Magn Reson Med Sci 2022; 21; 485–498
doi:10.2463/mrms.mp.2020-0160 Published Online: June 26, 2021

MAJOR PAPER

Combined Use of Texture Features and Morphological Classification Based on Dynamic Contrast-enhanced MR Imaging: Differentiating Benign and Malignant Breast Masses with High Negative Predictive Value

Shigeharu Ohyu[1*], Mitsuhiro Tozaki[2], Michiro Sasaki[3], Hisae Chiba[4], Qilin Xiao[5], Yasuko Fujisawa[1], and Yoshiaki Sagara[2]

Purpose: We evaluated the diagnostic performance of the texture features of dynamic contrast-enhanced (DCE) MRI for breast cancer diagnosis in which the discriminator was optimized, so that the specificity was maximized via the restriction of the negative predictive value (NPV) to greater than 98%.

Methods: Histologically proven benign and malignant mass lesions of DCE MRI were enrolled retrospectively. Training and testing sets consist of 166 masses (49 benign, 117 malignant) and 50 masses (15 benign, 35 malignant), respectively. Lesions were classified via MRI review by a radiologist into 4 shape types: smooth (S-type, 34 masses in training set and 8 masses in testing set), irregular without rim-enhancement (I-type, 60 in training and 14 in testing), irregular with rim-enhancement (R-type, 56 in training and 22 in testing), and spicula (16 in training and 6 in testing). Spicula were immediately classified as malignant. For the remaining masses, 298 texture features were calculated using a parametric map of DCE MRI in 3D mass regions. Masses were classified into malignant or benign using two thresholds on a feature pair. On the training set, several feature pairs and their thresholds were selected and optimized for each mass shape type to maximize specificity with the restriction of NPV > 98%. NPV and specificity were computed using the testing set by comparison with histopathologic results and averaged on the selected feature pairs.

Results: In the training set, 27, 12, and 15 texture feature pairs are selected for S-type, I-type, and R-type masses, respectively, and thresholds are determined. In the testing set, average NPV and specificity using the selected texture features were 99.0% and 45.2%, respectively, compared to the NPV (85.7%) and specificity (40.0%) in visually assessed MRI category-based diagnosis.

Conclusion: We, therefore, suggest that the NPV of our texture-based features method described performs similarly to or greater than the NPV of the MRI category-based diagnosis.

Keywords: *breast cancer, magnetic resonance imaging, benign mass, texture feature, diagnostic performance*

Introduction

Dynamic contrast-enhanced (DCE) MRI has a high sensitivity (88%–92%) and reasonable specificity (67%–77%) in the diagnosis of breast lesions.[1,2] For the differentiation of benign and malignant breast lesions, a kinetic curve analysis of DCE MRI and texture-based features of kinetic maps have been used in recent studies.[3–5] Karahaliou et al.[3] applied texture features of DCE MRI for the diagnosis of breast lesions, and an area under receiver-operating characteristic curve (ROC-AUC) greater than 0.9 was obtained by assessing the heterogeneity of features. Gibbs et al.[4] reported a sensitivity of 56%–67% and specificity of 97%–100% in the characterization of sub-1 cm breast lesions using texture features of DCE MRI. They obtained a negative predictive value (NPV) of 89%–91% in subjects, where 70%

[1]Research and Development Center, Canon Medical Systems Corporation, Ohtawara, Tochigi, Japan
[2]Department of Radiology, Sagara Hospital, Kagoshima, Kagoshima, Japan
[3]Department of Radiology, Sagara Perth Avenue Clinic, Kagoshima, Kagoshima, Japan
[4]MRI Sales Department, Canon Medical Systems Corporation, Kawasaki, Kanagawa, Japan
[5]Research & Development Center, Canon Medical Systems (China) Co., Ltd, Beijing, China

*Corresponding Author: Research and Development Center, Canon Medical Systems Corporation, 1385, Shimoishigami, Ohtawara, Tochigi 324-8550, Japan. Phone: +81-287-26-6210, Fax: +81-287-26-6066, E-mail: shigeharu.ohyu@medical.canon

Received: November 9, 2020 | Accepted: May 14, 2021

had benign lesions and 30% malignant lesions. These studies suggest that texture-based features of DCE MRI have the potential to achieve high diagnostic performance in the differentiation of benign and malignant breast lesions.

In studies discriminating benign and malignant lesions using features extracted from DCE MRI, ROC-AUC is frequently used as the criterion for optimizing the discriminator, including the selection of features.[3,5] However, the discriminator with the highest ROC-AUC does not always have the best balance of sensitivity and specificity necessary for a decision on patient management.

We postulated that the criterion for the optimization of the discriminator including the feature selection should be consistent with the clinical requirement. Breast imaging reporting and data system (BI-RADS)[6] is widely used in the reporting of breast MRI. In BI-RADS MRI, category 3 (probably benign) is described as an examination with a likelihood of malignancy $> 0\%$ but $\leq 2\%$ and that a short-interval of patient follow-up is recommended for management. Examinations in categories 4 (suspicious) and 5 (highly suggestive of malignancy) are described as having a likelihood of malignancy $\geq 2\%$ and that an invasive tissue diagnosis (biopsy) is recommended. In a diagnostic examination, those lesions with a category 3 or less are classified as negative (benign decision), and categories 4 and 5 are classified as positive (malignant decision). Thus, the NPV of diagnostic examination should optimally be 98% or larger. A very high NPV is required in order to reduce the potential for a malignant lesion to be classified inaccurately as negative, so the decision to do an invasive examination is not selected.

For the optimization of a discriminator including feature selection, a reasonable combination of criteria consistent with clinical demand is high NPV and specificity. A higher NPV reduces the mis-classifications of malignant lesions in negatively classified cases. And higher specificity reduces the false positive in benign lesions; thus, it reduces unnecessary biopsy.

Assesments of both morphological feature and kinetic pattern are efficient in classification of MRI category.[6,7] Integration of morphological features and kinetic features into the discriminator would be efficient to improve the discrimination performance. However, interobserver variability of morphological and kinetic features needs to be considered. Grimm et al.[8] reported interobserver variability of BI-RADS MRI descriptors. Mass margin and mass shape features showed large agreement ($\kappa = 0.78$ and $\kappa = 0.72$) between breast imaging specialists, and lesion kinetic curve assessments, such as persistent, plateau, and washout, showed lower agreement ($\kappa = 0.17$). From their studies, we presumed that combination use of texture features of kinetic maps and visual classification of shape and margin features of masses may improve diagnostic performance with smaller affection of interobserver variability.

We proposed the benign and malignant discrimination method on texture features of DCE MRI in which the NPV and specificity are consistently used as the criteria of optimization, and where the features and thresholds are selected to maximize specificity with restriction of NPV greater than 98%. Visual classifications of MRI mass shape and margin type are incorporated into discrimination process. Our expectation was that the proposed criteria in optimization would make the higher NPV possible. Our goal was to assess the performance of the proposed method in differentiating benign and malignant masses in DCE MRI.

Materials and Methods

Subjects

The training set consisted of 189 histologically verified lesions selected in January 2019 from 179 breast DCE MRI examinations acquired between November 2013 and November 2018. After the exclusion of 23 lesions (11 non-mass lesions, 7 lesions not identified on MRI, 4 lesions without contrast enhancement, and 1 lesion of the nipple), 166 mass lesions (49 benign and 117 malignant) in 159 patients were used for the optimization of the discriminator (Table 1).

The testing set consisted of 50 histologically verified masses in 47 patient examinations selected, applying the same exclusion criteria from breast DCE MRI examinations acquired before August 2019 excluding the masses in training set. The acquisition period of DCE MR in testing set was from July 2017 to March 2019 (Table 1).

Lesions enrolled in this study for both the training and testing sets were lesions where cancer was suspected on either mammography or ultrasound, and where mass lesions were identified in DCE MRI with a size of 5 mm or larger. All were histologically proven. Lesions were retrospectively enrolled from sequentially acquired lesions in the acquisition periods. Benign and malignant lesions were enrolled separately so that the malignant lesions accounted for 70% of the total number of masses. To adjust the fraction of benign and malignant lesions, several lesions in latter periods were not included in the data set. In the patient enrollment for testing set, lesions selected for training set were not selected. Hence, no duplicated lesions are contained in the training set and the testing set. Minimum, maximum, mean, and median diameters of masses are 5, 51, 19.3, and 17.4 mm, respectively, in the training set, whereas they are 5, 60, 17.2, and 13.2 mm, respectively, in the testing set.

This retrospective study was conducted with the approval of our institutional review board. All the patients were provided with written informed consent for our review of their medical records and access to images. Our study was performed using retrospectively collected MR images obtained from outpatient clinical examinations.

MRI protocol

Two different MRI scanners were used for the acquisition. Different acquisition parameters were used across scanners. However, both applied standard breast DCE MRI sequences using a 3D-gradient echo. The first scanner was a 3T MRI (Biograph mMR; Siemens Healthcare, Erlangen, Germany); spatial resolution was 0.80–0.89 mm; slice interval was 0.9–1.2 mm; RE was 3.09–3.68 ms; TE was 1.32–1.38 ms; matrix

Table 1 Characteristics of lesions in the training set and testing set

Histology type		Number of mass lesions	
		Lesions in training set for feature selection and threshold optimization	Lesions in testing set for validation
Benign	Benign phyllodes tumor	14	4
	Fibroadenoma	19	6
	Papilloma	1	3
	Intraductal papilloma	3	1
	Intracystic papilloma	5	1
	Ductal adenoma	1	0
	Adenomyoepithelioma	2	0
	Granular cell tumor	1	0
	Subareolar sclerosing duct hyperplasia	1	0
	Undefined benign lesion	2	0
	Total benign	49	15
Malignant	Ductal carcinoma *in situ*	7	0
	Papillotubular carcinoma	14	4
	Solid-tubular carcinoma	28	9
	Scirrhous carcinoma	51	18
	Mucinous carcinoma	4	1
	Invasive lobular carcinoma	6	0
	Adenoid cystic carcinoma	0	1
	Apocrine carcinoma	1	0
	Undefined malignant lesion	6	2
	Total malignant	117	35
Total		166	50

size was 448 × 448; and 160–192 slices were acquired in the coronal plane to cover the breasts bilaterally. After the acquisition of pre-contrast image, gadobutorol (Gadvist 0.1 mmol/kg; Bayer Schering Pharma, Berlin, Germany) was used for the post-contrast sequences acquired at 5 time points (60s/120s/180s/240s/300s). The second scanner was a 1.5T MRI (Optima MR360; GE Healthcare, Chicago, IL, USA); spatial resolution was 0.59–0.68 mm; slice interval was 1.0 mm; RE was 5.1–6.0 ms; TE was 2.4–2.9 ms; matrix size was 512 × 512; and 152–196 slices were acquired in the trans-axial plane to cover both breasts bilaterally. Gadopentetate dimeglumine (Magnevist 0.1 mmol/kg; Bayer Schering Pharma, Berlin, Germany) was used for contrast enhancement. Acquisition time points were similar to those used on the first scanner. For the training set, two different MRI scanners were used for the acquisition. The training set consists of 98 lesions from the first scanner and 68 lesions from the second scanner. All the 44 lesions in the testing set are acquired using the first scanner.

Image review

The masses assessed in the DCE MRI scans were reviewed by a radiologist (M.T) with 26 years of experience in breast radiology and who was not involved in the patient enrollment. The MRI category was assessed individually for each mass. Categories 1, 2, and 3 were treated as negative (benign decision) according to the BI-RADS negative diagnostic examination. Categories 4 and 5 were treated as positive (malignant decision). The masses were then classified into four mass shape types: S (smooth), I (irregular-without rim-enhancement), R (irregular with rim-enhancement), and spicula (Fig. 1). Tozaki et al. introduced the classification of mass shape and margin.[7] They classified masses into smooth margin, irregular lesion, and spiculated margin. We used similar classification; however, irregular lesion is further divided into two types by the presence or absence of rim-enhancement. In our method, the oval- or round-shaped masses with circumscribed margin masses were classified as S-type. In the masses with irregular shapes or

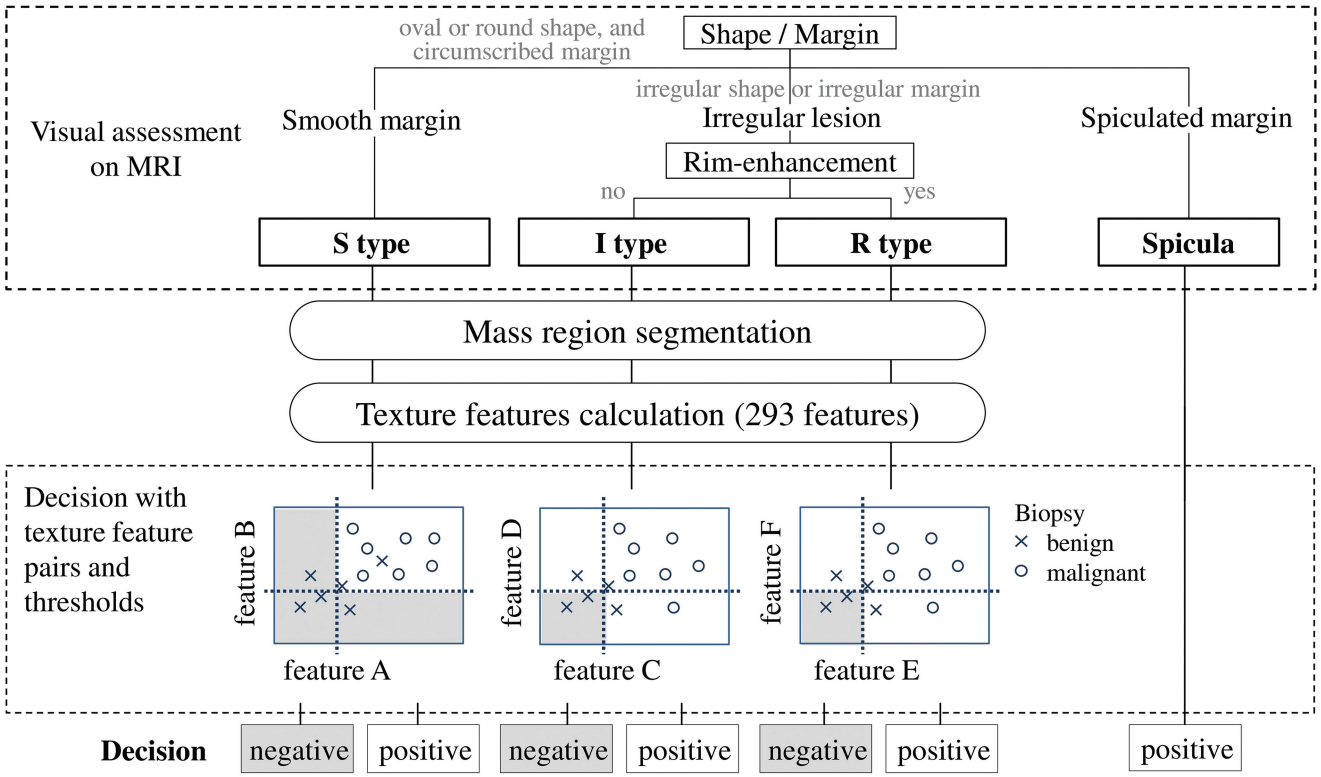

Fig. 1 Discrimination scheme of masses using texture values of DCE parametric maps. Masses in testing set are visually classified into S-type, I-type, and R-type on DCE MRI. Texture features (298) are calculated in extracted mass region. Two features and corresponding thresholds are used for the discrimination of negative (benign) or positive (malignant). Different feature pairs are applied for S-, I-, and R-type masses. Feature pairs for S-, I-, and R-types and corresponding thresholds are determined in the training set. DCE, dynamic contrast-enhanced; I, irregular without rim-enhancement; R, irregular with rim-enhancement; S, smooth.

irregular margins, masses without rim-enhancement were classified as I-type and masses with rim-enhancement were classified as R-type. The masses that had spiculated margins were immediately classified as positive, and they were excluded from case samples for following analysis.

Image processing

In-house software was used for the execution of image processing steps: motion correction, mass region extraction, calculation of DCE kinetic maps, and the calculation of texture features. The reviewer used the software for both image review and processing. Masses seen in DCE MRI were identified by a reviewer. No other part of the medical record was referred to throughout the image review except for the lesion position. Mass regions were extracted three-dimensionally using the semi-automatic mass region extraction tool. The software generates five candidate mass regions by applying five different thresholds of enhancement ratios. The reviewer determined one mass region from five candidate regions by choosing a region, which is similar to contrast-enhanced region of the mass. Rigid motion correction was applied if large motion was found around the mass in DCE MRI. DCE parametric maps, initial enhancement (IE),[3] post initial enhancement (PIE),[3] maximum slope (MS), peak percentage enhancement (PPE)[3], signal enhancement ratio (SER)[3], peak time (PT), maximum enhancement in initial phase (MEIP), subsequent enhancement to initial peak (SEIP), were generated. All of the maps are calculated as

$$\mathrm{IE}(\%) = 100 \times \frac{S_1 - S_0}{S_0},\ \mathrm{PIE}(\%) = 100 \times \frac{S_{last} - S_1}{S_1},$$

$$\mathrm{MS}(\%/\mathrm{min}) = 100 \times \max \frac{S_{i+1} - S_i}{S_0 \times (T_{i+1} - T_i)},$$

$$\mathrm{PPE}(\%) = 100 \times \frac{S_{peak} - S_0}{S_0},\ \mathrm{SER}(\%) = 100 \times \frac{S_1 - S_0}{S_{last} - S_0},$$

$$\mathrm{MEIP}(\%) = 100 \times \frac{S_{peak2} - S_0}{S_0},$$

$$\mathrm{SEIP}(\%) = 100 \times \frac{S_{last} - S_{peak2}}{S_{peak2}},$$

where S_0, S1, and S_{last} are signal intensity at pre-contrast phase, initial phase around 120 s, and last phase around 5

Table 2 Total 298 texture features calculated in detected mass region

Texture features of mass (298)	
Univariate statistics	14 formulas × 8 maps = 112 (mean, standard deviation, CV, skewness, kurtosis, Shannon index [entropy], Simpson index [energy], min, C5 [5 percentile], C25 [25 percentile], C50 [50 percentile], C75 [75 percentile], C95 [95 percentile], max) × (IE, MEIP, PPE, MS, PIE, SEIP, SER, PT)
Bivariate statistics	5 formulas × (4 maps × 4 maps + 2 maps × 2 maps) = 100 (covariance, correlation, Shannon index[entropy], Simpson index [energy], bivariate SD) × ([IE, MEIP, PPE, MS] × [PIE, SEIP, SER, PT], [PIE, SEIP] × [SER, PT])
3 × 3 histogram parameters of initial enhancement and post initial enhancement	31 features very_fast (percentage of IE ≥ 200%), typical_fast (200% > IE ≥ 100%), fast (IE ≥ 100%), medium (100% > IE ≥ 50%), slow(IE < 50%), washout (PIE < –10%), plateau (10% > PIE ≥ –10%), persistent (PIE ≥ 10%), very_fast-washout, very_fast-plateau, very_fast-persistent,typical_fast-washout, typical_fast-plateau, typical_fast-persistent, fast-washout, fast-plateau, fast-persistent, medium-washout, medium-plateau, medium-persistent, slow-washout, slow-plateau, slow-persistent, not_fast-not_washout, not_slow-washout, fast-not_persistent, not_slow-not_persistent, shannon 3 × 3, simpson 3 × 3, shannon 4 × 3, simpson 4 × 3
Shannon/Simpson combinations	8 formulas × 5 map combinations = 40 (shannon_multiply, shannon_ratio, shannon_diff, shannon_rel_diff, simpson_multiply, simpson_ratio, simpson_diff, simpson_rel_dif) × (IE_PIE, PPE_PIE, IE_SER, PPE_SER, SER_PIE)
Others	15 features Parameters from percentiles (C95_IEm 100 × C5_PIE, Max_IEm 100 × C5_PIE, C95_IEm 100 × Min_PIE, Max_IEm 100 × Min_PIE, C95_IE × C5_PIE, C75-C25 IE, C95-C5 IE, C75-C25 PPE, C95-C5 PPE, C75-C25 MS, C95-C5 MS, C75-C25 PIE, C95-C5 PIE, C75-C25 SER, C95-C5 SER)

C95_IEm 100 × C5_PIE = (C95_IE – 100) × C5_IE; Max_IEm 100 × C5_PIE = (max_IE–100) × C5_PIE; C95_IEm 100 × Min_PIE = (C95_IE–100) × min_PIE; Max_IEm 100 × Min_PIE = (max_IE–100) × min_PIE. CV, coefficient of variance; IE, initial enhancement; MEIP, maximum enhancement in initial phase; MS, maximum slope; PIE, post initial enhancement; PPE, peak percentage enhancement; PT, peak time; SD, standard deviation; SEIP, subsequent enhancement to initial peak; SER, signal enhancement ratio.

min; S_{peak} is maximum intensity; S_{peak2} is maximum intensity before initial phase; $(S_{i+1}-S_i)$ is intensity difference of adjoining phases; $T_{i+1}-T_i$ (min) is time between adjoining phases. PT (s) is defined as time to peak from injection. Then, 298 texture features of DCE parametric maps shown in Table 2 were calculated for each mass region. From single parametric maps, 112 univariate statistics are calculated by applying mean, standard deviation, and 12 other formulas to 8 maps. From 2 of 8 parametric maps, 100 bivariate statistics are calculated by 5 formulas and 20 combinations of parametric maps. From 3 × 3 cross-histogram of Initial Enhancement map and Post Initial Enhancement map in the mass region, 31 features are calculated. These features are frequency percentage of pixels in defined range of the cross-histogram. Remaining 55 features are defined as arithmetic of univariate statistics, bivariate statistics, and 3 × 3 cross-histogram features.

Masses were discriminated into negative (benign decision) or positive (malignant decision) by the application of two thresholds each corresponding with one of two features. Total 44253 = (298 × 297)/2 feature pairs were possible as the pair is a combination of 298 features. Eight different inequality determination formulas were able to be composed of two inequality comparison (less than, greater than) for the two features and two logical operations (AND, OR). Final selected inequality determination formulas were selected from the 44253 feature pairs and 8 inequality determination formulas in the optimization process described in the next section.

Optimization of discriminator on training set

In the optimization process, feature pairs were selected from 44253 pairs on training set. Feature pairs having larger specificity and NPV ≥ 98% were selected. Here, benign and malignant histopathologic results were referred in the calculation of specificity and NPV. Estimated 2D cumulative histograms were used in the evaluation of the NPV and specificity in the feature pair selection and threshold optimization. First, for benign masses in the training set, 2D histograms of a feature pair were estimated using a kernel density estimation with Gaussian kernels. A map of the true negative number of masses was obtained by integrating the 2D histogram of a feature pair of benign masses. Similarly, a map of true positive numbers was obtained by the integration of a 2D histogram of a feature pair of malignant masses. Next, the NPV map, specificity map, positive predictive value (PPV) map, and sensitivity map were calculated from the true negative map and true positive map as follows:

NPV = true negative/(true negative + false negative),
Specificity = true negative/(true negative + false positive),
PPV = true positive/(true positive + false positive),
Sensitivity = true positive/(true positive + false negative).

Table 3 Number of masses with respect to each shape type of mass and each MRI category classified on DCE MRI

Histology	MRI Category	Training set					Testing set				
		Shape type of mass					Shape type of mass				
		S	I	R	Spicula	Total	S	I	R	Spicula	Total
Benign	3	11	1	0	0	12	4		2		6
	4	12	15	9	0	36	3	3	3		9
	5	0	0	1	0	1					
		23	16	10	0	49	7	3	5		15
Malignant	3	2	0	0	0	2	1				1
	4	9	38	34	2	83		11	11		22
	5	0	6	12	14	32			6	6	12
		11	44	46	16	117	1	11	17	6	35
Total		34	60	56	16	166	8	14	22	6	50
			150					44			

DCE, dynamic contrast-enhanced; I, irregular without rim-enhancement; R, irregular with rim-enhancement; S, smooth.

Finally, a point having the maximum specificity in the area of NPV $\geq$ 98% was obtained from the NPV map and specificity map to determine the two thresholds of the feature pair. This process was repeated for all of the 44253 feature pairs; thus, the NPV and specificity of all feature pairs and thresholds were determined.

The training set was divided into two groups. The feature pairs were selected in two steps. In the first step, the top 1000 feature pairs were selected on the first division of the training set. Feature pairs with larger specificity were selected in restriction of NPV $\geq$ 98%. The thresholds of the feature pairs were determined in this step. In the second step, the NPV and specificity were calculated on the second division of the training set with the determined thresholds from the first step. Histogram estimation was not used in the second step. Feature pairs with a larger specificity were selected in the restriction of NPV $\geq$ 98%. Feature pair selection and threshold optimization consisted of these two steps and were repeated for each mass shape type.

Evaluation on testing set

Selected feature pairs and thresholds were applied to the masses in the testing set for each shape type, for the classification of either negative or positive. By referring to histopathologic benign and malignant results, the numbers of masses of true negatives, false positives, false negatives, and true positives were counted to produce contingency tables of corresponding shape type of mass. NPV, specificity, PPV, and sensitivity were calculated from contingency tables, and they were averaged over feature pairs selected for the shape type. From these averages, the NPVs for the three shape types and average NPV of total masses were calculated considering the number masses of S-, I-, and R-type masses. Similarly, average specificity, average PPV, and average sensitivity of total masses were calculated.

Results

Image review

MRI category and shape type of masses assessed by the reviewer are shown in Table 3. No masses assessed as category 1 or 2 were included. After exclusion of spiculated masses, the total numbers of S-type, I-type, and R-type masses were 150 (49 benign, 101 malignant) in the training set and 44 (15 benign, 29 malignant) in the testing set. The NPV and specificity in the MRI category were 85.7% and 40%, respectively (Table 4).

Image processing

During the image review and identification of masses, significant motion was identified in 20 masses in the training set and in 12 masses in the testing set. Rigid motion correction was applied for each of them. For each of the 150 masses in the training set and 44 masses in the testing set, 298 texture features were calculated.

Optimization of discriminator on training set

Sets of texture feature pairs were selected for S-type, I-type, and R-type in the optimization process for the training set. Selected feature pairs and corresponding thresholds are shown in Tables 5, 6, and 7. The numbers of selected feature pairs are 27, 12, and 15 for S-type, I-type, and R-type, respectively.

Table 4 NPV, Specificity, PPV, and Sensitivity of MRI category in the training set and testing set

	Training set		Testing set	
	Including Spicula	Excluding Spicula	Including Spicula	Excluding Spicula
NPV (%)	85.7	85.7	85.7	85.7
Specificity (%)	24.5	24.5	40.0	40.0
PPV (%)	75.7	72.8	79.1	75.7
Sensitivity (%)	98.3	98.0	97.1	96.6
Malignant/total lesions (%)	70.5	67.3	70.0	65.9

MRI category 3 and lower categories are regarded as negative, while categories 4 and 5 are regarded as positive. NPV, negative predictive value; PPV, positive predictive value.

Evaluation on testing set

NPV, specificity, PPV, and sensitivity calculated for selected feature pairs are shown in four columns on the right side of Tables 5, 6, and 7. In S-type feature pairs (Table 5), NPV and sensitivity for all 27 pairs were 100%. Specificity ranged from 43% to 71%. Note that the reliabilities of NPV and specificity were restricted by the number of masses. In S-type masses, the total number of masses was 8, and 7 of those masses were benign. In I-type feature pairs (Table 6), the NPV was 100% and specificity was 33% for 10 pairs in 12 pairs. The remaining 2 pairs could not identify any benign masses from a total of 3 I-type benign masses. In R-type feature pairs (Table 7), the NPV was 100% for 14 pairs in 15 pairs. Specificity ranged from 20% to 60%. The number of R-type masses was 22, and 5 of those masses were benign.

The average NPV, specificity, PPV, and sensitivity in the testing set of S-type, I-type, R-type, and total masses are shown in Table 8. The average NPVs for S-type, I-type, and R-type were 100% or a value close to 100%. Specificity in S-type masses was relatively larger than the specificity for I-type and R-types because of the larger fraction of benign masses. Average NPV and average specificity of the total masses were 99.0% and 45.2%, respectively. These values were comparable with the diagnostic values for the MRI category (NPV 85.7%, specificity 40.0%) in Table 4. Average PPV and average sensitivity were 77.9% and 99.8%, respectively.

Feature distributions in testing set

Feature values of benign and malignant masses in testing set are shown in scatter plots (Fig. 2). Horizontal and vertical axes correspond with values of feature pairs selected on the training set for S-type, I-type and R-type masses. Benign masses and malignant masses are plotted with crosses and circle, respectively. Dotted lines in the graphs indicate the cut-off values, and masses in gray area are classified as negative (benign).

Figure 2a shows two scatter plots with two examples of feature pairs for S-type masses. DCE images of masses indicted by a1 and a2 in scatter plots are shown in Fig. 3a1 and 3a2, respectively. Feature values of a1 and a2 are listed in Table 9a1 and 9a2. Two masses are proven as benign by biopsy and classified as negative with two feature pairs. The feature shannon_diff PPE_SER represents the subtraction of Shannon entropy of PPE from Shannon entropy of SER. With this feature, masses with homogeneous SER and inhomogeneous PPE were classified as negative. This property confirms the inhomogeneous enhancement and uniform enhancement change after initial phase as seen in DCE images of Fig. 3a1 and 3a2. Refer schematic diagrams in Fig. 4, for example, time-intensity curve patterns relating with inhomogeneous enhancement and uniform enhancement change after initial phase. The feature shannon_index_2 IE_PIE represents joint entropy of IE and PIE. With this feature, masses with a lower correlation in IE and PIE are classified as negative. This property confirms the inhomogeneous enhancement and uniform enhancement change after the initial phase as seen in Fig. 3a1 and 3a2. The feature skewness IE has a higher value if regions with very large IEs are broader than the regions with very small IEs. Masses with a lower skewness IE value were classified as negative. This property confirms that the fewer regions in Fig. 3a1 and 3a2 have very large enhancement. Figure 4 also illustrates this type of time-intensity curves that few curves plotted in upper area of graph indicated with light gray.

The I-type masses indicated by b in Fig. 2b are shown in Fig. 3b in DCE images. Feature values of the mass b are listed in Table 9b. This mass was benign and classified as negative. With the feature C95 PT, I-type masses were classified as negative if the 95th percentile of peak time was nearby at the time of last phase, or with the feature shannon index PIE, I-type masses were classified as negative if the initial enhancement was uniform. DCE image in Fig. 3b

Table 5 NPV, specificity, PPV, and sensitivity of individual texture future pairs of S-type masses in the testing set

Feature 1	Feature 2	Sign 1	Threshold 1	Sign 2	Threshold 2	Operation	NPV (%)	Specificity (%)	PPV (%)	Sensitivity (%)
skewness IE	shannon_diff PPE_SER	>	1.8132	>	-0.7151	OR	100	57.14	25	100
skewness IE	shannon_rel_diff PPE_SER	>	1.8132	>	-0.2196	OR	100	71.43	33.33	100
skewness MEIP	shannon_diff PPE_SER	>	1.7875	>	-0.7151	OR	100	57.14	25	100
skewness MEIP	shannon_rel_diff PPE_SER	>	1.7875	>	-0.2196	OR	100	71.43	33.33	100
skewness PPE	shannon_diff PPE_SER	>	1.9054	>	-0.7151	OR	100	57.14	25	100
skewness PPE	shannon_rel_diff PPE_SER	>	1.9054	>	-0.2196	OR	100	71.43	33.33	100
shannon_diff PPE_SER	kurtosis IE	>	-0.7376	>	7.3338	OR	100	57.14	25	100
shannon_diff PPE_SER	kurtosis PPE	>	-0.7151	>	7.5401	OR	100	57.14	25	100
shannon_rel_diff PPE_SER	kurtosis IE	>	-0.2281	>	7.3338	OR	100	71.43	33.33	100
shannon_rel_diff PPE_SER	kurtosis PPE	>	-0.2196	>	7.5401	OR	100	71.43	33.33	100
shannon_rel_diff PPE_SER	shannon_index_2 MS_PIE	>	-0.2196	<	5.4489	OR	100	57.14	25	100
shannon_index_2 IE_PIE	shannon_diff PPE_SER	<	5.5455	>	-0.6701	OR	100	57.14	25	100
shannon_index_2 IE_SEIP	shannon_diff PPE_SER	<	5.5629	>	-0.6701	OR	100	57.14	25	100
shannon_index_2 MEIP_PIE	shannon_diff PPE_SER	<	5.5578	>	-0.6701	OR	100	57.14	25	100
shannon_index_2 MEIP_SEIP	shannon_diff PPE_SER	<	5.5612	>	-0.6701	OR	100	57.14	25	100
shannon_index_2 PPE_PIE	shannon_diff PPE_SER	<	5.5594	>	-0.6701	OR	100	57.14	25	100
shannon_index_2 PPE_SEIP	shannon_diff PPE_SER	<	5.588	>	-0.6701	OR	100	57.14	25	100
shannon_index_2 MS_PIE	shannon_diff PPE_SER	<	5.4489	>	-0.7151	OR	100	57.14	25	100
shannon_index_2 MS_SEIP	shannon_diff PPE_SER	<	5.4266	>	-0.7151	OR	100	57.14	25	100
simpson_index_2 IE_PIE	shannon_diff PPE_SER	<	0.9951	>	-0.6926	OR	100	57.14	25	100
simpson_index_2 IE_SEIP	shannon_diff PPE_SER	<	0.9952	>	-0.6926	OR	100	42.86	20	100
simpson_index_2 MEIP_PIE	shannon_diff PPE_SER	<	0.9951	>	-0.6701	OR	100	57.14	25	100
simpson_index_2 MEIP_SEIP	shannon_diff PPE_SER	<	0.9952	>	-0.6926	OR	100	42.86	20	100
simpson_index_2 PPE_PIE	shannon_diff PPE_SER	<	0.9953	>	-0.6701	OR	100	57.14	25	100
simpson_index_2 PPE_SEIP	shannon_diff PPE_SER	<	0.9953	>	-0.6701	OR	100	57.14	25	100
simpson_index_2 MS_PIE	shannon_diff PPE_SER	<	0.9946	>	-0.6926	OR	100	42.86	20	100
simpson_index_2 MS_SEIP	shannon_diff PPE_SER	<	0.9945	>	-0.6926	OR	100	57.14	25	100

IE, initial enhancement; MEIP, maximum enhancement in initial phase; MS, maximum slope; NPV, negative predictive value; PIE, post initial enhancement; PPE, peak percentage enhancement; PPV, positive predictive value; PT, peak time; S, smooth; SEIP, subsequent enhancement to initial peak; SER, signal enhancement ratio.

Table 6 NPV, specificity, PPV, and sensitivity of individual texture future pairs of I-type masses in the testing set

Feature 1	Feature 2	Sign 1	Threshold 1	Sign 2	Threshold 2	Operation	NPV (%)	Specificity (%)	PPV (%)	Sensitivity (%)
very_fast-persistent	Max PIE	<	11.2607	>	31.171	AND	100	0	78.57	100
very_fast-persistent	simpson_index_2 IE_PIE	<	11.4937	>	0.9924	AND	100	0	78.57	100
standard deviation PIE	C95 PT	>	6.4225	<	324.2233	AND	100	33.33	84.62	100
standard deviation PIE	Max PT	>	6.2248	<	320.5147	AND	100	33.33	84.62	100
shannon_index PIE	C95 PT	>	2.5621	<	324.2233	AND	100	33.33	84.62	100
shannon_index PIE	Max PT	>	2.5621	<	321.5429	AND	100	33.33	84.62	100
shannon_multiply IE_SER	C95 PT	>	6.8708	<	323.6521	AND	100	33.33	84.62	100
shannon_multiply IE_SER	Max PT	>	6.7838	<	320.5147	AND	100	33.33	84.62	100
Max PIE	C95 PT	>	31.171	<	323.0809	AND	100	33.33	84.62	100
Max PT	Max PIE	<	320.0006	>	31.171	AND	100	33.33	84.62	100
Max PT	C95-C5 PIE	<	320.5147	>	20.0022	AND	100	33.33	84.62	100
C95-C5 PIE	C95 PT	>	20.6713	<	324.2233	AND	100	33.33	84.62	100

I, irregular without Rim-enhancement; IE, initial enhancement; NPV, negative predictive value; PIE, post initial enhancement; PPV, positive predictive value; PT, peak time; SER, signal enhancement ratio.

shows the persistent pattern in an entire region of the mass. This enhancement pattern makes C95 PT larger leading to a negative decision. Persistent enhancement pattern and uniform enhancement confirm with clinical observations for benign masses.

DCE images of R-type benign mass indicated by c in Fig. 2c are shown in Fig. 3c. Feature values of the mass c are listed in Table 9c. On the left side of the graph, R-type masses with large values of a fast-plateau feature and a large value of shannon_diff SER_PIE are classified as negative. The fast-plateau feature denotes the fraction of the pixels with higher IE (IE $\geq$ 100%) and PIE at around zero (-10% $\leq$ PIE < 10%) in the mass region. shannon_diff SER_PIE denotes the difference of Shannon entropies of PIE and SER. This feature represents a positive value if the pre-contrast image is uniform in the region of the mass (Fig. 4). These features led to a negative decision, which was confirmed with observations of the mass region in Fig. 3c with a fast-plateau enhancement pattern and uniform signal intensity in the pre-contrast phase.

On the right-side graph, R-type masses with smaller values of C75-C25 SER and smaller values of skewness PPE are classified as negative. C75-C25 SER represents a difference of 75th percentile and 25th percentile of SER. These features led to a negative decision confirming with observations that the mass region in Fig. 3c had a uniform signal enhancement after the initial phase, and that the mass region had a smaller area with very large enhancement than the area with the very small amount of enhancement. Figure 4 illustrates time-intensity curves corresponding with lower value of C75-C25 SER and lower value of skewness IE.

Discussion

In this study, we obtained high NPV (99.0%) in our testing process in 44 masses in benign and malignant discrimination using texture-based features of DCE MRI. The obtained NPV was higher than that for the MRI categorization (85.7%). Specificity (45.2%) was similar to that of the MRI categorization (40.0%). Features and thresholds used in the testing set were optimized in training set. In the proposed method, the NPV and specificity are consistently used as the criteria in the optimization of the discriminator, where specificity is maximized with restriction of the NPV greater than 98% in the optimization. Such optimization criteria were introduced according to the clinical demand: reduction in falsely positive benign lesions, potentially leading to a reduction in unnecessary biopsies of benign lesions. According to our findings, we have shown that we could obtain a high NPV (99.0%). We considered that a high NPV could be obtained by consistent criteria in optimization process because features associated with a high NPV were selected effectively in the optimization process.

Features selected for characterization of benign and malignant masses were measures of uniformity and persistent enhancement. Evaluation of these features in the discrimination confirmed the clinically known findings of

Table 7 NPV, specificity, PPV, and sensitivity of individual texture future pairs of R-type masses in the testing set

Feature 1	Feature 2	Sign 1	Threshold 1	Sign 2	Threshold 2	Operation	NPV (%)	Specificity (%)	PPV (%)	Sensitivity (%)
skewness PPE	C75-C25 SER	>	2.4667	>	17.3308	OR	100	60	89.47	100
skewness PT	plateau	>	-1.3824	<	29.3175	OR	100	40	85	100
skewness PT	fast-plateau	>	-1.3824	<	24.9494	OR	100	20	80.95	100
kurtosis PPE	C75-C25 SER	>	10.1444	>	18.0996	OR	100	60	89.47	100
shannon_index PIE	C25 PT	>	2.8617	<	124.5146	OR	100	20	80.95	100
shannon_index PIE	C75-C25 SER	>	2.843	>	22.7126	OR	66.67	40	84.21	94.12
shannon_index SEIP	correlation IE_SER	>	2.8714	<	0.0698	OR	100	60	89.47	100
shannon_multiply PPE_PIE	correlation SEIP_PT	>	10.6554	>	0.6815	OR	100	40	85	100
shannon_ratio SER_PIE	fast-plateau	<	1.2677	<	25.8442	OR	100	20	80.95	100
shannon_diff SER_PIE	plateau	<	0.6367	<	30.8883	OR	100	20	80.95	100
shannon_diff SER_PIE	fast-plateau	<	0.6367	<	27.6338	OR	100	20	80.95	100
shannon_rel_diff SER_PIE	fast-plateau	<	0.2364	<	25.8442	OR	100	20	80.95	100
shannon_index_2 PIE_SER	correlation IE_PIE	>	3.4689	<	-0.4061	OR	100	40	85	100
shannon_index_2 PIE_SER	correlation IE_SEIP	>	3.4689	<	-0.3798	OR	100	40	85	100
simpson_index_2 PPE_PT	shannon_index SEIP	>	0.9645	>	2.8339	AND	100	60	89.47	100

IE, initial enhancement; NPV, negative predictive value; PIE, post initial enhancement; PPE, peak percentage enhancement; PPV, positive predictive value; PT, peak time; R, irregular with Rim-enhancement; SEIP, subsequent enhancement to initial peak; SER, signal enhancement ratio.

Table 8 Discrimination performance of proposed method in testing set

Shape type of mass		S-type	I-type	R-type	Total
Number of feature pairs		27	12	15	–
Number of masses	benign	7	3	5	15
	malignant	1	11	17	29
Average NPV (%)		100.0	100.0	97.8	99.0
Average specificity (%)		58.2	27.8	37.3	45.2
Average PPV (%)		26.0	83.6	84.5	77.9
Average sensitivity (%)		100.0	100.0	99.6	99.8

NPV, specificity, PPV, and sensitivity of feature pairs are averaged to compute average NPV, average specificity, average PPV, and average sensitivity for S-type, I-type, and R-type. Combined performances are computed from the number of masses in S-type, I-type, and R-type. I, irregular without Rim-enhancement; NPV, negative predictive value; PPV, positive predictive value; R, irregular with Rim-enhancement; S, smooth.

benign and malignant masses. In S-type masses, masses with inhomogeneous enhancement and with uniform enhancement after the initial phase were classified as negative, and masses that had few regions of very large enhancement were classified as negative too. In I-type masses, masses with a persistent enhancement pattern and with uniform enhancement were classified as negative. In R-type masses, masses with a uniform signal enhancement after the initial phase were classified as negative if the mass region had a smaller area with very large levels of enhancement.

The high NPV aspect of the proposed method could have a contribution to improve the diagnostic performance of breast cancer characterization in DCE MRI by referring to the classification of the result in patients with suspicious lesions on mammography or ultrasound. It has also the potential to reduce the number of unnecessary biopsies performed on benign lesions.

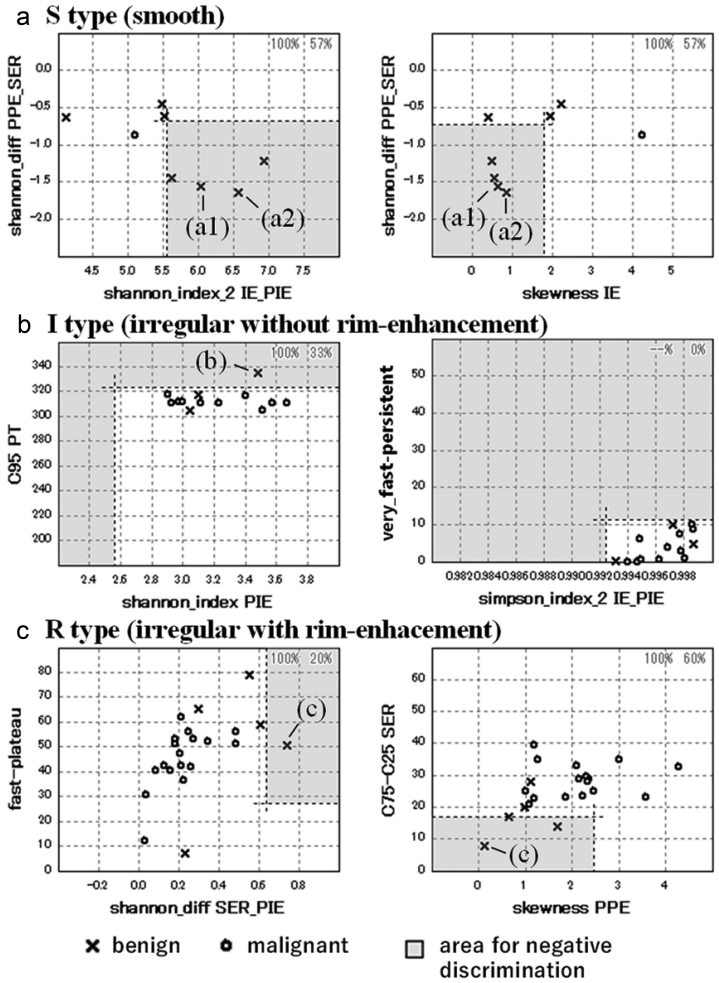

Fig. 2 Six examples of scatter plots of S-type masses (**a**), I-type masses (**b**), and R-type masses in the testing set. Horizontal and vertical axes are corresponding with the values of feature pairs selected on the training set for S-type masses (**a**), for I-type masses (**b**), and R-type masses (**c**). Masses in gray area are classified as negative (benign). DCE MRI of masses indicated by (a1), (a2), (b), and (c) are shown in Fig. 3 (a1), (a2), (b), and (c), respectively. See body texts for individual features. DCE, dynamic contrast-enhanced; I, irregular without Rim-enhancement; IE, initial enhancement; PIE, post initial enhancement; PPE, peak percentage enhancement; PT, peak time; R, irregular with Rim-enhancement; S, smooth; SER, signal enhancement ratio.

Another possible application of the proposed method is MRI surveillance in high-risk groups of breast cancer. Breast MRI could play an important role in the surveillance of breast cancer due to its high sensitivity. Quantitative analysis would be an efficient method to improve the sensitivity and the NPV by reducing the interobserver variability in breast MRI analysis. To apply the proposed method in MRI surveillance, validation of the diagnostic performance of the proposed method would be needed in a larger number of subjects within a high breast cancer risk. The larger study should also include smaller lesions and non-mass lesions.[9,10] The selection of features and threshold designed for high-risk groups is an important challenge for our future study.

This study was subject to several limitations. First, the number of case samples was limited. We could not statistically prove the similar or superior NPV of the proposal method in comparison with the NPV in MRI categorization. The confidence interval of NPV in the proposed method was estimated at (53% and 100%) in the Clopper-Pearson's exact confidence interval, estimating NPV 99% and specificity 50% in a total of 44 masses with 70% of prevalence. The NPV of the proposed method was not demonstrated to be similar or superior to the NPV in MRI categorization (85.7%). Further validation would be needed with larger case samples in a multi-center study. The limitation of case samples also affects the selection of feature pairs.

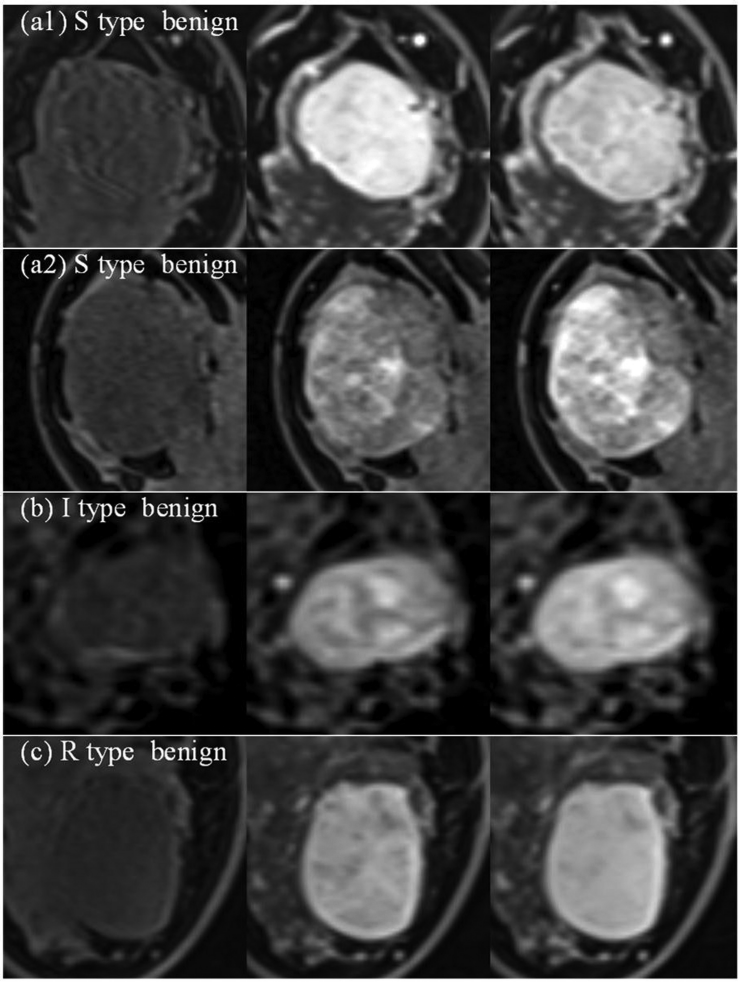

Fig. 3 DCE MRI of masses corresponding with benign masses indicated by (a1, a2, b, and c) in Fig. 2 are shown in (**a1**), (**a2**), (**b**), and (**c**), respectively. Figures at left, center, and right corresponding with pre-contrast phase, initial phase around 2 min after injection, and last phase around 5 min after injection, respectively. I, irregular without Rim-enhancement; R, irregular with Rim-enhancement; S, smooth.

Feature pairs were selected from a large number of combinations with a limited number of cases in training set, and they were validated with a limited number of cases in the testing set. Feature pairs also needs further validation with larger case samples. Second is the bias in lesions characteristics. Case samples in this study consisted of a larger number of lesions with an MRI category 4, and a smaller number of MRI category 3. No lesions with MRI categories 1 and 2 were included. The case selection of this study centered on cancer-suspected lesions in either mammography or ultrasound, and histologically proven benign and malignant lesions were enrolled. Most of the typical benign cases were excluded in this case selection. Further validation would be desirable in a study design with a larger variation of benign lesions more in line with the lesion characteristics seen in routine examinations. The third limitation is the dependency on the MRI systems. Kinetic curves measured in DCI MRI can have inter-system variability.[11] In this study, two MRI systems and two contrast medium injection protocols were used in the acquisition of the training set to reduce the inter-system variability; however, we used a single MRI system (Biograph mMR) with a field strength of 3.0T for the testing set. Further validation is required to establish any validity in multiple MRI systems. Especially, MRI systems with different field strengths and cases with different types of contrast medium should be included in the validation. Finally, the mass shape types of lesions were visually classified by one MRI reviewer. Discrimination of malignancy in the proposed method depends on mass shape type, so any variation of the mass shape type will cause variation in the discrimination performance. Reducing the possibility of interobserver variability is also key to the discrimination process by characterizing the mass shape analysis with morphology features.

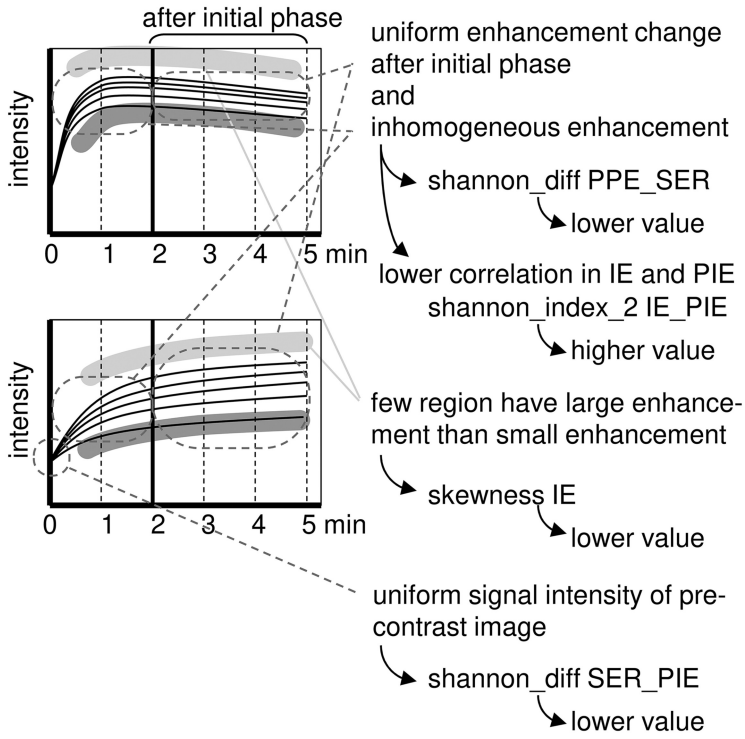

Fig. 4 Two graphs for illustrating variation of time-intensity curves at different points distributed in a mass and their relationship with features. Upper graph is a schematic diagram of time-intensity curves of mass with inhomogeneous enhancement and uniform enhancement change after initial phase. shannon_diff PPE_SER is lower value and shannon_index_2 IE_PIE is higher value in this type of time-intensity curves. Lower graph is another example of same time-intensity curves of mass with inhomogeneous enhancement and uniform enhancement change after initial phase. If the number of pixels corresponding with the curves in light gray area is smaller than the curves in dark gray area, skewness IE is smaller value. Uniform signal intensity of pre-contrast image makes shannon_diff SER_PIE smaller. These properties lead to negative decision. IE, initial enhancement; PIE, post initial enhancement; PPE, peak percentage enhancement; SER, signal enhancement ratio.

Table 9 Feature values of masses in Figures 2 and 3

Feature	Value	Range of positive decision	
(a1) S-type benign			
shannon_index_2 IE_PIE	6.03	< 5.55	OR
shannon_diff PPE_SER	-1.57	> -0.67	
skewness IE	0.64	> 1.81	OR
shannon_diff PPE_SER	-1.57	> -0.72	
(a2) S type benign			
shannon_index_2 IE_PIE	6.56	< 5.55	OR
shannon_diff PPE_SER	-1.65	> -0.67	
skewness IE	0.86	> 1.81	OR
shannon_diff PPE_SER	-1.65	> -0.72	
(b) I type benign			
shannon_index PIE	3.48	> 2.56	AND
C95 PT	335.64	< 324.22	
simpson_index_2 IE_PIE	0.999	> 0.992	OR
very_fast-persistent	4.90	< 11.49	
(c) R type benign			
shannon_diff SER_PIE	0.74	< 0.64	OR
fast-plateau	50.67	< 27.63	
skewness PPE	0.124	> 2.467	OR
C75-C25 SER	8.16	> 17.33	

IE, initial enhancement; PIE, post initial enhancement; PPE, peak percentage enhancement; SER, signal enhancement ratio.

Conclusion

In this study, a benign and malignant discrimination method for texture features of breast DCE MRI was presented in which the specificity is maximized with the restriction of NPV greater than 98% in optimization process of discriminator in training. According to the criteria used in the optimization process, a high NPV and a reasonable specificity were obtained in the testing process. We, therefore, suggest that the NPV of the present method has the potential to be similar to or greater than the NPV in MRI category-based diagnosis by the expert radiologist.

Conflicts of Interest

Shigeharu Ohyu, Hisae Chiba, and Yasuko Fujisawa are employee of Canon Medical Systems Corporation. Qilin Xiao is employee of Canon Medical Systems (China) Co., Ltd. Mitsuhiro Tozaki, Michiro Sasaki, and Yoshiaki Sagara have no conflicts of interest.

References

1. Peters NH, Borel Rinkes IH, Zuithoff NP, et al. Meta-analysis of MR imaging in the diagnosis of breast lesions. Radiology 2008; 246:116–124.
2. Zhang L, Tang M, Min Z, et al. Accuracy of combined dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted imaging for breast cancer detection: a meta-analysis. Acta Radiol 2016; 57:651–660.
3. Karahaliou A, Vassiou K, Arikidis NS, et al. Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis. Br J Radiol 2010; 83:296–309.
4. Gibbs P, Onishi N, Sadinski M, et al. Characterization of Sub-1 cm breast lesions using radiomics analysis. J Magn Reson Imaging 2019; 50:1468–1477.
5. Bhooshan N, Giger ML, Jansen SA, et al. Cancerous breast lesions on dynamic contrast-enhanced MR images: computerized characterization for image-based prognostic markers. Radiology 2010; 254:680–690.
6. American College of Radiology. ACR BI-RADS atlas. Breast imaging reporting and data system: Mammography, ultrasound, magnetic resonance imaging, follow-up and outcome monitoring, data dictionary, 5th ed. Reston:American College of Radiology, 2013.
7. Tozaki M, Fukuma E. 1H MR spectroscopy and diffusion-weighted imaging of the breast: are they useful tools for characterizing breast lesions before biopsy? AJR Am J Roentgenol 2009; 193:840–849.
8. Grimm LJ, Anderson AL, Baker JA, et al. Interobserver variability between breast imagers using the fifth edition of the BI-RADS MRI Lexicon. AJR Am J Roentgenol 2015; 204:1120–1124.
9. Tozaki M, Nakamura S, Kitagawa D, et al. Ductal carcinoma *in situ* detected during prospective MR imaging screening of a woman with a BRCA2 mutation: the first case report in Japan. Magn Reson Med Sci 2017; 16:265–269.
10. Tozaki M, Nakamura S. Current status of breast cancer screening in high-risk women in Japan. Breast Cancer 2021; 28:1181–1187.
11. Jansen SA, Shimauchi A, Zak L, et al. Kinetic curves of malignant lesions are not consistent across MRI systems: need for improved standardization of breast dynamic contrast-enhanced MRI acquisition. AJR Am J Roentgenol 2009; 193:832–839.