Appendix 1

Radiomic features

(I) Shape features (2D): in this group of features we included descriptors of the two-dimensional size and shape of the region of interest (ROI).

(II) First-order statistics: first-order statistics describe the distribution of voxel intensities within the image region defined by the mask through commonly used and basic metrics.

(III) Textural features: textural features describe the spatial distribution of voxel intensities and are calculated from five typical matrices—the gray level co-occurrence matrix (GLCM), the gray-level run-length matrix (GLRLM), the gray-level size zone matrix (GLSZM), the gray-level dependence matrix (GLDM), and the neighborhood gray-tone difference matrix (NGTDM).

(IV) Wavelet-based features: wavelet-based features were derived from wavelet decompositions of the original ultrasound (US) images using the "Coiflet 1" wavelet function. Each image was filtered using either a high band-pass filter (H) or low-band pass filter (L) in x and y directions, yielding 4 different combinations of decompositions. The wavelet decompositions of the original image X were labeled as XLL, XLH, XHL, XHH.

The feature algorithms can also be found online (https://pyradiomics.readthedocs.io/en/latest/features.html).

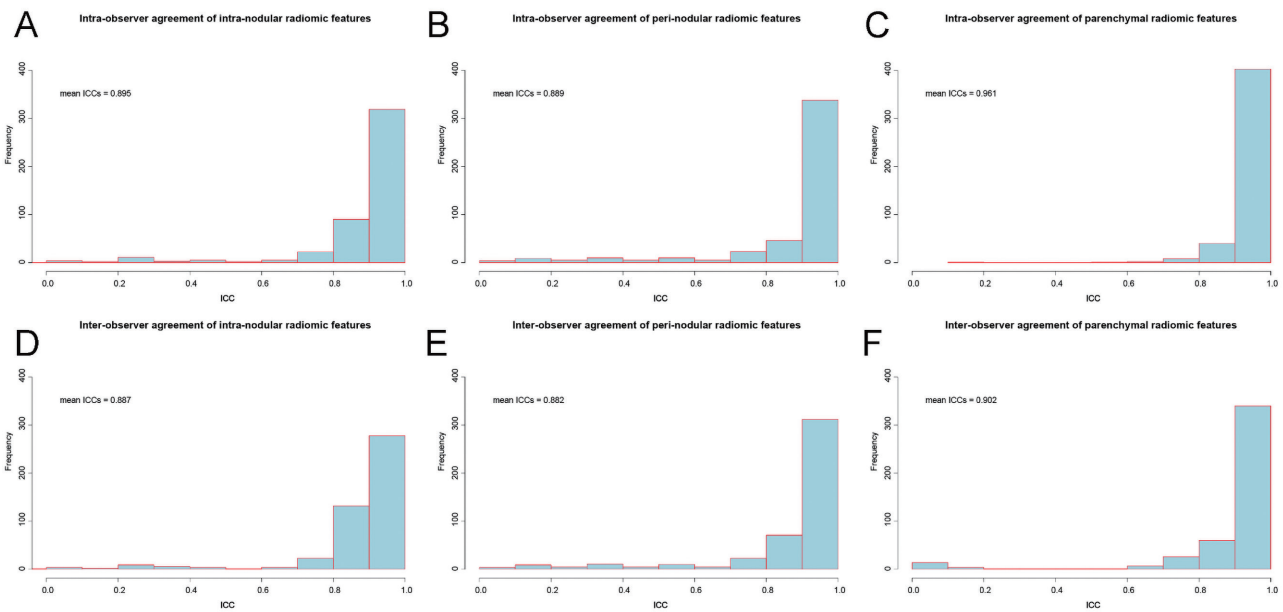

Figure S1 Histograms of ICCs used to assess the intra-observer and inter-observer reproducibility of intratumoral (A,D), peritumoral (B,E) and parenchymal (C,F) radiomic features. ICCs, intraclass correlation coefficients.

https://dx.doi.org/10.21037/qims-22-939

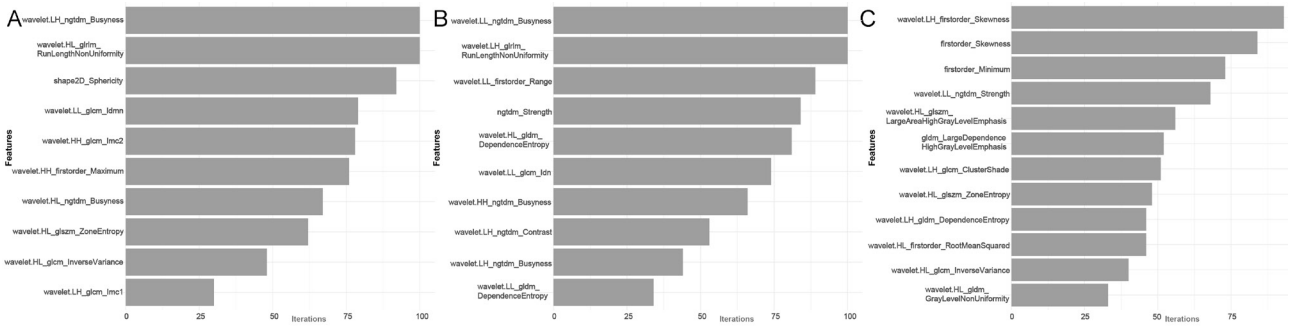

Figure S2 Frequency of radiomic feature selection in intratumoral (A), peritumoral (B), and parenchymal (C) regions.

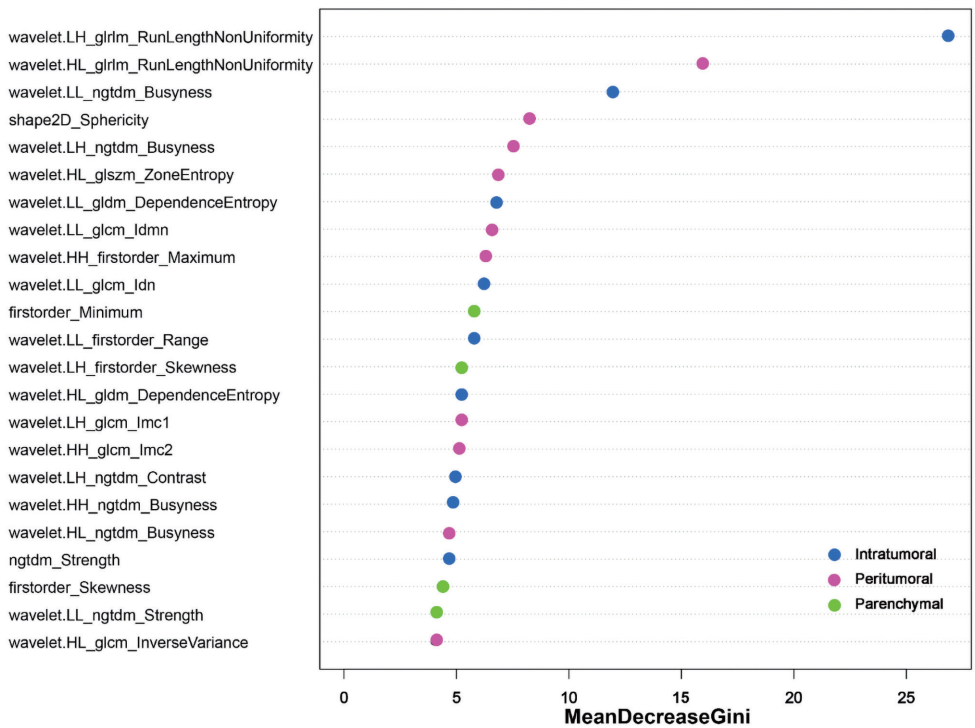

Figure S3 Importance of radiomic features of the In&Peri&P model in the random forest classifier process. In&Peri&P = intratumoral & peritumoral & parenchymal model.

https://dx.doi.org/10.21037/qims-22-939

Table S1 Ultrasound acquisition parameters of institutions

Institution	Manufacturer	Machine	Probe
1	Mindray	Resona 7	3–11 linear array probe
	Mindray	DC-8	3–12 linear array probe
	Philips	EPIQ 7	5–12 linear array probe
	Philips	iU Elite	3–9 linear array probe
	Siemens	Sequoia	6–18 linear array probe
	SuperSonic	Aixplorer	4–15 linear array probe
	TOSHIBA	Aplio 500	5–14 linear array probe
2	GE	LOGIQ E9	6–15 linear array probe
	Siemens	S3000	6–18 linear array probe
	SuperSonic	Aixplorer	4–15 linear array probe
	TOSHIBA	Aplio 500	5–14 linear array probe

Table S2 R packages used in this study

Functions	R package
AUC and Delong test	pROC
Bar diagrams, box diagrams, ROC curves, and waterfall plots	ggplot2, ROCR
Intra/interclass correlation	irr
Calibration curves	rms
Hosmer-Lemeshow test	ResourceSelection
Index IDI	PredictABEL
Random forest	randomForest
Spearman's rank correlation	cor
Recursive feature elimination	caret
Boruta	Boruta
Proportions comparison	epiR, STAT

AUC, area under the ROC curve; IDI, integrated discrimination improvement; ROC, receiver operator characteristic.

Table S3 Selected radiomic features of models

Model	Selected features
Intratumoral	wavelet.HL_glrlm_RunLengthNonUniformity, wavelet.LH_ngtdm_Busyness, shape2D_Sphericity, wavelet.LL_glcm_Idmn, wavelet.HH_glcm_Imc2, wavelet.HH_firstorder_Maximum, wavelet.HL_ngtdm_Busyness, wavelet.HL_glszm_ZoneEntropy, wavelet.HL_glcm_InverseVariance, wavelet.LH_glcm_Imc1
Peritumoral	wavelet.LH_glrlm_RunLengthNonUniformity, wavelet.LL_ngtdm_Busyness, wavelet.LL_firstorder_Range, ngtdm_Strength, wavelet.HL_gldm_DependenceEntropy, wavelet.LL_glcm_Idn, wavelet.HH_ngtdm_Busyness, wavelet.LH_ngtdm_Contrast, wavelet.LL_gldm_DependenceEntropy
Parenchymal	wavelet.LH_firstorder_Skewness, firstorder_Skewness, firstorder_Minimum, wavelet.LL_ngtdm_Strength, wavelet.HL_firstorder_RootMeanSquared
In&Peri	Intratumoral features: wavelet.HL_glrlm_RunLengthNonUniformity, wavelet.LH_ngtdm_Busyness, shape2D_Sphericity, wavelet.LL_glcm_Idmn, wavelet.HH_glcm_Imc2, wavelet.HH_firstorder_Maximum, wavelet.HL_ngtdm_Busyness, wavelet.HL_glszm_ZoneEntropy, wavelet.HL_glcm_InverseVariance, wavelet.LH_glcm_Imc1
	Peritumoral features: wavelet.LH_glrlm_RunLengthNonUniformity, wavelet.LL_ngtdm_Busyness, wavelet.LL_firstorder_Range, ngtdm_Strength, wavelet.HL_gldm_DependenceEntropy, wavelet.LL_glcm_Idn, wavelet.HH_ngtdm_Busyness, wavelet.LH_ngtdm_Contrast, wavelet.LL_gldm_DependenceEntropy
In&P	Intratumoral features: wavelet.HL_glrlm_RunLengthNonUniformity, wavelet.LH_ngtdm_Busyness, shape2D_Sphericity, wavelet.LL_glcm_Idmn, wavelet.HH_glcm_Imc2, wavelet.HH_firstorder_Maximum, wavelet.HL_ngtdm_Busyness, wavelet.HL_glszm_ZoneEntropy, wavelet.HL_glcm_InverseVariance, wavelet.LH_glcm_Imc1
	Parenchymal features: wavelet.LH_firstorder_Skewness, firstorder_Skewness, firstorder_Minimum, wavelet.LL_ngtdm_Strength
In&Peri&P	Intratumoral features: wavelet.HL_glrlm_RunLengthNonUniformity, wavelet.LH_ngtdm_Busyness, shape2D_Sphericity, wavelet.LL_glcm_Idmn, wavelet.HH_glcm_Imc2, wavelet.HH_firstorder_Maximum, wavelet.HL_ngtdm_Busyness, wavelet.HL_glszm_ZoneEntropy, wavelet.HL_glcm_InverseVariance, wavelet.LH_glcm_Imc1
	Peritumoral features: wavelet.LH_glrlm_RunLengthNonUniformity, wavelet.LL_ngtdm_Busyness, wavelet.LL_firstorder_Range, ngtdm_Strength, wavelet.HL_gldm_DependenceEntropy, wavelet.LL_glcm_Idn, wavelet.HH_ngtdm_Busyness, wavelet.LH_ngtdm_Contrast, wavelet.LL_gldm_DependenceEntropy
	Parenchymal features: wavelet.LH_firstorder_Skewness, firstorder_Skewness, firstorder_Minimum, wavelet.LL_ngtdm_Strength

In&Peri: intratumoral & peritumoral model; In&P = intratumoral & parenchymal model; In&Peri&P = intratumoral & peritumoral & parenchymal model.

Table S4 Performance of radiomic models in the differentiation of malignant from benign breast lesions using the training cohort

Model	Sen (%)	Spe (%)	Acc (%)	AUC
Intratumoral	77 [70–83]	82 [75–88]	79 [75–84]	0.855 [0.813–0.896]
Peritumoral	74 [67–80]	86 [79–91]	79 [75–84]	0.858 [0.819–0.898]
Parenchymal	75 [68–81]	56 [47–64]	66 [61–71]	0.689 [0.633–0.746]
In&Peri	76 [69–82]	88 [81–92]	81 [77–85]	0.884 [0.848–0.919]
In&P	85 [80–90]	70 [62–77]	78 [74–83]	0.867 [0.829–0.906]
In&Peri&P	85 [79–90]	83 [76–89]	84 [80–88]	0.921 [0.895–0.948]

Data in parentheses are 95% confidence intervals. In&Peri = intratumoral & peritumoral model; In&P = intratumoral & parenchymal model; In&Peri&P = intratumoral & peritumoral & parenchymal model. Acc, accuracy; AUC, area under the receiver operator characteristic curve; Sen, sensitivity; Spe, specificity.

https://dx.doi.org/10.21037/qims-22-939

Table S5 Calibration of radiomic models evaluated by the Hosmer-Lemeshow test

Cohort	Hosmer-Lemeshow P value					
	Intratumoral	Peritumoral	Parenchymal	In&Peri	In&P	In&Peri&P
Training cohort	0.641	0.016	0.996	0.994	0.037	0.070
Internal test cohort	0.087	0.544	0.085	0.282	0.065	0.175
External test cohort	0.407	0.437	0.002	0.149	0.074	0.069

In&Peri = intratumoral & peritumoral model; In&P = intratumoral & parenchymal model; In&Peri&P = intratumoral & peritumoral & parenchymal model.

Table S6 Univariate and multivariate analyses for the Rad-score and clinical factors in the differentiation of malignant from benign breast lesions

Characteristics	Univariate analysis		Multivariate analysis	
	OR (95% CI)	P value	OR (95% CI)	P value
Training cohort (n=339)				
Age (years)	1.10 (1.07–1.13)	<0.001	1.08 (1.04–1.13)	<0.001
Lesion size (mm)	1.06 (1.04–1.08)	<0.001	0.99 (0.95–1.02)	0.394
Tumor location (outer quadrant/subareolar)	2.24 (1.42–3.57)	<0.001	0.82 (0.34–1.93)	0.654
Distance from the nipple (mm)	1 (0.99–1.01)	0.985	1.00 (0.97–1.02)	0.698
BI-RADS category	5.33 (3.82–7.79)	<0.001	3.25 (2.14–5.28)	<0.001
Rad-score	3,384.39 (803.31–17,347.60)	<0.001	1,762.40 (283.57–14,383.87)	<0.001
Internal test cohort (n=146)				
Age (years)	1.10 (1.06–1.14)	<0.001	1.13 (1.06–1.22)	0.001
Lesion size (mm)	1.05 (1.02–1.08)	0.002	1.00 (0.94–1.06)	0.993
Tumor location (outer quadrant/subareolar)	1.11 (0.55–2.26)	0.770	0.64 (0.14–2.63)	0.541
Distance from the nipple (mm)	0.98 (0.96–1.00)	0.036	0.98 (0.94–1.01)	0.233
BI-RADS category	3.85 (2.56–6.31)	<0.001	2.20 (1.27–4.09)	0.007
Rad-score	2,760.87 (1,719.16–9,139.61)	<0.001	1,863.07 (708.16–14,439.58)	< 0.001
External test cohort (n=106)				
Age (years)	1.10 (1.07–1.13)	<0.001	1.08 (1.04–1.13)	<0.001
Lesion size (mm)	1.06 (1.04–1.08)	<0.001	0.99 (0.95–1.02)	0.392
Tumor location (outer quadrant/subareolar)	2.24 (1.42–3.57)	<0.001	0.82 (0.34–1.93)	0.657
Distance from the nipple (mm)	1.00 (0.99–1.01)	0.985	1.00 (0.97–1.02)	0.695
BI-RADS category	5.33 (3.82–7.79)	<0.001	3.25 (2.14–5.28)	<0.001
Rad-score	2,755.22 (478.98–22,964.10)	<0.001	1,792.02 (287.39–14,688.18)	<0.001

BI-RADS, Breast Imaging Reporting and Data System; CI, confidence interval; OR, odds ratio; Rad-score, radiomic signature score.

Table S7 Diagnostic performance of the Rad-BI-RADS category in the differentiation of malignant from benign breast lesions using the internal and external test cohorts

Model	Internal test cohort (n=146)			External test cohort (n=106)		
	Acc (%)	Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)
BI-RADS	73 [65–80]	96 [89–99]	39 [27–51]	76 [67–84]	97 [90–100]	46 [29–63]
Rad-score	82 [75–88]*	82 [72–90]*	82 [71–90]*	80 [71–87]	84 [73–92]*	73 [56–86]*
Rad-BI-RADS category	83 [76–89]#	93 [86–99]	71 [58–81]#	86 [78–92]#	90 [80–97]	75 [59–85]#

Data in parentheses are 95% confidence intervals. *, significantly different from the BI-RADS category and Rad-score; #, significantly different from the BI-RADS category and Rad-BI-RADS category. Acc, accuracy; BI-RADS, Breast Imaging Reporting and Data System; Rad-score, radiomic signature score; Sen, sensitivity; Spe, specificity.

https://dx.doi.org/10.21037/qims-22-939