British Journal of Hospital Medicine

Br. J. Hosp. Med. (Lond) **2026**; 87(7): 56311
https://doi.org/10.31083/BJHM56311

Article

Integrated Intratumoral and Peritumoral Ultrasound Radiomics Models for Breast Nodule Diagnosis Using Machine Learning

Xinru Yang[1], Jiaxin Zuo[1], Cuiping Liu[2], Huihong Zhou[1], Ping Xiong[1,*]

[1]Department of Ultrasound Medicine, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, 200011 Shanghai, China
[2]Department of Ultrasound Medicine, Shanghai Second People's Hospital, 200011 Shanghai, China
*Correspondence: xiongpxp@163.com (Ping Xiong)
Academic Editor: Raghavan Vidya
Submitted: 30 December 2025 Revised: 11 March 2026 Accepted: 27 March 2026 Published: 24 July 2026

Abstract

Aims/Background: The differentiation of benign and malignant breast nodules, particularly those categorized as Breast Imaging Reporting and Data System (BI-RADS) 3–4, remains a clinical challenge due to the subjectivity and operator dependence of conventional ultrasound assessment. This study aimed to evaluate the diagnostic value of intratumoral and peritumoral ultrasound radiomics features in distinguishing benign from malignant BI-RADS 3–4 breast nodules and to construct interpretable machine learning models. **Methods:** Ultrasound images and parameters (BI-RADS classification) of breast nodules were retrospectively collected from female patients at two institutions between January 2021 and June 2024: 571 patients (880 nodules) from Institution 1 (Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine) and 333 patients (351 nodules) from Institution 2 (Shanghai Second People's Hospital). Included patients had BI-RADS 3–4 nodules confirmed by pathology or, for BI-RADS 3 nodules, stable findings on follow-up for at least 2 years. Nodules from Institution 1 were randomly divided into a training set (n = 615) and an internal validation set (n = 265) at a 7:3 ratio, while patients from Institution 2 constituted an external test set (n = 351). Radiomics features of intratumoral and peritumoral regions (3 and 5 pixels wide) were extracted using PyRadiomics. Features were screened using the independent *t*-test, Spearman correlation, and least absolute shrinkage and selection operator (LASSO) regression. Machine learning models—including random forest (RF), multilayer perceptron (MLP), extra trees (ET), support vector machine (SVM), logistic regression (LR), k-nearest neighbor (KNN), eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), Gradient Boosting Decision Tree (GBDT), and Light Gradient Boosting Machine (LightGBM)—were trained and compared using area under the curve (AUC) and other metrics. The best-performing model was then used to compare intratumoral versus peritumoral regions. SHapley Additive exPlanations (SHAP) was applied to interpret feature importance. **Results:** Across the training, internal validation, and external test sets, SVM demonstrated the most stable and balanced performance. The 3 pixels transitional zone region of interest support vector machine (EI3_SVM) model achieved a higher AUC than the tumor core region of interest (T) model in the external test set (0.875 vs. 0.787, $p < 0.01$), with accuracy 78.1%, sensitivity 42.2%, specificity 97.0%, and Brier score 0.166, outperforming other peritumoral models. Most models incorporating peritumoral features demonstrated AUCs that were higher than or comparable to those of the T model. SHAP analysis indicated that the high specificity was mainly driven by texture and shape features. **Conclusion:** Integrating intratumoral and peritumoral radiomics features significantly improves the diagnostic accuracy and objectivity for differentiating benign and malignant breast nodules. This approach may aid clinical decision-making, reduce unnecessary biopsies, and support early precision diagnosis of breast cancer.

Keywords: breast nodule; breast neoplasms; ultrasonography; radiomics; machine learning

1. Introduction

In recent years, the incidence of breast cancer (BC) has been steadily increasing worldwide, making it the most commonly diagnosed cancer and the second leading cause of cancer-related death among women, with an estimated 2.3 million new cases and 670,000 deaths in 2022 [1]. This year, the International Agency for Research on Cancer reported in Nature Medicine that newly diagnosed breast cancer cases account for 25% of all new female cancer cases worldwide, and the mortality rate represents 15.5% of female cancer deaths [2]. Therefore, early detection and accurate characterization of breast nodules are crucial for improving breast cancer survival.

Currently, the most commonly used imaging modalities for indirectly predicting breast cancer are ultrasound, mammography, and magnetic resonance imaging (MRI); however, distinguishing between different molecular subtypes of breast cancer remains challenging. Ultrasound, as an accurate, cost-effective, and easily repeatable imaging technique, has become a first-line clinical tool for breast disease screening and follow-up. It holds significant clinical value for early breast cancer detection, lymph node metastasis assessment, and postoperative patient monitoring. Moreover, it has been shown to simultaneously improve breast cancer detection rates while reducing unnecessary biopsies [3]. Although the Breast Imaging Reporting and Data System (BI-RADS)–based ultrasound evaluation

system is well established, it still faces significant issues with false positives and false negatives in the differential diagnosis of category 3 and 4 nodules. It is highly dependent on operator experience and limited by subjectivity and insufficient diagnostic reliability [4]. Thus, there is a need to introduce objective and stable intelligent-assisted tools to improve diagnostic accuracy and consistency.

Radiomics enables the high-throughput extraction of quantitative features from medical images, including morphological, intensity, texture, and functional information. These features are difficult for the human eye to recognize or quantify and have shown promising potential in tumor diagnosis across imaging modalities such as computed tomography (CT) and MRI [5]. Over the past few years, radiomics has seen considerable progress in breast cancer, particularly in tumor diagnosis, evaluating lymph node involvement, and predicting patient outcomes [6,7]. Progress in the field of radiology is quickly reshaping how breast cancer is diagnosed and treated, especially by providing insights into the complex heterogeneity of the tumor microenvironment.

Current studies often focus exclusively on features extracted from the tumor interior, overlooking potentially valuable biological information in the surrounding tissue. The breast tumor microenvironment (TME) comprises not only malignant cells but also stromal cells, immune cells, and vascular components, all of which play key roles in tumor growth, progression, and metastasis [8,9]. A recent imaging radiomics study has demonstrated that combining intratumoral and peritumoral features improves the accuracy of differentiating benign from malignant breast lesions [10].

This study aims to investigate the value of intratumoral and peritumoral ultrasound radiomics features at varying widths and to develop an interpretable machine learning model. The goal is to provide an objective and robust intelligent tool for distinguishing between benign and malignant breast nodules, reducing misdiagnosis, and guiding clinical treatment strategies. Based on this, we systematically defined and compared intratumoral, peritumoral, and transitional region features with varying widths (3/5 pixels) in breast ultrasound radiomics for the first time. We then developed an interpretable machine learning model focused on high specificity with the aim of reducing unnecessary biopsies. We anticipate that this work will provide a novel tool and insights for the objective and precise diagnosis of BI-RADS 3–4 lesions.

2. Methods

2.1 Patient Selection

This retrospective, dual-center study included 571 patients (880 nodules; 763 benign, 117 malignant) from Institution 1 (Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine) and 333 patients (351 nodules; 230 benign, 121 malignant) from Institution 2 (Shanghai Second People's Hospital), collected between January 2021 and June 2024.

Inclusion criteria were as follows: (1) patients who underwent ultrasound examination before surgery or biopsy; (2) patients with BI-RADS category 3–4 lesions; (3) patients with a definitive pathological diagnosis obtained through surgery or biopsy; and for BI-RADS 3 lesions not subjected to surgery or biopsy, patients with at least 2 years of follow-up at our hospital showing no change in classification.

Exclusion criteria were as follows: (1) patients with a history of breast cancer; (2) patients who had received neoadjuvant chemotherapy or radiotherapy prior to the ultrasound examination; (3) patients who were pregnant or lactating at the time of examination; (4) poor-quality or incomplete ultrasound images, or incomplete clinical data.

Clinical and pathological data were collected for all included patients, including age, tumor size, BI-RADS classification, and postoperative or biopsy pathology results, which were used for subsequent statistical analyses. Nodules from Institution 1 were randomly divided into a training set (n = 615) and an internal validation set (n = 265) in a 7:3 ratio, while nodules from Institution 2 served as an external test set (n = 351). The study workflow is illustrated in Figs. 1,2.

2.2 Ultrasound Image Acquisition and Preprocessing

Breast ultrasound images were acquired using various ultrasound systems to ensure data diversity: MyLab Twice (Esaote S.p.A., Genoa, Italy), PHILIPS-IU ELITE (Philips Ultrasound, Inc., Bothell, WA, USA), TOSHIBA-Aplio 500 (Canon Medical Systems Corporation, Otawara, Japan), Mindray Resona R9S (Mindray Medical International Limited, Shenzhen, China), and HITACHI ALOKA Noblus (Hitachi, Ltd., Tokyo, Japan) etc. All devices were fitted with a 7–12 MHz linear array probe, which enabled the capture of detailed grayscale images. Standard imaging protocols were followed to minimize variability, including consistent probe positioning, patient supine posture with full exposure of the chest and axilla, and careful centering of the target area on the screen. For each identified breast nodule, the grayscale image displaying the largest diameter in the long-axis view was included in the analysis, and the maximum longitudinal dimension of each lesion was documented to assess size-related features. Following the American College of Radiology (ACR) BI-RADS classification, nodule characteristics, including location, margin, shape, and internal echogenicity, were systematically evaluated.

Image preprocessing was performed using Python 3.7.1 (Python Software Foundation, Wilmington, DE, USA) for grayscale conversion and pixel value normalization. Regions of interest (ROIs) were delineated on 3D Slicer 5.6.2 (Slicer Community, Boston, MA, USA). Two radiologists specializing in breast ultrasound (Reader 1: 5 years of experience; Reader 2: 2 years of experience) in-

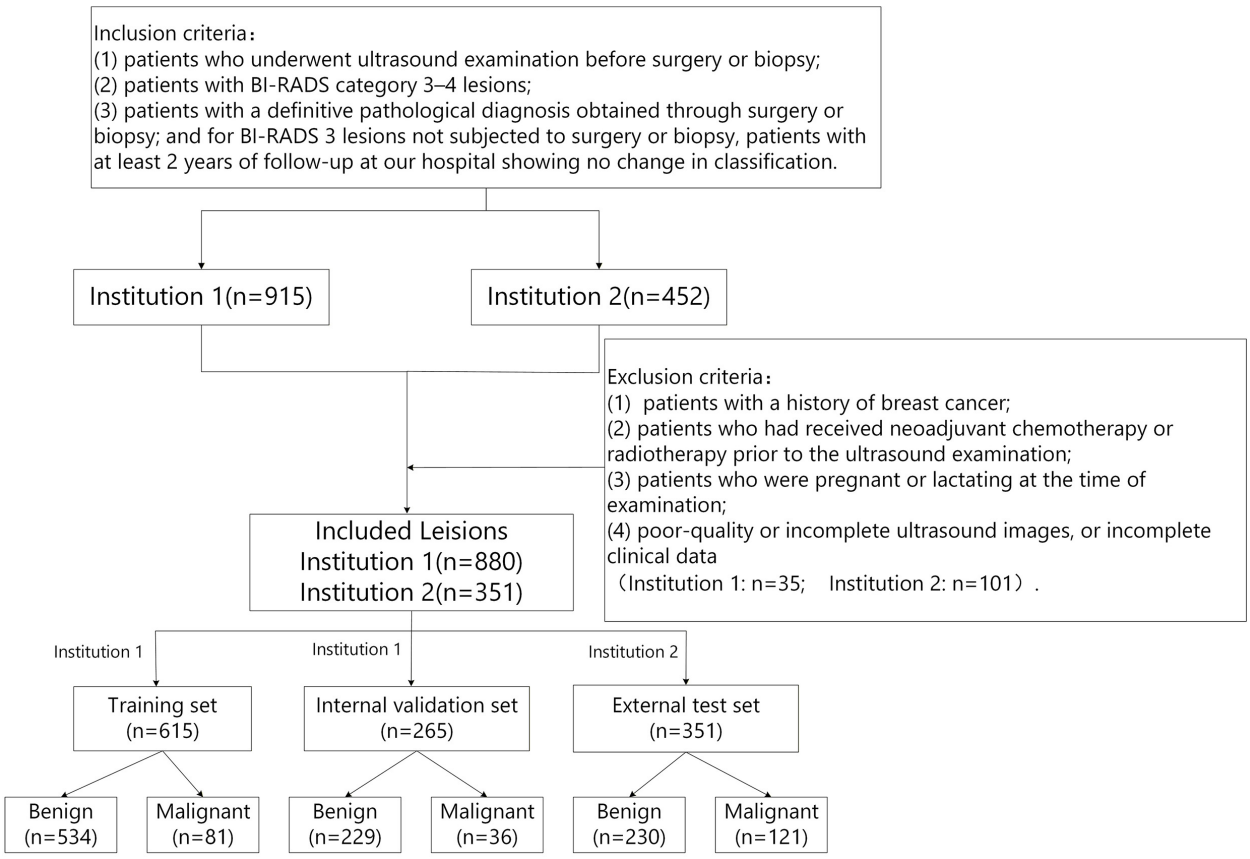

Fig. 1. The process of breast nodule enrollment. BI-RADS, Breast Imaging Reporting and Data System. Institution 1, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine; Institution 2, Shanghai Second People's Hospital.

dependently performed blinded segmentation, unaware of the pathological results, to obtain the tumor core ROI (T). To capture contextual information from surrounding tissue, multiple peripheral ROIs were generated by expanding or contracting the tumor ROI boundary by either 3 or 5 pixels. These peritumoral ROIs were automatically generated using the "Margin" tool in 3D Slicer to ensure consistency across all samples.

Nine ROI groups were defined based on tumor boundaries (Fig. 3):

(1) Tumor core ROI (T)—delineated along the inner edge of the tumor capsule, covering the entire tumor parenchyma.

(2) 3 pixels peritumoral ROI (E3)—defined by outwardly expanding the tumor capsule by 3 pixels, including the entire tumor region.

(3) 5 pixels peritumoral ROI (E5)—defined by outwardly expanding the tumor capsule by 5 pixels, including the entire tumor region.

(4) 3 pixels inner margin ROI (TI3)—a concentric intratumoral ring adjacent to the tumor edge, created by contracting the tumor capsule inward by 3 pixels.

(5) 5 pixels inner margin ROI (TI5)—a concentric intratumoral ring adjacent to the tumor edge, created by contracting the tumor capsule inward by 5 pixels.

(6) 3 pixels transitional zone ROI (EI3)—a 6-pixel-wide band centered on the tumor capsule, extending 3 pixels inward and 3 pixels outward from the capsule.

(7) 5 pixels transitional zone ROI (EI5)—a 10-pixel-wide band centered on the tumor capsule, extending 5 pixels inward and 5 pixels outward from the capsule.

(8) 3 pixels outer peritumoral margin ROI (ET3)—the outermost 3 pixels ring of the E3 region adjacent to the tumor capsule, excluding the tumor itself.

(9) 5 pixels outer peritumoral margin ROI (ET5)—the outermost 5 pixels ring of the E5 region adjacent to the tumor capsule, excluding the tumor itself.

During ROI delineation, care was taken to ensure that all contours closely followed the actual tumor boundaries, minimizing potential displacement caused by operational error. For nodules with diameters smaller than the target range, inward contraction was limited by the effective tumor radius.

2.3 Feature Extraction and Selection

The delineated ROI images were then imported into PyRadiomics (v3.1.0) (The Computational Imaging & Bioinformatics Laboratory, Harvard Medical School, Boston, MA, USA) for feature extraction, while preserv-

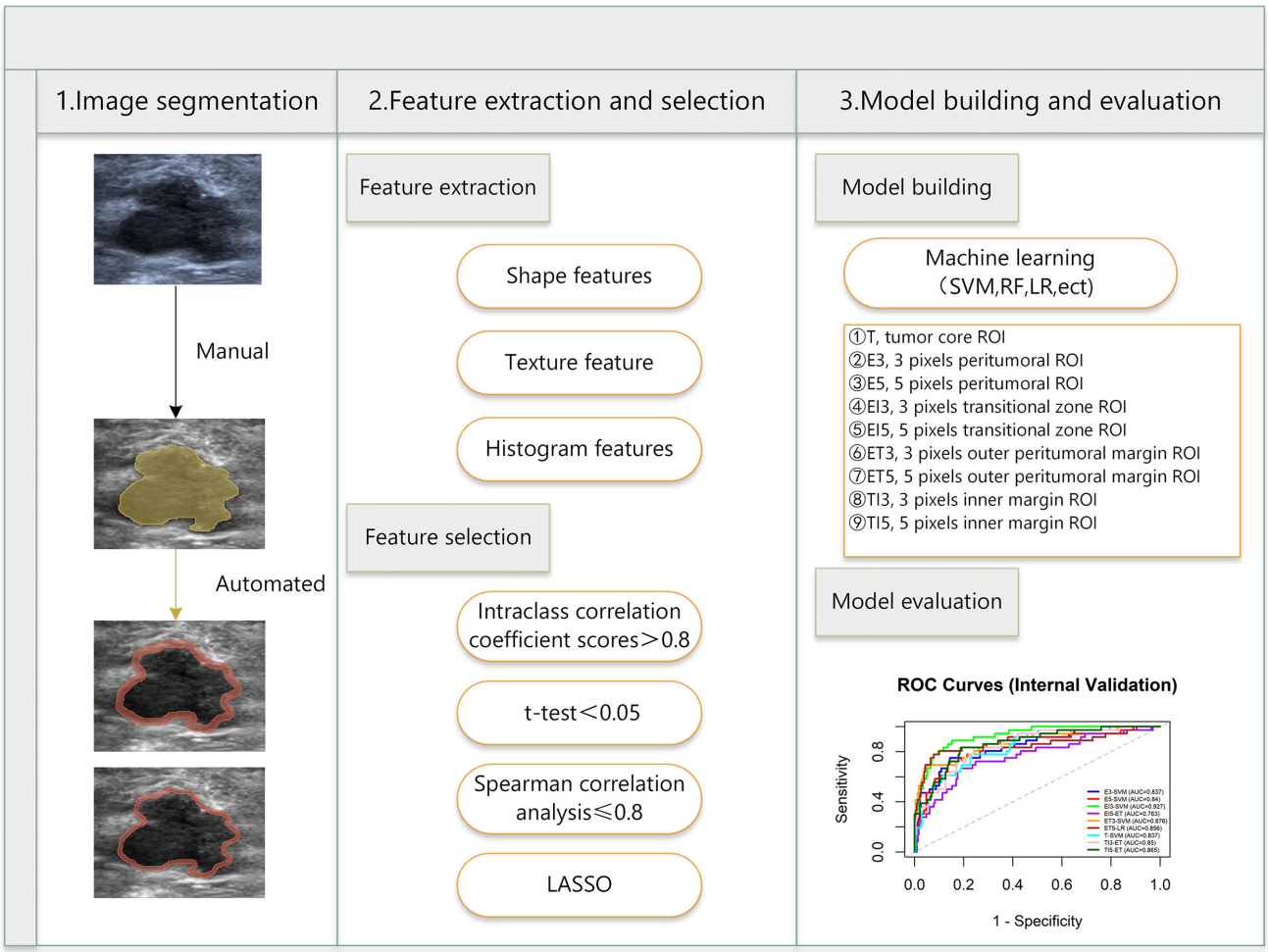

Fig. 2. Study flowchart. SVM, support vector machine; RF, random forest; LR, logistic regression; ROI, region of interest; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; ET, extra trees; AUC, area under the area.

ing the original ultrasound images. Extracted features included morphological features, first-order statistics, and texture features derived from the Gray Level Size Zone Matrix (GLSZM), Gray Level Co-occurrence Matrix (GLCM), Neighboring Gray Tone Difference Matrix (NGTDM), Gray Level Dependence Matrix (GLDM), and Gray Level Run Length Matrix (GLRLM). Additionally, frequency-domain features were extracted using wavelet transforms, from which first-order and texture features were further computed, thereby expanding the feature set to capture multi-scale spatial-frequency information. The consistency of ROI delineation between the two radiologists was evaluated using the intraclass correlation coefficient (ICC), with features showing ICC ≤0.8 being excluded.

Features were Z-score normalized and subjected to independent *t*-tests, removing those with $p > 0.05$. Features exhibiting strong pairwise Spearman correlations ($r > 0.80$) or the highest average correlation with all other features were further excluded. Radiomics signatures were constructed using the least absolute shrinkage and selection operator (LASSO) regression model on the discovery dataset. LASSO shrinks all regression coefficients toward zero and sets the coefficients of irrelevant features to zero based on the weighting parameter λ. The optimal λ was determined using 10-fold cross-validation with the minimum standard, with the final λ corresponding to the minimum cross-validation error. The combination of non-zero coefficient features retained by LASSO constituted the radiomics signature. All features were standardized (Z-score) using only the mean and standard deviation calculated from the training set, which were then consistently applied to the internal validation and external test sets to prevent data leakage.

2.4 Model Building and Evaluation

Based on the selected optimal radiomics features, ten machine learning algorithms were employed to construct intratumoral and peritumoral radiomics models: random forest (RF), multilayer perceptron (MLP), extra trees (ET), support vector machine (SVM), logistic regression (LR), k-nearest neighbors (KNN), eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), Gradi-

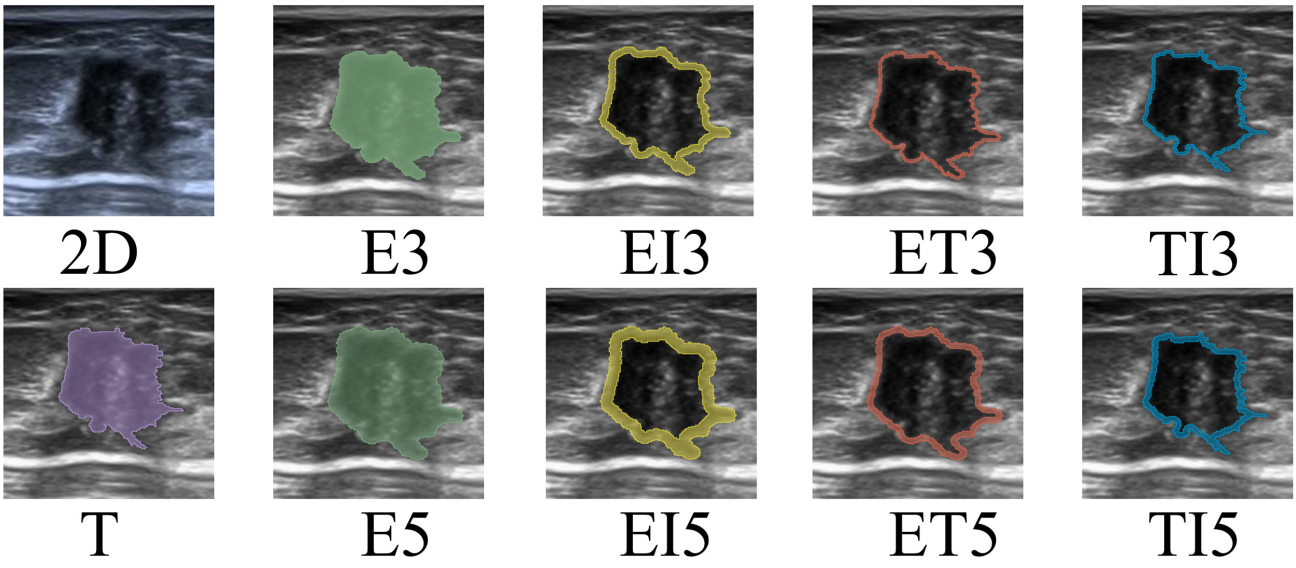

Fig. 3. Examples of ROI delineation on the original 2D grayscale ultrasound images in 9 representative cases. T, tumor core ROI; E3, 3 pixels peritumoral ROI; E5, 5 pixels peritumoral ROI; EI3, 3 pixels transitional zone ROI; EI5, 5 pixels transitional zone ROI; ET3, 3 pixels outer peritumoral margin ROI; ET5, 5 pixels outer peritumoral margin ROI; TI3, 3 pixels inner margin ROI; TI5, 5 pixels inner margin ROI.

ent Boosting Decision Tree (GBDT), and Light Gradient Boosting Machine (LightGBM). 10-fold cross-validation was performed on the training set to train and evaluate the performance of these models, identify the optimal hyperparameters for each algorithm, and generate receiver operating characteristic (ROC) curves to compare the area under the curve (AUC) of all ten radiomics models.

To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively within each fold of the cross-validation on the training set, using k_neighbors = 5 and random_state = 42 for reproducibility. This ensured no data leakage to the validation or test sets. SMOTE was incorporated into a pipeline with each classifier, guaranteeing that resampling was performed independently within every cross-validation fold, thereby preventing data leakage between the training and evaluation phases.

Following the identification of the optimal model using a two-stage modeling strategy, we compared its diagnostic performance between the intratumoral region and the peritumoral-inclusive region. Specifically, the optimal model was applied to feature sets derived from the tumor core ROI (T) and the optimal peritumoral-inclusive region to assess the incremental value of peritumoral features. This comparison was conducted on an independent external test set to ensure an unbiased evaluation of model generalizability. Threshold Selection Strategy: To align with the clinical objective of maximizing specificity while maintaining clinically acceptable sensitivity, we implemented a two-stage threshold optimization approach. During 10-fold cross-validation on the training set, optimal thresholds within each fold were determined for each fold by maximizing the Youden index under the constraint that specificity remained above 90%. For independent validation and testing, a single optimal threshold was established using the internal validation set, applying the same optimization criterion (maximizing the Youden index with specificity ≥90% constraint). This threshold was then fixed and applied directly to the external test set without further adjustment, ensuring an unbiased assessment of model generalizability. No absolute minimum sensitivity threshold was set, as the primary clinical aim was to develop a high-specificity tool to reduce unnecessary biopsies.

The robustness of performance metrics was assessed using bootstrap resampling (1000 repetitions) to calculate 95% confidence intervals (CIs). Model performance was further evaluated on the independent validation and test sets using multiple metrics, including accuracy, sensitivity, specificity, F1 score, and Brier score. The F1 score—the harmonic mean of precision and recall—was used to provide a balanced assessment of model performance, particularly given the class imbalance in the dataset. ROC curves were used to visualize model performance, and the AUC with corresponding CIs was used to quantify predictive accuracy.

2.5 Statistical Analysis

Statistical analyses were performed using Python 3.7.1 (Python Software Foundation, Wilmington, DE, USA) and R 4.5.0 (R Foundation for Statistical Computing, Vienna, Austria).

Model development and evaluation: the diagnostic performance of each model was evaluated using the area under the curve (AUC). Calibration accuracy of predicted

probabilities was assessed using the Brier score. Pairwise comparisons of model performance were conducted using the DeLong test, with *p*-values adjusted for multiple comparisons using the Bonferroni method. ROC curve analysis was used to determine the optimal threshold, defined by maximizing the Youden index. Based on the optimal cutoff, accuracy, sensitivity, and specificity were calculated for each model. Calibration curves were used to assess model agreement, while decision curve analysis (DCA) quantified the net benefit across different threshold probabilities to evaluate clinical utility.

Model interpretability: to interpret the optimal model and quantify feature contributions, SHapley Additive exPlanations (SHAP) analysis was performed, providing both global feature importance rankings and local explanations for individual predictions. A two-tailed significance level of $\alpha = 0.05$ was applied.

3. Results

3.1 Clinical and Imaging Baseline Characteristics

Between 1 January 2021, and 30 June 2024, a total of 571 patients with 880 breast nodules meeting the inclusion and exclusion criteria were collected from Institution 1, including 117 malignant nodules (13.3%). The training set comprised 534 benign and 81 malignant nodules, while the internal validation set included 229 benign and 36 malignant nodules. From Institution 2, 333 patients with 351 nodules (230 benign, 121 malignant) were collected as the external test set.

A total of 1033 radiomic features were extracted from ultrasound images of these breast nodules using PyRadiomics, including 33 shape features, 198 first-order statistics, 792 texture features, and 744 frequency-domain features. Some overlap existed between texture and frequency-domain features, yielding 1033 unique features. To ensure robustness and reproducibility, inter-observer agreement was assessed using the intraclass correlation coefficient (ICC). Features with good reproducibility (ICC >0.80) were retained. Accordingly, 796 stable features were selected for subsequent analysis, while 237 features (ICC ≤0.80) were excluded. After feature selection using *t*-tests, Spearman correlation analysis, and LASSO regression, the final numbers of non-zero coefficient radiomic features selected for the T and peritumoral areas (denoted as E3/E5/EI3/EI5/ET3/ET5/TI3/TI5) were 21, 18, 22, 20, 17, 25, 21, 25, and 19, respectively. Following feature selection, the ultrasound radiomic features exhibited low-to-moderate correlations ($|r| < 0.6$), confirming minimal residual collinearity. The selected optimal features were subsequently used to construct the ultrasound radiomic models.

3.2 Diagnostic Performance of Multiple Machine Learning Models

Ultrasound radiomics models were constructed using multiple machine learning algorithms, including random forest, XGBoost, LightGBM, SVM, MLP, Gradient Boosting, KNN, logistic regression, extra trees, and AdaBoost. We systematically compared the performance of these models across different ROI datasets (T, EI3, EI5, E3, E5, ET3, ET5, TI3, TI5). Overall, SVM demonstrated stable and well-rounded performance. Although SVM showed slightly lower sensitivity in some groups (test set sensitivity ranged from approximately 0.165 to 0.422), its specificity was generally high (approximately 0.944–0.987), effectively reducing the false-positive rate. Metrics that balance sensitivity and specificity, such as accuracy and F1 score, indicated that SVM achieved a good trade-off and robustness in both the validation and test sets. For instance, when applied to the EI3, ET5, and TI3 feature sets, it achieved test set AUCs of 0.875, 0.849, and 0.861, respectively, coupled with consistently high specificity (0.961–0.983). This consistent performance highlights the reliability of the SVM classifier in this radiomics task, although it did not always yield the absolute highest AUC in every group (as shown in Table 1, where other algorithms such as LR and ET performed better in specific groups). Detailed results are presented in Tables 1,2 (Fig. 4).

3.3 Comparison of Model Performance Across Different ROI Groups

After a comprehensive comparison of multiple machine learning models, the SVM model consistently demonstrated superior performance across cross-validation, independent validation, and external test sets. Consequently, we further evaluated the classification performance of the SVM model across different ROI groups. Compared with the T group (tumor region), the EI3 group (intratumoral–peritumoral 3 pixels transition zone) exhibited notably better performance. In the external test set, the EI3 group achieved a significantly higher AUC than the T group (0.875 vs. 0.787; $p < 0.01$), indicating its enhanced ability to discriminate between positive and negative cases. Additionally, the EI3 group showed an accuracy of 78.1%, sensitivity of 42.2%, and specificity of 97.0%, with a Brier score of 0.166, surpassing the other groups. In comparison, the T group achieved an accuracy of 71.2%, sensitivity of 24.0%, and specificity of 96.1% (Table 2; Fig. 5).

In terms of model calibration, the EI3_SVM model outperformed the T_SVM model, with a lower Brier score (0.166 vs. 0.238), indicating greater calibration accuracy and more reliable predictive performance. Taken together, the EI3_SVM model demonstrated superiority across multiple metrics, including AUC, accuracy, specificity, and calibration, highlighting its robust generalization and predictive stability. Decision curve analysis further revealed that both the EI3 and TI3 models provided higher net benefit than the T group across most threshold probabilities, particularly at lower decision thresholds, underscoring their potential clinical utility.

Table 1. Diagnostic performance of the optimal model in each group.

Model	Group	AUC (95% CI)	Accuracy	Sensitivity	Specificity	F1 score
T_SVM	Internal validation	0.837 (0.767, 0.900)	0.849	0.556	0.895	0.86
	External test	0.787 (0.731, 0.840)	0.712	0.240	0.961	0.66
E3_SVM	Internal validation	0.837 (0.758, 0.910)	0.860	0.667	0.891	0.87
	External test	0.804 (0.756, 0.849)	0.698	0.165	0.978	0.62
E5_SVM	Internal validation	0.840 (0.760, 0.908)	0.857	0.583	0.900	0.86
	External test	0.772 (0.717, 0.822)	0.704	0.248	0.944	0.65
EI3_SVM	Internal validation	0.927 (0.882, 0.963)	0.872	0.806	0.882	0.88
	External test	0.875 (0.837, 0.910)	0.781	0.422	0.970	0.76
EI5_ET	Internal validation	0.763 (0.669, 0.848)	0.800	0.528	0.843	0.82
	External test	0.866 (0.823, 0.904)	0.784	0.413	0.978	0.76
ET3_SVM	Internal validation	0.876 (0.808, 0.937)	0.879	0.694	0.908	0.89
	External test	0.828 (0.780, 0.869)	0.752	0.306	0.987	0.71
ET5_LR	Internal validation	0.856 (0.763, 0.938)	0.887	0.806	0.900	0.90
	External test	0.869 (0.826, 0.906)	0.809	0.570	0.935	0.80
TI3_ET	Internal validation	0.850 (0.785, 0.907)	0.834	0.611	0.869	0.85
	External test	0.869 (0.827, 0.905)	0.781	0.413	0.974	0.75
TI5_ET	Internal validation	0.865 (0.801, 0.922)	0.834	0.722	0.852	0.85
	External test	0.872 (0.833, 0.913)	0.778	0.405	0.974	0.75

AUC, area under the curve; CI, confidence interval; T, tumor core ROI; E3, 3 pixels peritumoral ROI; E5, 5 pixels peritumoral ROI; EI3, 3 pixels transitional zone ROI; EI5, 5 pixels transitional zone ROI; ET3, 3 pixels outer peritumoral margin ROI; ET5, 5 pixels outer peritumoral margin ROI; TI3, 3 pixels inner margin ROI; TI5, 5 pixels inner margin ROI; ET, extra trees.

Table 2. Diagnostic performance of the SVM model for different intratumoral and peritumoral categories.

Model (SVM)	Group	AUC (95% CI)	Accuracy	Sensitivity	Specificity	F1 score
T_SVM	Internal validation	0.837 (0.767, 0.900)	0.849	0.556	0.895	0.86
	External test	0.787 (0.731, 0.840)	0.712	0.240	0.961	0.66
E3_SVM	Internal validation	0.837 (0.758, 0.910)	0.860	0.667	0.891	0.87
	External test	0.804 (0.756, 0.849)	0.698	0.165	0.978	0.62
E5_SVM	Internal validation	0.840 (0.760, 0.908)	0.857	0.583	0.900	0.86
	External test	0.772 (0.717, 0.822)	0.704	0.248	0.944	0.65
EI3_SVM	Internal validation	0.927 (0.882, 0.963)	0.872	0.806	0.882	0.88
	External test	0.875 (0.837, 0.910)	0.781	0.422	0.970	0.76
EI5_SVM	Internal validation	0.757 (0.661, 0.843)	0.819	0.444	0.878	0.83
	External test	0.806 (0.753, 0.853)	0.766	0.380	0.970	0.74
ET3_SVM	Internal validation	0.876 (0.808, 0.937)	0.879	0.694	0.908	0.89
	External test	0.828 (0.780, 0.869)	0.752	0.306	0.987	0.71
ET5_SVM	Internal validation	0.838 (0.734, 0.918)	0.857	0.667	0.887	0.87
	External test	0.849 (0.803, 0.891)	0.772	0.413	0.961	0.75
TI3_SVM	Internal validation	0.852 (0.779, 0.912)	0.834	0.611	0.869	0.85
	External test	0.861 (0.816, 0.902)	0.749	0.306	0.983	0.71
TI5_SVM	Internal validation	0.860 (0.783, 0.925)	0.891	0.722	0.917	0.90
	External test	0.841 (0.794, 0.885)	0.729	0.298	0.957	0.69

To systematically assess differences in model performance, DeLong's test was used to compare the AUCs of each model with the T group in the test set. Models incorporating peritumoral information consistently showed significant advantages: the EI3 group ($p < 0.01$) and TI3 group ($p < 0.01$) significantly outperformed the T group. After Bonferroni correction for multiple comparisons, the superiority of the EI3 and TI3 models remained statistically significant (adjusted $p < 0.01$), confirming the robust and reliable contribution of peritumoral features to improved diagnostic performance.

3.4 Feature Importance Analysis of EI3_SVM Using SHAP

SHAP value analysis of the EI3_SVM model revealed that its high specificity in the EI3 group (intratumoral–peritumoral 3 pixels transition zone) is primarily driven

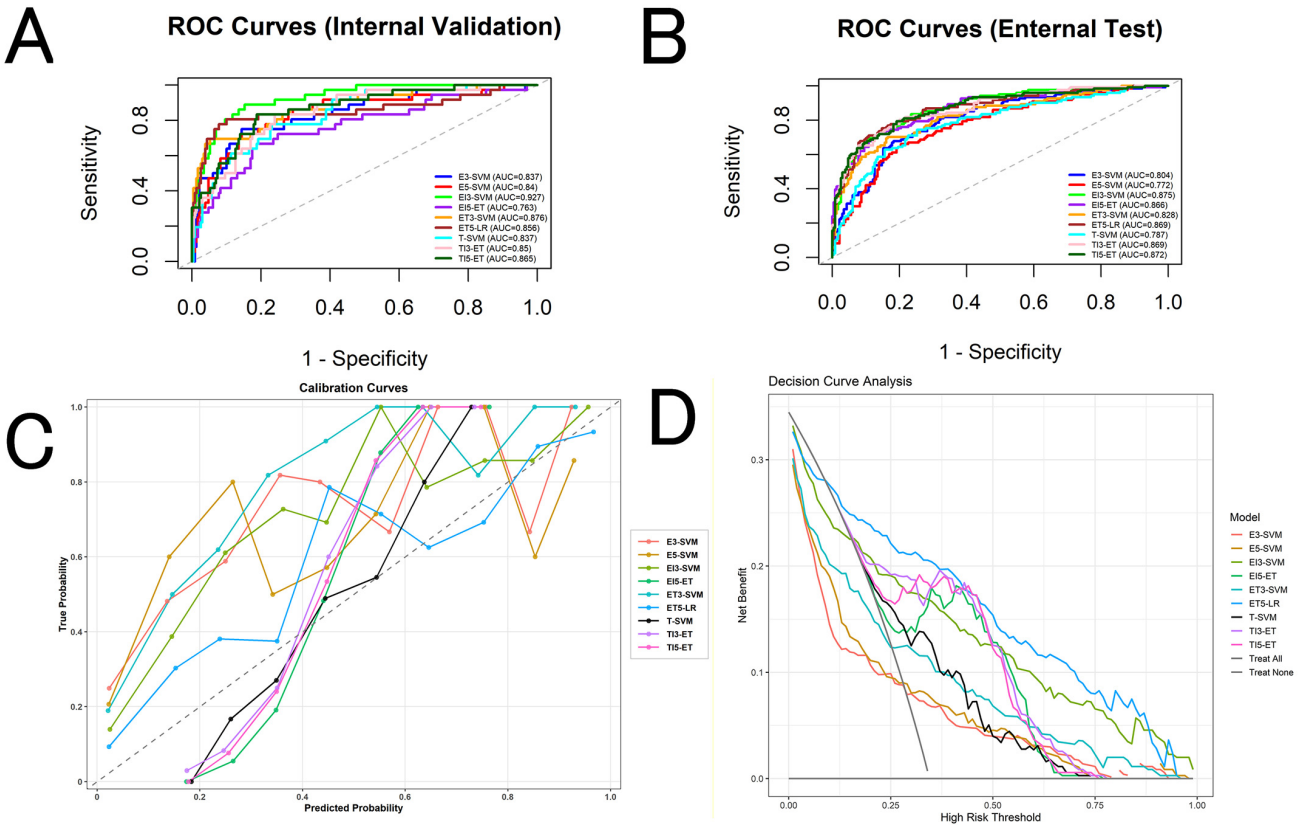

Fig. 4. Performance comparison of optimal machine learning models across different ROI groups. Notes: Comparison of the receiver operating characteristic (ROC) of optimal machine learning models for differentiating breast nodules across different tumor ROI models in the internal validation set (A) and external test set (B). Calibration curves (C) and decision curve analysis (DCA) curves (D) of the optimal machine learning models in each group in the external test set.

by sensitivity to texture and shape features (Fig. 6). Specifically, the model relies heavily on peritumoral texture features, such as gradient_glcm_Imc1 and gradient_ngtdm_Busyness. Tumor morphological features, including original_shape2D_MajorAxisLength and original_shape2D_Elongation, also contribute significantly to its decisions. Texture features derived from GLCM, NGTDM, and GLRLM quantify the complexity and heterogeneity of intralesional gray-level distribution, whereas shape features—such as Major Axis Length and Elongation—encode information on lesion size and contour irregularity. The model's sensitivity to these features likely underlies its ability to reduce false positives and accurately identify non-lesional tissue in high-specificity tasks. Importantly, the ranking of these predictive features corresponds closely with clinically recognized malignant ultrasound signs, such as spiculated margins and irregular shapes, thereby enhancing the interpretability of the model and its alignment with established diagnostic criteria.

4. Discussion

With the widespread adoption of screening and advances in ultrasound technology, the detection rate of breast nodules has been increasing year by year [11]. Although the use of the BI-RADS system improves diagnostic accuracy, it is highly subjective and relies on the expertise of the ultrasound physician. The diagnostic performance of the physician largely depends on individual experience, which introduces considerable subjectivity [12]. This diagnostic uncertainty highlights the need for more objective assessment methods to accurately differentiate benign from malignant nodules classified as ACR BI-RADS 3–4. Benign and malignant breast nodules often share certain features in their ultrasound imaging appearances [13].

Moreover, medical images often contain abundant information that reflects the biological behavior of tumors, which is difficult for the human eye to discern. Radiomics, as an emerging non-invasive technique, can extract more objective and precise quantitative features from medical images through high-throughput computation. It enables the use of objective, quantitative data, such as tumor heterogeneity, to characterize tumor phenotypes, providing valuable information for tumor diagnosis and treatment [14]. In Hong et al.'s study [15], the radiomics model designed to identify BI-RADS 4–5 breast nodules demonstrated an AUC of 0.886 in the test cohort. In fact, some BI-RADS 3–4 nodules are challenging to identify using only intratumoral radiomics. Incorporating peritumoral features im-

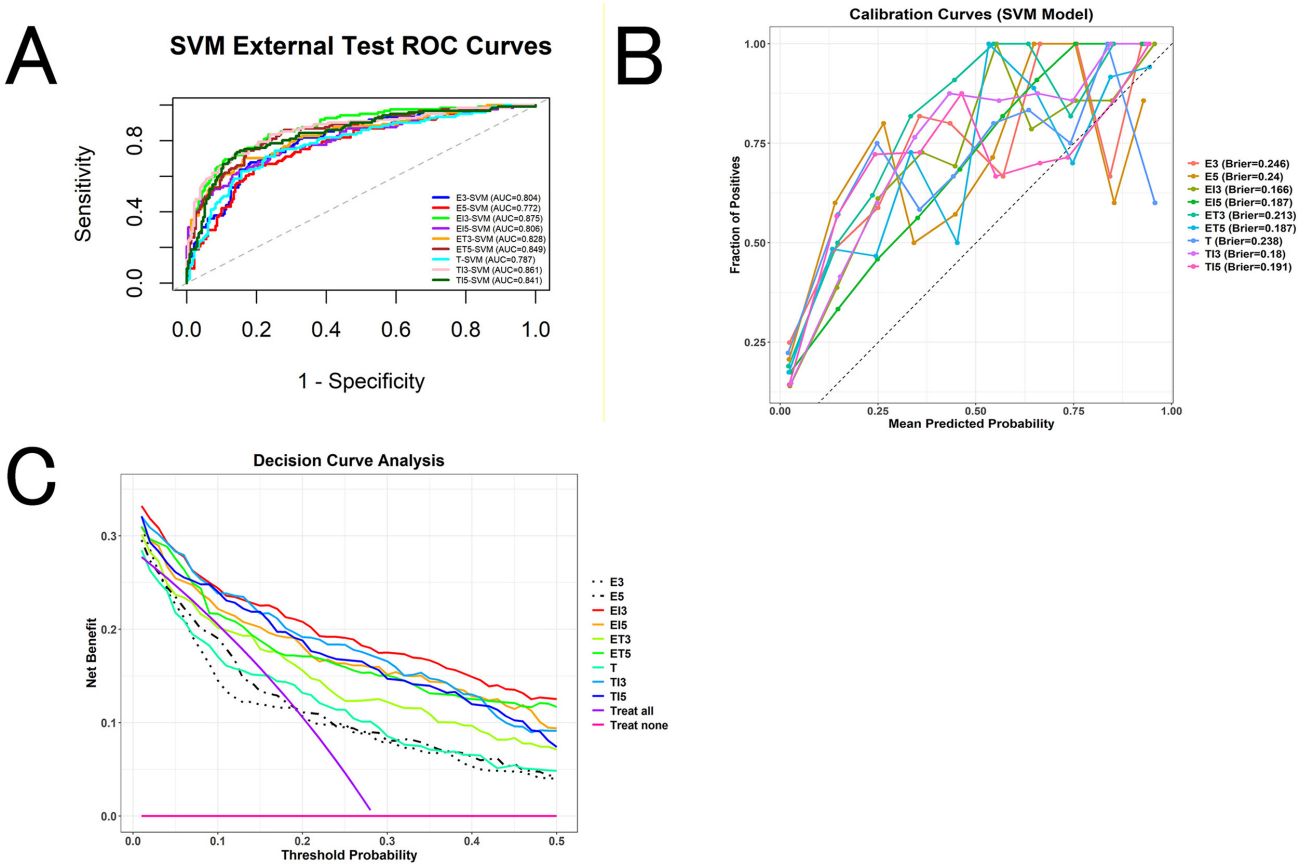

Fig. 5. Performance comparison of SVM-based models across different ROI groups. Notes: Receiver operating characteristic (ROC) curves of SVM-based models using different tumor ROI models for differentiating breast nodules in the external test set (A). Calibration curves (B) and decision curve analysis (DCA) curves (C) of the SVM-based models with different tumor ROI models in the external test set.

proves detection [16]. Previous radiomics studies of breast tumors have largely focused on intratumoral imaging features, with relatively little attention paid to the peritumoral region. However, studies using MRI-based radiomics have shown that the peritumoral area may harbor critical biological information, such as stromal responses, immune cell infiltration, and invasion of peritumoral lymphatic vessels and blood vessels [17]. Wu et al. [18] reported associations between peritumoral heterogeneity and certain gene features related to immune cell recruitment and inflammation. In the context of lung computed tomography, intratumoral and peritumoral texture heterogeneity features have been linked to immune responses at both molecular and morphological levels [19,20]. Yao et al. [21] found that the RF-based Peritumoral and Intratumoral Ultrasound Radiomics Signatures (P-IURS) model achieved high predictive accuracy in the human epidermal growth factor receptor 2 $(HER2)^{+}$ subtype, suggesting that integrating intratumoral and peritumoral ultrasound radiomics can improve the prediction of axillary pathologic complete response (PCR) after neoadjuvant chemotherapy (NAC) in breast cancer. A previous study has also demonstrated that incorporating peritumoral radiomic features derived from dilated ROIs (e.g., 2–8 mm beyond the tumor margin) improves model performance for predicting clinically relevant outcomes compared with intratumoral features alone, likely because these extended regions contain additional biologically relevant information [22]. Therefore, peritumoral ultrasound radiomics features may provide valuable complementary information to intratumoral radiomics.

In this study, we explored the optimal peritumoral range by expanding or contracting the tumor region by 3 or 5 pixels along the tumor boundary. The results showed that the radiomics model based on the EI3_SVM group achieved the best diagnostic performance for distinguishing benign from malignant breast nodules, significantly outperforming the model based solely on the T. This model improved specificity (reducing unnecessary biopsies), with a sensitivity of 42.2%. The EI3_SVM model achieved 78.1% accuracy and 97.0% specificity on the test set, demonstrating its substantial potential for this purpose. These findings indicate that the peritumoral region is crucial for differentiating benign from malignant breast nodules, and that combining the intratumoral margin (3 pixels) and the peritumoral margin (3 pixels), i.e., the EI3 region, is the most suitable. This region captures both intratumoral ultrasound features

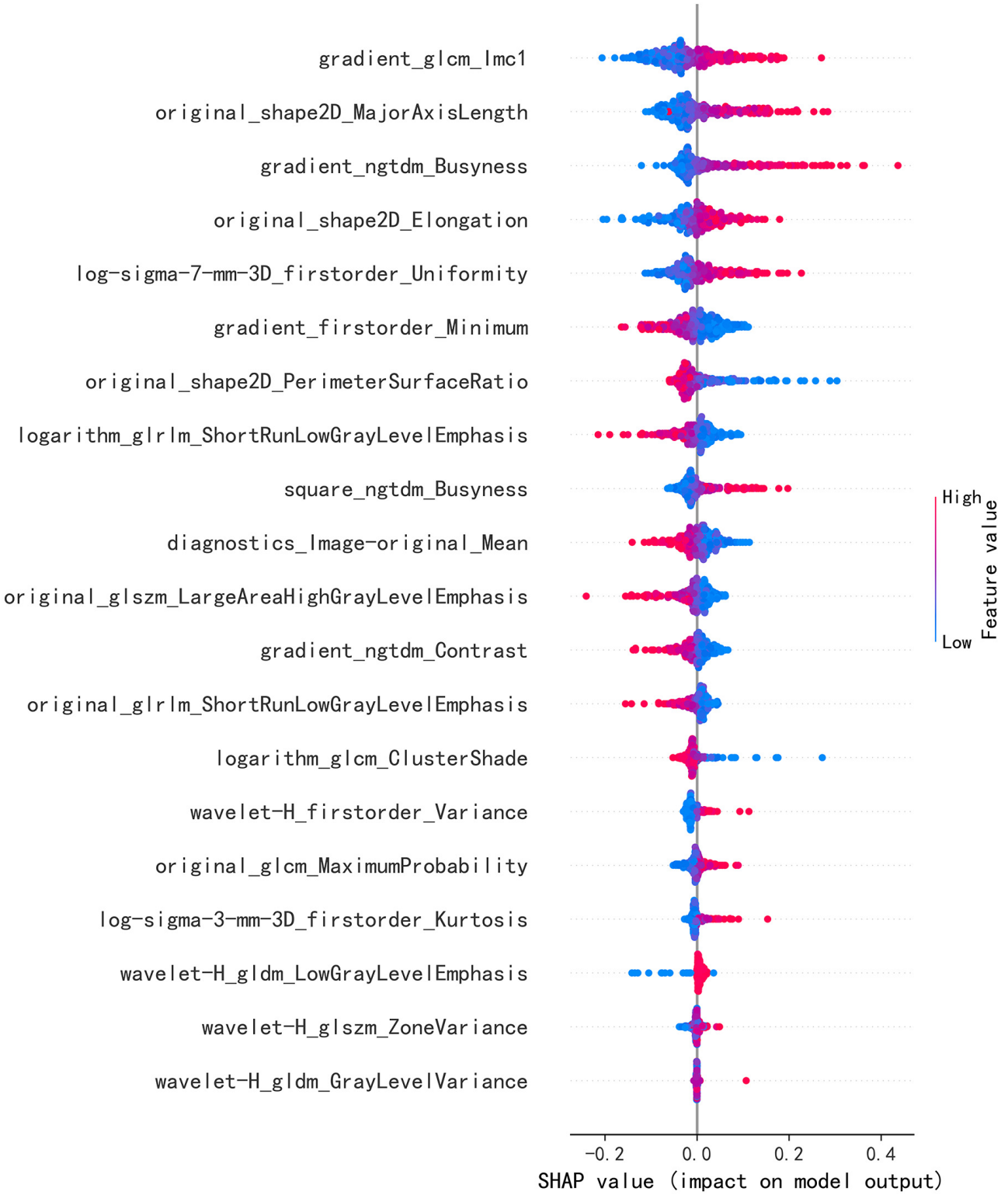

Fig. 6. SHapley Additive exPlanations (SHAP) beeswarm plot of the EI3_SVM model on the external test set.

and distinctive peritumoral characteristics. In contrast, a 5-pixel peritumoral expansion may be overly large, including more normal breast tissue, which could be detrimental to radiomics analysis. Compared with the optimal EI3 region, the TI3, TI5, and ET5 margins may not cover sufficient tissue affected by invasion, providing less comprehensive information for radiomic evaluation, although they still outperform the T model, confirming the value of peritumoral features. An intriguing finding is that the E3 model, which includes the entire intratumoral region, exhibits lower diag-

nostic performance than the EI3 model, which incorporates only 3 pixels each of intratumoral and peritumoral regions. Notably, as shown in Table 2, several other peritumoral-inclusive models (e.g., ET5, TI3, TI5) also outperformed E3, with ET5 achieving the second-highest AUC (0.849) after EI3. This suggests that simply expanding the ROI outward without refining the intratumoral component may introduce redundant information that dilutes critical signals from the tumor periphery. Our analysis suggests that the tumor core in the E3 model occupies a disproportionately large area, potentially diluting the critical signal from the peripheral region and even introducing substantial "informational noise". In contrast, the EI3 region simultaneously captures the most active peripheral intratumoral zone and the earliest microenvironmental changes in the peritumoral region, providing the model with "dual evidence" for assessing tumor invasiveness, thereby achieving superior discriminative performance. Guo et al. [23] attempted a radiomics study of benign and malignant breast lesions using dual-modality imaging combining conventional breast ultrasound and strain elastography. The model incorporating intratumoral, peritumoral, and parenchymal features (In&Peri&P) achieved the highest diagnostic performance. In summary, our study confirms the value of peritumoral features in radiomics and demonstrates that the extent of the peritumoral region has a significant impact on the predictive outcomes of radiomic analysis.

This SVM model distinguishes benign from malignant breast nodules by integrating multidimensional features from both intratumoral and peritumoral regions for a comprehensive assessment. The model primarily relies on peritumoral texture features (gradient_glcm_Imc1 and gradient_ngtdm_Busyness), which reflect the complexity of the textural structure. Additionally, tumor shape irregularity, extreme signals in the peritumoral region, intratumoral gray-level heterogeneity (kurtosis), and intensity information also contribute significantly to the classification. This finding is consistent with the study by Singh et al. [24], which showed that combining texture and geometric features can significantly improve classification performance. It is noteworthy that the class imbalance in the training data (benign-to-malignant ratio $\approx$ 7:1) was addressed using the Synthetic Minority Oversampling Technique (SMOTE). In medical imaging analysis, positive cases (the minority class) often carry greater clinical significance than negative cases, as the consequences of a missed diagnosis are generally more severe than those of a false positive. SMOTE was applied exclusively within each fold of cross-validation on the training set, while the validation and test sets retained their original imbalanced distributions. This approach ensures that the performance evaluation reflects the model's robustness and clinical applicability in real-world scenarios where class imbalance is inherent. It is important to clarify the intended clinical application of our high-specificity model. Rather than serving as a standalone screening tool, the model is designed as an adjunctive decision-support tool for BI-RADS 4 lesions. In current clinical practice, BI-RADS 4 lesions routinely undergo biopsy due to the inability to confidently exclude malignancy. The high specificity of our model (97.0%) means that when the model classifies a nodule as benign, the probability of it being truly benign is extremely high, allowing clinicians to safely avoid biopsy in a subset of patients who would otherwise undergo unnecessary invasive procedures. We fully recognize that the trade-off for this high specificity is a lower sensitivity (42.2%), meaning some malignant nodules may be missed. However, within the intended use context—as an adjunctive tool for BI-RADS 4 lesions—nodules classified as benign by the model could be considered for short-interval follow-up rather than immediate biopsy, while nodules with indeterminate or malignant predictions would proceed to biopsy as per standard care. This approach can significantly reduce unnecessary biopsies while maintaining the safety net of clinical follow-up for cases that are not definitively benign. Notably, the use of SMOTE may also contribute to the model's moderate sensitivity observed in the external test set, a point that warrants further optimization in future work.

As a retrospective, dual-center study, the core data that could be uniformly and completely obtained from all participating institutions were pathological results and ultrasound images, which ensured the validity of modeling and analysis. However, records of certain baseline clinical information (such as patient age and detailed morphological descriptions of nodules) were incomplete or inconsistently formatted across different centers. Consequently, we were unable to perform a comprehensive comparability analysis of baseline characteristics among the training, internal validation, and external test sets. To minimize potential selection bias, we adopted the following research design: (1) samples from the same center were strictly divided in a 7:3 ratio following the principle of randomization; (2) data from a completely independent center were used to constitute the external test set, with the specific aim of evaluating the model's generalizability in real-world scenarios where distributional differences may exist. Future prospective studies should systematically collect and report all baseline data. Furthermore, the models in this study, particularly the optimal EI3_SVM model, demonstrated high specificity (97.0%) but moderate sensitivity (42.2%) on the external test set. This performance profile suggests that it is more suitable for clinical scenarios aimed at "reducing false positives"—that is, serving as an auxiliary tool to avoid unnecessary biopsies for benign nodules—rather than functioning as a high-sensitivity screening tool. The limited sensitivity may stem from: (1) severe class imbalance in the training data (benign:malignant $\approx$ 7:1); (2) the use of conventional grayscale ultrasound only, without incorporating functional information such as elastography; and (3) model optimization that prioritized high specificity.

This study also has several limitations. First, it is inherently retrospective, and the possibility of selection bias cannot be excluded; future prospective studies are needed to validate our model. Second, the intratumoral and peritumoral ROIs were analyzed only in two dimensions, ignoring the three-dimensional characteristics of tumors, which may somewhat limit the model's predictive capability. Third, advanced techniques such as shear-wave elastography and contrast-enhanced ultrasound were not incorporated. Fourth, the calibration curves of some models (Fig. 5B) exhibited deviations and fluctuations, likely attributable to the limited sample size of the external test set, particularly the modest number of malignant cases, leading to instability in probability estimates. Future studies with larger, prospective cohorts are required to optimize model calibration. Finally, in this study, peritumoral radiomic features were used solely to predict benign versus malignant nodules. Other important prognostic factors, such as lymph node metastasis, molecular subtypes, chemotherapy response, and recurrence rate, were not considered. Future studies should incorporate these diverse clinical and pathological features for further validation.

5. Conclusion

This study explored the application of intratumoral and peritumoral features in radiomics analysis of tumors. Our findings confirm that peritumoral features contain critical information relevant to the tumor itself and that the choice of peritumoral region size significantly affects the predictive accuracy of radiomics models. Our study highlights the importance of optimizing peritumoral feature extraction in future radiomics research to enhance overall predictive performance, with the potential to improve clinical diagnostic accuracy and reduce the rate of unnecessary biopsies.

Key Points

• Incorporating peritumoral radiomic features significantly enhances the diagnostic objectivity and accuracy for BI-RADS 3–4 breast nodules, effectively reducing reliance on subjective physician interpretation.

• The SVM model with a 3-pixel intratumoral-peritumoral-transition zone achieved a high specificity of 97.0% in external testing, demonstrating strong potential for avoiding unnecessary biopsies.

• Model interpretability via SHAP analysis confirms that decisions are driven by clinically meaningful features such as peritumoral texture and tumor shape irregularity, aligning with established malignant ultrasound signs.

• The width of the peritumoral region substantially influences model performance, highlighting the importance of optimizing this parameter in future radiomics research on breast ultrasound.

Abbreviations

BI-RADS, Breast Imaging Reporting and Data System; ROI, region of interest; ICC, intraclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; SMOTE, Synthetic Minority Oversampling Technique; SVM, support vector machine; SHAP, SHapley Additive exPlanations; AUC, area under the curve; DCA, decision curve analysis.

Availability of Data and Materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions

XRY contributed to the collection, analysis, interpretation, and drafting of the manuscript, as well as to data analysis. JXZ was involved in data acquisition, analysis, and review of the submitted version. HHZ and CPL contributed to data acquisition. PX participated in the study design, data acquisition and analysis, data interpretation, reviewing the submitted version, and supervision. All authors contributed to important intellectual revisions of the manuscript. All authors have read and approved the final version of the manuscript. All authors were fully involved in the work and agreed to be accountable for all aspects of the study.

Ethics Approval and Consent to Participate

This study was conducted in accordance with the Declaration of Helsinki. This dual-center observational study employed a retrospective cohort design and was approved by the Ethics Committee of the Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine (No. SH9H-2025-T462-1). Considering its retrospective nature and that all data have been anonymized, the Ethics Committee of the Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine exempted the requirement for patients' informed consent.

Acknowledgment

The authors would like to acknowledge the support and assistance of the Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine and Shanghai Second People's Hospital.

Funding

This work was supported by the Discipline Development Project of Medical Imaging, the Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine (XKQJS202403).

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a Cancer Journal for Clinicians. 2024; 74: 229–263. https://doi.org/10.3322/caac.21834

[2] Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nature Medicine. 2025; 31: 1154–1162. https://doi.org/10.1038/s41591-025-03502-3

[3] Wei Q, Yan YJ, Wu GG, Ye XR, Jiang F, Liu J, et al. The diagnostic performance of ultrasound computer-aided diagnosis system for distinguishing breast masses: a prospective multicenter study. European Radiology. 2022; 32: 4046–4055. https://doi.org/10.1007/s00330-021-08452-1

[4] Giess CS, Frost EP, Birdwell RL. Difficulties and errors in diagnosis of breast neoplasms. Seminars in Ultrasound, CT, and MR. 2012; 33: 288–299. https://doi.org/10.1053/j.sult.2012.01.007

[5] Hong M, Fan S, Xu Z, Fang Z, Ling K, Lai P, et al. MRI radiomics and biological correlations for predicting axillary lymph node burden in early-stage breast cancer. Journal of Translational Medicine. 2024; 22: 826. https://doi.org/10.1186/s12967-024-05619-4

[6] Magnuska ZA, Roy R, Palmowski M, Kohlen M, Winkler BS, Pfeil T, et al. Combining Radiomics and Autoencoders to Distinguish Benign and Malignant Breast Tumors on US Images. Radiology. 2024; 312: e232554. https://doi.org/10.1148/radiol.232554

[7] Li SY, Li YM, Fang YQ, Jin ZY, Li JK, Zou XM, et al. Contrast-enhanced ultrasound radiomics model for predicting axillary lymph node metastasis and prognosis in breast cancer: a multicenter study. BMC Cancer. 2025; 25: 1315. https://doi.org/10.1186/s12885-025-14632-9

[8] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144: 646–674. https://doi.org/10.1016/j.cell.2011.02.013

[9] Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nature Medicine. 2013; 19: 1423–1437. https://doi.org/10.1038/nm.3394

[10] Yu T, Yu R, Liu M, Wang X, Zhang J, Zheng Y, et al. Integrating intratumoral and peritumoral radiomics with deep transfer learning for DCE-MRI breast lesion differentiation: A multicenter study comparing performance with radiologists. European Journal of Radiology. 2024; 177: 111556. https://doi.org/10.1016/j.ejrad.2024.111556

[11] Berg WA, Vourtsis A. Screening Breast Ultrasound Using Handheld or Automated Technique in Women with Dense Breasts. Journal of Breast Imaging. 2019; 1: 283–296. https://doi.org/10.1093/jbi/wbz055

[12] Schwab F, Redling K, Siebert M, Schötzau A, Schoenenberger CA, Zanetti-Dällenbach R. Inter- and Intra-Observer Agreement in Ultrasound BI-RADS Classification and Real-Time Elastography Tsukuba Score Assessment of Breast Lesions. Ultrasound in Medicine & Biology. 2016; 42: 2622–2629. https://doi.org/10.1016/j.ultrasmedbio.2016.06.017

[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. Ultraschall in Der Medizin. 2020; 41: 390–396. https://doi.org/10.1055/a-0917-6825

[14] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016; 278: 563–577. https://doi.org/10.1148/radiol.2015151169

[15] 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. Frontiers in Oncology. 2022; 12: 894476. https://doi.org/10.3389/fonc.2022.894476

[16] Huang Z, Mo S, Wu H, Kong Y, Luo H, Li G, et al. Optimizing breast cancer diagnosis with photoacoustic imaging: An analysis of intratumoral and peritumoral radiomics. Photoacoustics. 2024; 38: 100606. https://doi.org/10.1016/j.pacs.2024.100606

[17] Kim ES, Ham S, Seo BK, Lee JY, Sun W, Jeon M, et al. Optimized prediction of breast cancer tumor microenvironment using MRI-based intratumoral and peritumoral radiomics: a prospective study. Frontiers in Oncology. 2025; 15: 1654508. https://doi.org/10.3389/fonc.2025.1654508

[18] Wu J, Li B, Sun X, Cao G, Rubin DL, Napel S, et al. Heterogeneous Enhancement Patterns of Tumor-adjacent Parenchyma at MR Imaging Are Associated with Dysregulated Signaling Pathways and Poor Survival in Breast Cancer. Radiology. 2017; 285: 401–413. https://doi.org/10.1148/radiol.2017162823

[19] Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. eLife. 2017; 6: e23421. https://doi.org/10.7554/eLife.23421

[20] Beig N, Khorrami M, Alilou M, Prasanna P, Braman N, Orooji M, et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. Radiology. 2019; 290: 783–792. https://doi.org/10.1148/radiol.2018180910

[21] Yao J, Jia X, Zhou W, Zhu Y, Chen X, Zhan W, et al. Predicting axillary response to neoadjuvant chemotherapy using peritumoral and intratumoral ultrasound radiomics in breast cancer subtypes. iscience. 2024; 27: 110716. https://doi.org/10.1016/j.isci.2024.110716

[22] Ding J, Chen S, Serrano Sosa M, Cattell R, Lei L, Sun J, et al. Optimizing the Peritumoral Region Size in Radiomics Analysis for Sentinel Lymph Node Status Prediction in Breast Cancer. Academic Radiology. 2022; 29: S223–S228. https://doi.org/10.1016/j.acra.2020.10.015

[23] Guo S, Huang X, Xu C, Yu M, Li Y, Wu Z, et al. Multiregional radiomic model for breast cancer diagnosis: value of ultrasound-based peritumoral and parenchymal radiomics. Quantitative Imaging in Medicine and Surgery. 2023; 13: 3127–3139. https://doi.org/10.21037/qims-22-939

[24] Singh H, Sharma V, Singh D. Comparative analysis of proficiencies of various textures and geometric features in breast mass classification using k-nearest neighbor. Visual Computing for Industry, Biomedicine, and Art. 2022; 5: 3. https://doi.org/10.1186/s42492-021-00100-1