incorporated into a unified predictive framework, achieving comprehensive coverage of ultrasonic imaging information spanning morphology, tissue stiffness, and the tumor invasive microenvironment. This provides a more holistic radiomics feature system for the objective differentiation of breast lesions. Notably, this study is the first to systematically compare the diagnostic efficacy of three peritumoral annular zones (1 mm/3 mm/5 mm) in BI-RADS category 4 breast lesions, thereby addressing a critical gap in ultrasonic multi-modal radiomics research for this specific lesion subset. Of particular importance, our model demonstrated numerically higher specificity than the conventional BI-RADS classification which highlights its potential to reduce unnecessary biopsies and limit invasive interventions in patients with diagnostically challenging breast tumors. Furthermore, this study preliminarily clarifies the potential added value of elastographic features and peritumoral characteristics in the stratification of BI-RADS category 4 lesions, offering a more comprehensive imaging basis for risk assessment in complex cases. Compared with population-wide studies, the present work exhibits stronger clinical targeting and practical guiding significance for the management of BI-RADS 4 lesions specifically.

The frontier of breast radiomics has evolved from traditional machine learning (ML) to the integration of deep learning (DL) and radiomics. Recent cutting-edge studies have confirmed that DL can better capture subtle texture and spatial features of ultrasonic images, achieving an AUC of up to 0.94 in the diagnosis of BI-RADS category 4 lesions.[25,26]

In this study, only traditional ML algorithms were employed, without attempting DL models. While this choice ensures model interpretability—a key advantage for clinical translation—it also introduces inherent limitations in automatic feature extraction and complex texture recognition, leading to a potential gap with the latest technological advancements. Notably, traditional machine learning-based radiomics models yield superior explainable artificial intelligence (XAI) compared with deep learning models. XAI not only enhances clinical users' understanding and trust in prediction outcomes by visualizing model decision bases (eg, saliency maps, feature attribution) but also provides a technical foundation for the safety and regulatory compliance of artificial intelligence systems. Unlike end-to-end deep learning black-box models, shallow models based on manually extracted radiomic features generally provide more explicit correspondences between quantitative features and imaging phenotypes, facilitating the interpretation of classification rationale. Meanwhile, such models can be integrated with interpretable mechanisms within the XAI framework to improve clinical applicability.

Even with a moderate sample size (264 cases), conventional radiomics approaches enabled accurate feature extraction and reduced overfitting risks via cross-validation, despite their well-recognized limitations in capturing complex high-dimensional patterns in medical images. The shallow learning models adopted herein (eg, logistic regression, support vector machine) possess intrinsic interpretability: the selected radiomic features (eg, first-order statistics, texture matrices) allow radiologists to trace model decisions back to quantifiable imaging features. In contrast, although deep learning models exhibit greater potential in complex feature extraction, they lack intrinsic interpretability and usually require post-hoc XAI methods to uncover decision drivers. XAI approaches have demonstrated diverse explanatory capabilities and limitations across different

tasks, suggesting that future research can explore the integration of feature-driven models with advanced interpretation mechanisms to achieve a better balance between performance and interpretability.[27–29]

The integration of peritumoral features in this study—with optimal efficacy observed for the 3 mm margin—further underscores the value of interpretability design: this finding is highly consistent with established clinical imaging insights (eg, peritumoral hyperechoic halos associated with malignant lesions). Mechanistically, the peritumoral region can reflect key biological processes such as stromal reactions and angiogenesis, thereby associating radiomics features with underlying pathological mechanisms. This biological interpretability enhances the clinical relevance of the model, aligning the study results with known tumor progression pathways. Additionally, cross-validation and calibration across multi-modal analyses (B-mode US + elastography) improved diagnostic accuracy and reliability, which may facilitate early and precise diagnosis and treatment of BI-RADS 4 lesions.

Despite the encouraging outcomes, this study has several inherent limitations that should be acknowledged. First, the relatively modest sample size (264 patients) may restrict the generalizability of the results, as smaller cohorts are less likely to capture the full spectrum of BI-RADS 4 lesion heterogeneity. Second, as a single-center retrospective analysis, the study is inherently prone to selection bias—specifically, the overrepresentation of lesions with higher clinical suspicion (given the surgical cohort inclusion criterion). Therefore, external validation in independent, multi-center cohorts with diverse patient populations is warranted to confirm the model's performance across different clinical settings.

It is important to note that the study sample comprises patients who underwent surgical treatment, meaning the study population may be inherently biased toward lesions with higher clinically assessed risk. This population selection limits the direct extrapolation of study conclusions to screening populations or patients who did not undergo surgery. However, it simultaneously ensures the accuracy and consistency of outcome determination (using surgical pathology as the gold standard), providing a robust foundation for analyzing the relationship between imaging features and pathological results —an advantage that enhances the internal validity of the study.

Manual segmentation was used for lesion delineation in this study, a common approach in ultrasonic radiomics research that helps ensure consistency between lesion boundaries and pathological findings. To mitigate the impact of inter-observer variability on feature stability, a subset of 30 randomly selected cases was independently segmented by two board-certified radiologists with 5 years of breast imaging experience, respectively. Consistency metrics (ICC > 0.9 for all key features) were calculated, demonstrating good reproducibility. Nevertheless, manual segmentation may inherently involve subtle subjective recognition of tumor morphological features, which could potentially influence certain morphology-related radiomics features. Future studies could incorporate automatic or semi-automatic segmentation algorithms to further enhance the objectivity and reproducibility of the method, while also comparing the performance of manual versus automated segmentation to quantify the impact of segmentation method on model efficacy.

Conclusion

Integrating multi-modal imaging features encompassing B-mode ultrasound, elastography, and peritumoral radiomics can markedly improve the diagnostic accuracy for breast tumors. This comprehensive approach not only can strengthen early detection but also provide clinicians with a more reliable, objective, and data-driven tool to guide personalized treatment strategies. Future research will continue to advance this field toward more precise and efficient diagnostic strategies. The incorporation of artificial intelligence and advanced machine learning algorithms may further enhance predictive performance and enable automated, real-time diagnostic support. Additionally, combining imaging-based radiomics with genomic, proteomic, and liquid biopsy data could pave the way for integrated multi-omics diagnostic platforms, offering deeper insights into tumor biology and treatment response.

Acknowledgments

We would like to express our sincere gratitude to Professor Qi Chen from the First Affiliated Hospital of Harbin Medical University for her meticulous guidance throughout this study. We also acknowledge the generous support and assistance from all colleagues in the Departments of Ultrasound, CT and Surgery of the hospital. This work was funded by the Foundation of the First Affiliated Hospital of Harbin Medical University (2023T01) and the Open Fund of the Key Laboratory of Hepatosplenic Surgery, Ministry of Education, the First Affiliated Hospital of Harbin Medical University

(GPKF202504), and we are grateful for the financial support of these foundations. Finally, we extend our heartfelt thanks to all individuals and institutions that have contributed to this study.

Funding

1.Supported by Foundation of The First Affiliated Hospital of Harbin Medical University, Harbin, China [2023T01]; 2. Supported by the Open Fund of Key Laboratory of Hepatosplenic Surgery, Ministry of Education, The First Affiliated Hospital of Harbin Medical University, Harbin, China [GPKF202504].

Disclosure

The authors report no conflicts of interest in this work.

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 Cancer J Clin*. 2021;71(3):209–249. doi:10.3322/caac.21660
2. Vijayaraghavan GR, Vedantham S, Santos-Nunez G, Hultman R. Unifocal invasive lobular carcinoma: tumor size concordance between pre-operative ultrasound imaging and postoperative pathology. *Clin Breast Cancer*. 2018;18(6):e1367–e1372. doi:10.1016/j.clbc.2018.07.017
3. Sadigh G, Carlos RC, Neal CH, Wojcinski S, Dwamena BA. Impact of breast mass size on accuracy of ultrasound elastography vs. conventional B-mode ultrasound: a meta-analysis of individual participants. *Eur Radiol*. 2013;23:1006–1014. doi:10.1007/s00330-012-2682-0
4. Liu B, Zheng Y, Huang G, et al. Breast lesions: quantitative diagnosis using ultrasound shear wave elastography-a systematic review and meta--analysis. *Ultrasound Med Biol*. 2016;42(4):835–847. doi:10.1016/j.ultrasmedbio.2015.10.024
5. Barr RG. The role of sonoelastography in breast lesions. *Semin Ultrasound CT MR*. 2018;39(1):98–105. doi:10.1053/j.sult.2017.05.010
6. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to radiomics. *J Nucl Med*. 2020;61(4):488–495. doi:10.2967/jnumed.118.222893
7. Gong Y, Cheng Y, Liu Y, et al. Diagnostic value of ultrasound radiomic features in differentiating benign and malignant breast lesions. *J Ultrasound*. 2025;28(3):645–652. doi:10.1007/s40477-025-01025-8
8. Lyu S, Zhang M, Zhang B, et al. The value of radiomics model based on ultrasound image features in the differentiation between minimal breast cancer and small benign breast masses. *J Clin Ultrasound*. 2023;51(9):1536–1543. doi:10.1002/jcu.23556
9. Yu H, Meng X, Chen H, et al. Correlation between mammographic radiomics features and the level of tumor-infiltrating lymphocytes in patients with triple-negative breast cancer. *Front Oncol*. 2020;10:412. doi:10.3389/fonc.2020.00412
10. Li R. Peritumoral radiomics and predicting treatment response. *JAMA Network Open*. 2020;3(9):e2016125. doi:10.1001/jamanetworkopen.2020.16125
11. Wang S, Sun Y, Li R, et al. Diagnostic performance of perilesional radiomics analysis of contrast-enhanced mammography for the differentiation of benign and malignant breast lesions. *Eur Radiol*. 2022;32(1):639–649. doi:10.1007/s00330-021-08134-y
12. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: an overview. *Int J Cancer*. 2021;149(4):778–789.
13. Youk JH, Kwak JY, Lee E, Son EJ, Kim JA. Grayscale ultrasound radiomic features and shear-wave elastography radiomic features in benign and malignant breast masses. Radiomische Merkmale im B-Bild und in der scherwellen-elastografie bei benignen und malignen Raumforderungen Der Brust. *Ultraschall Med*. 2020;41(4):390–396. doi:10.1055/a-0917-6825
14. Mann RM, Hooley R, Barr RG, Moy L. Novel approaches to screening for breast cancer. *Radiology*. 2020;297(2):266–285. doi:10.1148/radiol.2020200172
15. Zhong L, Shi L, Zhou L, Liu X, Gu L, Bai W. Development of a nomogram-based model combining intra- and peritumoral ultrasound radiomics with clinical features for differentiating benign from malignant in Breast Imaging Reporting and Data System category 3-5 nodules. *Quant Imaging Med Surg*. 2023;13(10):6899–6910. doi:10.21037/qims-23-283
16. Li J, Bu Y, Lu S, et al. Development of a deep learning-based model for diagnosing breast nodules with ultrasound. *J Ultrasound Med*. 2021;40(3):513–520. doi:10.1002/jum.15427
17. Hong ZL, Chen S, Peng XR, Li JW, Yang JC, Wu SS. Nomograms for prediction of breast cancer in breast imaging reporting and data system (BI-RADS) ultrasound category 4 or 5 lesions: a single-center retrospective study based on radiomics features. *Front Oncol*. 2022;12:894476. doi:10.3389/fonc.2022.894476
18. Soysal SD, Tzankov A, Muenst SE. Role of the tumor microenvironment in breast cancer. *Pathobiology*. 2015;82(3–4):142–152. doi:10.1159/000430499
19. 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. doi:10.1186/s13058-017-0846-1
20. Guo S, Huang X, Xu C, et al. Multiregional radiomic model for breast cancer diagnosis: value of ultrasound-based peritumoral and parenchymal radiomics. *Quant Imaging Med Surg*. 2023;13(5):3127–3139. doi:10.21037/qims-22-939
21. Zhang X, Liang M, Yang Z, et al. Deep learning-based radiomics of B-Mode ultrasonography and shear-wave elastography: improved performance in breast mass classification. *Front Oncol*. 2020;10:1621. doi:10.3389/fonc.2020.01621
22. Zhang Q, Song S, Xiao Y, et al. Dual-mode artificially-intelligent diagnosis of breast tumours in shear-wave elastography and B-mode ultrasound using deep polynomial networks. *Med Eng Phys*. 2019;64:1–6. doi:10.1016/j.medengphy.2018.12.005
23. Zhang L, Xu J, Wu H, et al. Screening breast lesions using shear modulus and its 1-mm shell in sound touch elastography. *Ultrasound Med Biol*. 2019;45(3):710–719. doi:10.1016/j.ultrasmedbio.2018.11.013
24. La Rocca LR, Caruso M, Stanzione A, et al. Machine learning-based discrimination of benign and malignant breast lesions on US: the contribution of shear-wave elastography. *Eur J Radiol*. 2024;181:111795. doi:10.1016/j.ejrad.2024.111795
25. Li Y, Li C, Yang T, et al. Multiview deep learning networks based on automated breast volume scanner images for identifying breast cancer in BI-RADS 4. *Front Oncol*. 2024;14:1399296. doi:10.3389/fonc.2024.1399296

26. Yang Y, Zhong Y, Li J, et al. Deep learning combining mammography and ultrasound images to predict the malignancy of BI-RADS US 4A lesions in women with dense breasts: a diagnostic study. *Int J Surg*. 2024;110(5):2604–2613. doi:10.1097/JS9.0000000000001186
27. Gipikis R, Tsai CW, Kurasova O. Explainable AI (XAI) in image segmentation in medicine, industry, and beyond: a survey. *ICT Express*. 2024;10 (6):1331–1354. doi:10.1016/j.icte.2024.09.008
28. Rundo L, Militello C. Image biomarkers and explainable AI: handcrafted features versus deep learned features. *Eur Radiol Exp*. 2024;8(1):130. doi:10.1186/s41747-024-00529-y.
29. Abrantes J, Rouzrokh P. Explaining explainability: the role of XAI in medical imaging. *Eur J Radiol*. 2024;173:111389. doi:10.1016/j. ejrad.2024.111389