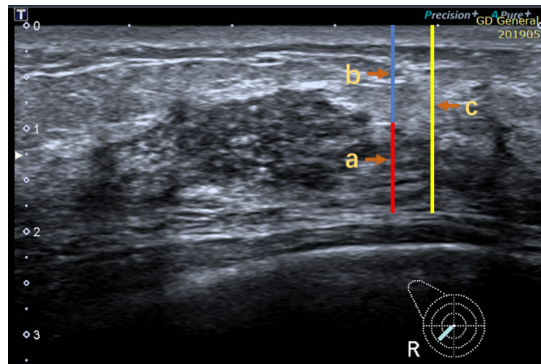


## Appendix 1

According to the criteria of the ACR BI-RADS fifth edition and our previous studies (37,38), the detailed descriptions of imaging features are as follows.

### Breast BEP

- (I) BEP contains fat and fibro-glandular parenchyma (homogeneous and heterogeneous). Homogeneous echotexture consists of fat or fibroglandular homogeneity. Heterogeneity can be either focal or diffuse. The breast echotexture is characterized by multiple small areas of increased and decreased echogenicity and may be accompanied by shadowing.
- (II) The qualitative parameters of breast parenchyma and fat adjacent to lesions were measured in the maximal diameter of the tumor's image. The mean diameter was based on 3 measurements. Each mean diameter should be obtained based on three measurements as follows.



$RPT = a/c$ . RPT is the ratio (the largest thickness of TBP/thickness of pectoralis fascia).  $RPF = a/b$ . RPF is the ratio (the thickness of TBP/thickness of mammary fat).

### Shape

- (I) Oval: a mass that is elliptical or egg-shaped (may include 2 or 3 undulations; i.e., gently lobulated or macrolobulated).
- (II) Round: a round mass is one that is spherical, ball-shaped, circular, or globular. It has an anteroposterior diameter equal to its transverse diameter; to qualify as a round mass, it must be circular in perpendicular projections.
- (III) Irregular: this lesion shape is neither round nor oval.

### Orientation

- (I) Parallel: the long axis of the mass parallels the skin line. Masses that are only slightly obliquely oriented might be considered parallel.
- (II) Not parallel: the long axis of the mass is not parallel to the skin line. The anterior-posterior or vertical dimension is greater than the transverse or horizontal dimension. These masses can also be obliquely oriented to the skin line. Round masses are not parallel in their orientation.

### Margin

- (I) Circumscribed: a circumscribed margin is one that is well defined, with an abrupt transition between the lesion and the

surrounding tissue. For a mass to be described as circumscribed at US, its entire margin must be sharply defined. Most circumscribed lesions have round or oval shapes.

- (II) Indistinct: there is no clear demarcation of the entire margin or any portion of the margin from the surrounding tissue.
- (III) Angular: some or all of the margin has sharp corners, often forming acute angles.
- (IV) Microlobulated: the margin is characterized by short-cycle undulations, but the significant feature is that the margin of the mass is not circumscribed.

### ***Echogenic pattern***

- (I) Hypoechoic: the term “hypoechoic” is defined relative to subcutaneous fat; hypoechoic masses, which are less echogenic than fat, are characterized by low-level echoes throughout.
- (II) Isoechoic: isoechogenicity is defined as having the same echogenicity as subcutaneous fat.
- (III) Hyperechoic: hyperechogenicity is defined as having increased echogenicity relative to fat or equal to fibroglandular tissue.
- (IV) Complex cystic and solid: a complex mass contains both anechoic (cystic or fluid) and echogenic (solid) components.
- (V) Heterogeneous: a mixture of echogenic patterns within a solid mass.

### ***Posterior features***

- (I) No posterior features: no shadowing or enhancement is present deep to the mass; the echogenicity of the area immediately behind the mass is not different from that of the adjacent tissue at the same depth.
- (II) Enhancement sound transmission: there is no impediment in the passage through the mass. Enhancement appears as a column that is more echogenic (whiter) deep to the mass.
- (III) Shadowing: there is attenuation of the acoustic transmission. Sonographically, the area posterior to the mass appears darker.
- (IV) Combined pattern: these are lesions with more than one pattern of posterior attenuation.

### ***Calcifications***

- (I) Calcifications in a mass.
- (II) Calcifications outside of a mass: calcifications situated in fat or fibro-glandular tissue.
- (III) Intraductal calcifications.
- (IV) No calcification.

### ***Vascularity distribution***

- (I) Absent: cysts are the most common avascular lesions. Some solid masses also have little or no vascularity.
- (II) Vessels in rim: the blood vessels may be marginal, forming part or all of a rim around a mass.
- (III) Internal vascularity: blood vessels are present within the mass. Vessels may penetrate the margin of the mass or display an orderly or disorderly pattern within the mass.

### ***Vascularity grade (39)***

- (I) Grade I: nonvascularity.
- (II) Grade II: less than 1 mm in diameter.
- (III) Grade III: a main vessel is visible in the area, and/or several small vessels are visible.
- (IV) Grade IV: four or more vessels are visible.

### *Reported lymph node metastasis*

- (I) No: US doctor did not report lymph node metastasis.
- (II) Yes: US doctor reported lymph node metastasis.

### *Tumor size*

The largest diameter is the only feature used. The largest measurement should represent the longest axis of a lesion.

## **Appendix 2**

The hyperparameters used for RF, SVM, KNN, MLP, and RR.

RF: n\_estimators (200, test in [50, 100, 150, 200])  
max\_depth (5, test in [2, 3, 4, 5, 6])  
n\_features (none, since we want to test all features)

SVM: c (1, test in [0.1, 1, 10])  
Kernel ('rbf', test in ['linear', 'poly', 'rbf', 'sigmoid'])  
degree (3, test in [2, 3])

KNN: n\_neighbors (40, test in [3, 5, 10, 20, 30, 40, 50])  
weights ('uniform', test in ['uniform', 'distance'])

MLP: hidden\_layer\_sizes (128, test in [32, 64, 128])  
activation ('relu', test in ['logistic', 'relu'])

RR: alpha (1.6, test in [0, 0.1, 0.2, ..., 1.9, 2.0])

Parameters not mentioned above were set as the default values.

## **Appendix 3**

### *Pairwise comparison of clinical and ultrasonic characteristics*

Characteristics	HIS	B1	B2	B3	B5
Age	B1		*	*	*
	B2	*		*	*
	B3	*	*		*
	B5	*	*	*	
Height	B1				
	B2				*
	B3				
	B5			*	

(continued)

(continued)

Characteristics	HIS	B1	B2	B3	B5
Weight	B1				
	B2				*
	B3				
	B5		*		
BMI	B1		*		*
	B2	*		*	*
	B3		*		*
	B5	*	*	*	
Echo pattern	B1				
	B2				*
	B3				
	B5		*		
Shape	B1		*		*
	B2	*			*
	B3				
	B5	*	*		
Margin	B1				*
	B2				*
	B3				*
	B5	*	*	*	
Orientation	B1				
	B2				*
	B3				
	B5		*		
Posterior feature	B1				*
	B2				*
	B3				*
	B5	*	*	*	
Calcification	B1				*
	B2				*
	B3				*
	B5	*	*	*	

(continued)

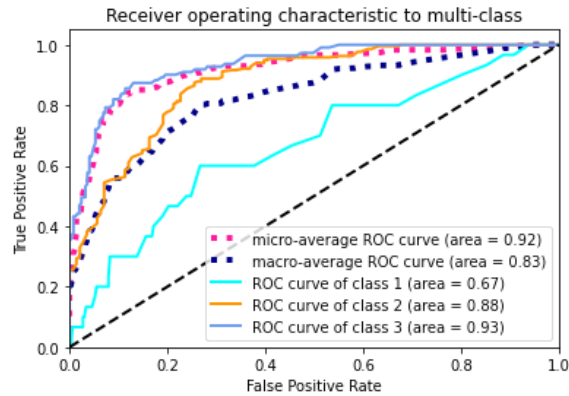
(continued)

Characteristics	HIS	B1	B2	B3	B5
Vascularity distribution	B1		*	*	*
	B2	*		*	*
	B3	*	*		
	B5	*	*		
Vascularity grade	B1		*		*
	B2	*			*
	B3				*
	B5	*	*	*	
Lymph node metastasis reported	B1				*
	B2				*
	B3				*
	B5	*	*	*	
BEP	B1				
	B2				
	B3				
	B5				
BI-RADS category	B1		*		*
	B2	*			*
	B3				*
	B5	*	*	*	
Tumor size	B1		*	*	*
	B2	*			*
	B3	*	*		*
	B5	*	*	*	
TBP	B1				*
	B2				*
	B3				*
	B5	*	*	*	
RPT	B1				
	B2				*
	B3				
	B5			*	
RPF	B1				
	B2				*
	B3				
	B5			*	

\*, P<0.05. HIS, actual histological result; BMI, body mass index; BEP, background echotexture of parenchyma; BI-RADS, Breast Imaging Reporting and Data System; TBP, anteroposterior thickness of breast parenchyma; RPT, thickness ratio of breast parenchyma to tissue before pectoralis fascia; RPF, thickness ratio of breast parenchyma to mammary fat.

## Appendix 4

ROC curves, areas under the curve, and accuracy of the RF model with overrepresentation of B1 and B3 to the same class (class 1).



PRE: class 1 (B1 + B3), class 2 (B2), class 3 (B5), AUCs, accuracy.

HIS: class 1 (B1 + B3), [0, 26, 4, 0.67], 0.81

class 2 (B2) [0, 200, 15, 0.88]

class 3 (B5) [0, 24, 87, 0.93]

## References

37. Liang T, Cong S, Yi Z, Liu J, Huang C, Shen J, Pei S, Chen G, Liu Z. Ultrasound-Based Nomogram for Distinguishing Malignant Tumors from Nodular Sclerosing Adenoses in Solid Breast Lesions. *J Ultrasound Med* 2021;40:2189-200.
38. Liang T, Shen J, Zhang S, Cong S, Liu J, Pei S, Shang S, Huang C. Using Ultrasound-Based Multilayer Perceptron to Differentiate Early Breast Mucinous Cancer and its Subtypes From Fibroadenoma. *Front Oncol* 2021;11:724656.
39. Adler DD, Carson PL, Rubin JM, Quinn-Reid D. Doppler ultrasound color flow imaging in the study of breast cancer: preliminary findings. *Ultrasound Med Biol* 1990;16:553-9.