

Appendix 1

MRI scan parameters

Breast MRI examinations were performed on 1.5 T (Avanto, Siemens, Germany) and 3.0 T MRI scanners (Verio, Siemens, Germany; Discovery MR750, GE Healthcare, USA). The details of MRI scan parameters are listed in Table S1.

Selection of mpMRI-based decision rules

We searched previous published studies from PubMed and referred to the study (17) that summarized decision rules for characterization of breast lesions. Among 16 mpMRI-based decision rules, only 3 were involved in our study. The selection of mpMRI-based decision rules is listed in Table S2.

Details of all decision rules involved in the study

Three mpMRI-based decision rules by Kim *et al.* (19), Istomin *et al.* (17) and Zhong *et al.* (18) and Kaiser (20) score were analysed in our study. Additionally, BI-RADS category, which was assigned based on radiologists' experience, was retrieved from radiology reports and was also analysed. The details of these decision rules for characterizing breast lesions are illustrated below.

- ❖ Kim *et al.* (19): it was developed based on imaging features on DCE MRI and signal intensity on T2WI-DWI. The total score for a breast lesion was the sum of DCE score and DWI-T2WI set score and it ranged from 3 to 10 points. A lesion with total score >5 points was considered as malignant, otherwise, it was benign. The details of DCE score and DWI-T2WI set score were illustrated as follows: (I) DCE score: it was equal to DCE BI-RADS category and was given based on morphology, enhancement pattern and kinetics. A focal mass with round, oval shape; smooth margin; non-enhancing internal septations; and a persistent TIC was categorized as BI-RADS 2 category. Focal masses with round, oval shape; smooth margin; persistent TIC; and asymmetric non-mass enhancement of focal or regional distribution were categorized as BI-RADS 3 category. A mass with irregular shape; irregular or spiculated margin; heterogeneous or rim enhancement; non-mass lesions with clumped and clustered ring enhancement at linear or segmental distribution; and plateau or wash-out time TIC was categorized as BI-RADS 4 category. A mass or non-mass lesion with suspicious morphology; more than 90% wash-in rate; and wash-out TIC was categorized as BI-RADS 5 category. (II) DWI-T2WI set score: it ranged from 1 (definite benign) to 5 (highly suggestive of malignancy) on the basis of signal intensity of T2WI and DWI (high b value imaging) (Table S3).
- ❖ Istomin *et al.* (17): MRI features including morphology, enhancement pattern, kinetics, signal intensity on T2WI and ADC values were categorized into the minor, intermediate and major for malignancy (Table S4). A lesion with minor features was set as BI-RADS 3 category, a lesion with intermediate and no more than one major feature was set as BI-RADS 4 category, otherwise, it was set as BI-RADS 5 category.
- ❖ Zhong *et al.* (18): it was developed based on morphology, TIC type and ADC values on MRI. Each suspicious finding was assigned a score of 1 point. The scoring ranged from 0 points to 3 points. A lesion with 0 point and 1 point was set as BI-RADS 3 and 4 category, respectively. A lesion with 2 and 3 points was set as BI-RADS 5 category. The details of suspicious findings are shown in Table S5.
- ❖ Kaiser score (20): it was developed using Chi-squared automatic interaction detection method and demonstrated as decision tree structure. Five imaging features that were obtained from T2WI and DCE-MRI were incorporated in Kaiser score, namely spiculated sign, TIC type, margin, enhancement pattern and edema. Kaiser score ranged from 1 to 11 points, with greater score suggestive a high likelihood of malignancy. Lesions with score >4 points were considered as malignant. Of note, Kaiser score could be used for NME, although it showed discordant with BI-RADS descriptors for NME. The detail of Kaiser score referred to the review by Dietzel *et al.* (61).
- ❖ BI-RADS category: it was retrieved from radiology reports. No decision rules was given and assessment was performed based on radiologists' experience. In our department, breast MRI interpretation was double-reading by two radiologists (one junior and one senior) specialized in breast imaging. The assessment was established on multiple

sequences, including DCE-MRI, T2WI and DWI. During imaging interpretation, radiologists have access to measure ADC values on ADC maps on our workstation.

Descriptive statistics for all decision rules stratified by individual scores

Table S6 shows descriptive statistics for all decision rules stratified by individual scores. According to the decision rule, Kaiser score categorizes lesions with a score of 1–4 points as BI-RADS 2–3, 5–7 points as BI-RADS 4, and 8–11 points as BI-RADS 5. In the decision rule by Zhong *et al.*, a lesion with a score of 0 point is equivalent to a BI-RADS 3 category, a score of 1 point is equivalent to a BI-RADS 4 category, and a score of 2–3 points is equivalent to a BI-RADS 5 category. The corresponding BI-RADS category based on Kaiser score and the decision rule by Zhong *et al.* is shown in Table S7.

Details of 9 benign lesions that cannot be scored by DWI-T2WI set score in the decision rule by Kim *et al.*

In our study, 9 benign lesions cannot be scored by DWI-T2WI set score because of their signal intensity on DWI and T2WI does not align with the predefined DWI-T2WI set in Kim *et al.*, of which 77.8% (7/9) is fibroadenoma. The details are shown in Table S8. Examples are given in Figures S1,S2. Additionally, the correlation of DWI-T2WI set score with pathology results of the lesions are summarized in Table S9. Of 110 benign lesions, 73 lesions (66.4%) have false-positive findings on DWI-T2WI set. The sensitivity and specificity of DWI-T2WI alone for characterizing breast lesions is 96.0% (95/99) and 33.6% (37/110).

Comparison of diagnostic performance of the three mpMRI-based decision rules in original studies and ours

Sensitivity, specificity and AUC of the three mpMRI-based decision rules in original studies and ours are listed in Table S10. Malignancy rates in original studies were higher than ours (52.6–76.2% *vs.* 45.4%). The decision rules by Kim *et al.* and Zhong *et al.* showed a markedly lower specificity in our study compared to original studies (Kim *et al.*: 31.8% *vs.* 91.7%; Zhong *et al.*: 31.9% *vs.* 69.8%).

Table S1 Details of MRI scan parameters

Parameters	GE Discovery MR750	Siemens Avanto	Siemens Verio
Magnetic field	3.0 T	1.5 T	3.0 T
Breast coil	8-channel	16-channel	16-channel
T2WI-fat saturation			
Fat saturation	IDEAL	TIRM	TIRM
Slice thickness/slice gap	4 mm/5 mm	4 mm/4.8 mm	4 mm/4.4 mm
TR	5,560	5,600	4,000
TE	86	56	70
DWI			
Sequence	Single-shot EPI	Single-shot EPI	Single-shot EPI
Fat saturation	IR + water excitation	SPAIR	SPAIR
Slice thickness/slice gap	4 mm/5 mm	4 mm/6 mm	4 mm/6 mm
TR	3,003 ms	6,500 ms	8,400 ms
TE	57.3 ms	108 ms	84 ms
b value (s/mm ²)	0, 1,000	50, 400, 800	50, 400, 800
T1WI with non-fat saturation			
	N/A	Optional	Optional
Slice thickness/slice gap		1.0 mm/0 mm	1.0 mm/0 mm
TR		13	5.9
TE		4.7	2.2
Dynamic contrast-enhanced sequences			
FOV	340 mm	340 mm	340 mm
Slice thickness/slice gap	1.4 mm/0 mm	1.2 mm/0 mm	1.0 mm/0 mm
TR	4.48 ms	4.43 ms	4.51 ms
TE	1.79 ms	1.5 ms	1.61 ms
Temporal resolution	60 s	60 s	60 s
In-plane spatial resolution	1.3 mm × 0.8 mm	1.0 mm × 0.8 mm	1.1 mm × 0.8 mm
Phase	1 precontrast phase followed by 6 postcontrast phases	1 precontrast phase followed by 5 postcontrast phases	1 precontrast phase followed by 5 postcontrast phases

MRI, magnetic resonance imaging; TR, time of repetition; TE, time of echo; T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging; T1WI, T1 weighted imaging; IDEAL, iterative decomposition of water and fat with echo asymmetry and least-squares estimation; EPI, echo-plane imaging; TIRM, turbo inversion recovery magnitude; IR, inversion recovery; SPAIR, spectral attenuated inversion recovery; FOV, field of view; N/A, not applicable.

Table S2 Selection of mpMRI-based decision rules

Authors	Published year	Study purpose	Inclusion	Exclusion reason
Pinker <i>et al.</i> (13)	2013	Differentiation of breast benign lesions from malignancy based on mpMRI	No	The decision rule was too complicated to use as 10 ADC thresholds were used
Sarica <i>et al.</i> (52)	2014	Characterizing breast lesions combining MRI descriptors and additional criteria	No	The decision rule could not apply to NME
Cai <i>et al.</i> (53)	2014	Differentiation of breast benign lesions from malignancy using DWI and DCE-MRI by machine learning	No	Texture features involved in the decision rule could not be obtained in daily work
Maltez de Almeida <i>et al.</i> (54)	2015	Subcategorization of BI-RADS 4 lesions using DCE-MRI and DWI	No	The decision rule only applied to BI-RADS 4 lesions and it was displayed by logistic regression formula
Baltzer <i>et al.</i> (14)	2016	Differentiation of breast benign lesions from malignancy using two combined reading for DCE-MRI and DWI	No	No detailed diagnostic criteria were provided for DCE-MRI BI-RADS category
Kawai <i>et al.</i> (55)	2018	Adding lesions size to BI-RADS descriptors for differentiation of breast solitary mass	No	The decision rule could not apply to NME and multiple breast masses
Fujiwara <i>et al.</i> (15)	2018	Grading breast mass using MRI descriptors and subcategorizing BI-RADS 4 lesions	No	The decision rule could not apply to NME and DWI was not involved
Kim <i>et al.</i> (19)	2018	Characterizing breast lesions by adding DWI and T2WI to DCE-MRI	Yes	–
Liu <i>et al.</i> (56)	2018	Evaluation of Fischer' score and ADC values for characterizing breast lesions	No	The Fischer' score involved in the decision rule was outdated as it was proposed in 1999
Asada <i>et al.</i> (57)	2018	Categorizing NME using MRI descriptors	No	The decision rule could not apply to mass; DWI was not involved
Zhang <i>et al.</i> (16)	2019	Differentiation of breast benign lesions from malignancy using DCE MRI and DWI	No	The decision rule could not apply to NME and it was not intuitive as it was displayed by logistic regression formula
Istomin <i>et al.</i> (17)	2020	Categorizing breast lesions using mpMRI	Yes	–
Ellmann <i>et al.</i> (58)	2020	Characterizing breast suspicious lesions by machine learning	No	The decision rule could only apply to breast suspicious mass
Sun <i>et al.</i> (59)	2021	Characterizing BI-RADS 4 lesions using mpMRI	No	The decision rule involved synthetic MRI parameters and could only apply to BI-RADS 4 lesions
Vassiou <i>et al.</i> (60)	2022	Categorizing breast lesions using DCE-MRI and DWI	No	The decision rule was displayed by logistic regression formula and it was not intuitive to use
Zhong <i>et al.</i> (18)	2022	Categorizing breast lesions using DCE-MRI and DWI	Yes	–

ADC, apparent diffusion coefficient; NME, non-mass enhancement; DWI, diffusion-weighted imaging; BI-RADS, breast imaging reporting and data system; T2WI, T2-weighted imaging; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; mpMRI, multiparametric magnetic resonance imaging; MRI, magnetic resonance imaging.

Table S3 DWI-T2WI set score in the decision rule by Kim *et al.*

DWI-T2WI set score	T2WI signal intensity	DWI signal intensity	Indication
1	High	Low	Definite benign
2	High	Intermediate	Probably benign
3	High	High	Possibly malignant
4	Intermediate	High	Probably malignant
5	Low	Very high	Definite malignant

DWI, diffusion-weighted imaging; T2WI, T2-weighted imaging.

Table S4 MRI features in the decision rule by Istomin *et al.*

MRI findings	Minor	Intermediate	Major
Morphology	Mass: oval shape, circumscribed margin NME: focal, multiple regions, diffuse	Mass: round or irregular shape, irregular margin NME: linear, regional	Mass: spiculated NME: segmental
Enhancement pattern	Mass: dark internal septations, homogeneous NME: homogeneous	Mass: heterogeneous NME: heterogeneous	Mass: rim enhancement NME: clumped, clustered ring
Kinetics	Persistent	Plateau	Washout
Signal intensity on T2WI	High	Intermediate or low	N/A
ADC values ($\times 10^{-3}$ mm ² /s)	N/A	N/A	≤ 0.69

MRI, magnetic resonance imaging; NME, non-mass enhancement; T2WI, T2-weighted imaging; ADC, apparent diffusion coefficient; N/A, not applicable.

Table S5 Details of suspicious findings in the decision rule by Zhong *et al.*

Suspicious findings	Mass	NME	Score
Morphology	Not-circumscribed margin, irregular shape, rim or heterogeneous enhancement	Linear or segmental distribution, clumped or cluster ring	1
TIC	Washout	Plateau or washout	1
ADC values ($\times 10^{-3}$ mm ² /s)	≤ 1.05	≤ 1.35	1

NME, non-mass enhancement; TIC, time signal intensity curve; ADC, apparent diffusion coefficient.

Table S6 Descriptive statistics for all decision rules stratified by individual scores

Decision rules	Benign (n=119)	Malignant (n=99)	Total numbers (n=218)	Malignancy rate
BI-RADS category				
Category 3	77	4	81	4.9%
Category 4	41	32	73	43.8%
Category 5	1	63	64	98.4%
Istomin <i>et al.</i>				
Category 3	25	0	25	0%
Category 4	91	43	134	32.1%
Category 5	3	56	59	94.9%
Zhong <i>et al.</i>				
0 point	38	0	38	0%
1 point	45	7	52	13.5%
2 points	25	34	59	57.6%
3 points	11	58	69	84.1%
Kim <i>et al.</i> [†]				
3 points	1	0	1	0%
4 points	8	0	8	0%
5 points	25	0	25	0%
6 points	18	4	22	18.2%
7 points	37	23	60	38.3%
8 points	18	42	60	70.0%
9 points	3	21	24	87.5%
10 points	0	9	9	100%
Kaiser score				
1 point	72	0	72	0%
2 points	13	5	18	27.8%
3 points	13	3	16	18.8%
4 points	6	0	6	0%
5 points	6	21	27	77.8%
6 points	2	2	4	50%
7 points	1	8	9	88.9%
8 points	4	21	25	84.0%
9 points	2	13	15	86.6%
10 points	0	11	11	100%
11 points	0	15	15	100%

[†], 9 benign lesions can not be scored by DWI-T2WI set score as their signal intensity on DWI and T2WI does not match the DWI-T2WI set in the decision rule by Kim *et al.* BI-RADS, breast imaging reporting and data system; DWI, diffusion-weighted imaging; T2WI, T2-weighted imaging.

Table S7 Descriptive statistics of the corresponding BI-RADS category based on Kaiser score and decision rule by Zhong *et al.*

Decision rules	Corresponding BI-RADS category	Benign	Malignant	Malignancy rate
Kaiser score				
1–4 points	2–3	104	8	7.1% (8/112)
5–7 points	4	9	31	77.5% (31/40)
8–11 points	5	6	60	90.9% (60/66)
Zhong <i>et al.</i>				
0 point	3	38	0	0% (0/38)
1 point	4	45	7	13.5% (7/52)
2–3 points	5	36	92	71.9% (92/128)

BI-RADS, breast imaging reporting and data system.

Table S8 Details of 9 benign lesions that cannot be scored by DWI-T2WI set score

Lesion numbers	Signal intensity on T2WI	Signal intensity on DWI	Pathology
1	Intermediate	Intermediate	Fibroadenoma
2	Intermediate	Low	Fibroadenoma
3	Intermediate	Low	Fibroadenoma
4	Low	Low	Fibroadenoma
5	Low	Low	Adenosis
6	Low	Intermediate	Fibroadenoma
7	Intermediate	Intermediate	Fibroadenoma
8	Intermediate	Intermediate	Fibroadenoma
9	Intermediate	Intermediate	Usual ductal hyperplasia

DWI, diffusion-weighted imaging; T2WI, T2-weighted imaging.

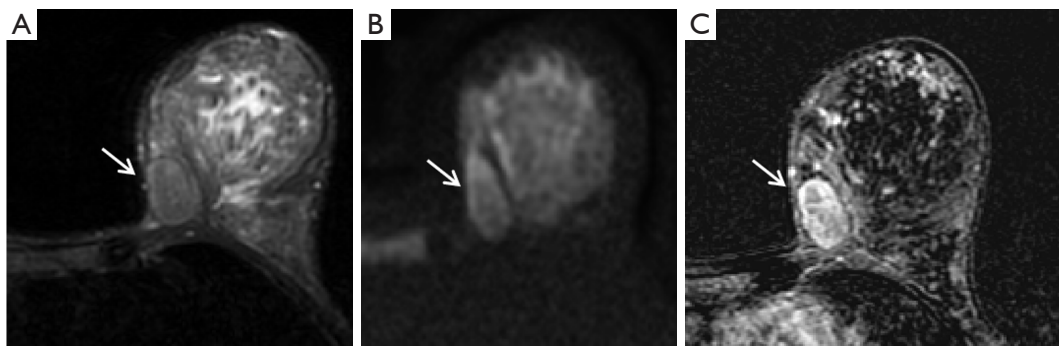


Figure S1 Fibroadenoma in a 40-year-old woman. The mass (white arrows) showed low signal intensity on T2WI with fat-saturation (A), intermediate signal intensity on DWI with b values of $1,000 \text{ s/mm}^2$ (B), in discordance with DWI-T2WI set in the decision rule by Kim *et al.* The lesion displayed oval shape, well-circumscribed margin, heterogeneous internal enhancement and plateau kinetics (C), with ADC values of $1.56 \times 10^{-3} \text{ mm}^2/\text{s}$. It was considered as malignant by the decision rules by Istomin *et al.* and Zhong *et al.*, and benign by Kaiser score and BI-RADS category. T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; BI-RADS, breast imaging reporting and data system.

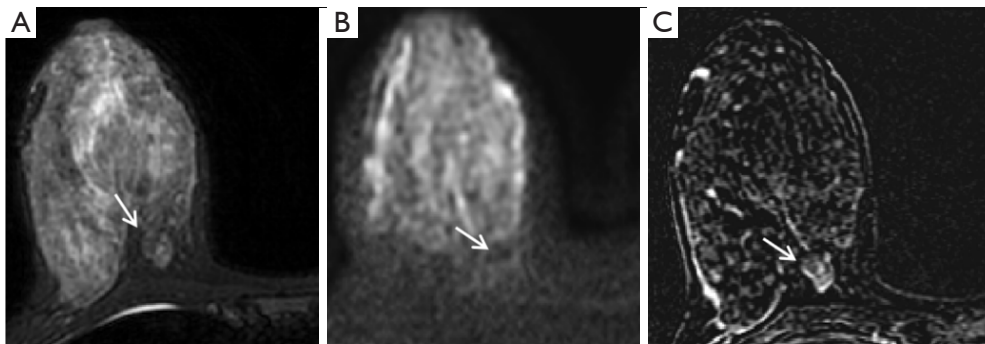


Figure S2 Fibroadenoma in a 41-year-old woman. The mass (white arrows) showed low signal intensity on both T2WI with fat-saturation (A) and DWI with b values of $1,000 \text{ s/mm}^2$ (B), in discordance with DWI-T2WI set in the decision rule by Kim *et al.* The lesion displayed oval shape, well-circumscribed margin, dark internal septation and persistent kinetics (C), with ADC values of $1.03 \times 10^{-3} \text{ mm}^2/\text{s}$. It was considered as benign by the decision rules by Istomin *et al.*, Zhong *et al.*, Kaiser score and BI-RADS category. T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; BI-RADS, breast imaging reporting and data system.

Table S9 Correlation of DWI-T2WI set score in the decision rule by Kim *et al.* with pathology results of the lesions

DWI-T2WI set score	Benign (n=110) [†]	Malignant (n=99)	Total
1 point: definite benign	6	0	6
2 points: probably benign	31	4	35
3 points: possibly malignant	49	28	77
4 points: probably malignant	22	49	71
5 points: definite malignant	2	18	20

[†], 9 benign lesions that cannot be scored by DWI-T2WI set score are not involved. DWI, diffusion-weighted imaging; T2WI, T2-weighted imaging.

Table S10 Diagnostic performance of the three mpMRI-based decision rules in original studies and ours

Decision rules	Study	Total lesions	Malignancy rate (%)	Sensitivity (%)	Specificity (%)	AUC
Istomin <i>et al.</i>	Istomin's	697	76.2 (531/697)	100	11.5	N/A
	Ours	218	45.4 (99/218)	100	21.0	0.81
Kim <i>et al.</i>	Kim's	169	71.6 (121/169)	83.5	91.7	0.922
	Ours	218	45.4 (99/218)	100	31.8	0.84
Zhong <i>et al.</i>	Zhong's	898	52.6 (472/898)	98.9	69.8	0.844
	Ours	218	45.4 (99/218)	100	31.9	0.82

The diagnostic performance is calculated for all type of breast lesions. High-risk lesions in these original studies and ours are grouped into the benign. Lesions with BI-RADS 4 and 5 category are considered malignant, while lesions with BI-RADS 2 and 3 category are benign for calculation of sensitivity and specificity. mpMRI, multiparametric magnetic resonance imaging; AUC, area under the curve; BI-RADS, breast imaging reporting and data system; N/A, not applicable.

References

52. Sarica O, Kaymaz G, Ucar N. Effectiveness of additional diagnostic parameters in magnetic resonance mammography: a comparative study with the BI-RADS classification and scoring system. *J Comput Assist Tomogr* 2014;38:985-91.
53. Cai H, Peng Y, Ou C, Chen M, Li L. Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted MR: a machine learning approach. *PLoS One* 2014;9:e87387.
54. Maltez de Almeida JR, Gomes AB, Barros TP, Fahel PE, de Seixas Rocha M. Subcategorization of Suspicious Breast Lesions (BI-RADS Category 4) According to MRI Criteria: Role of Dynamic Contrast-Enhanced and Diffusion-Weighted Imaging. *AJR Am J Roentgenol* 2015;205:222-31.
55. Kawai M, Kataoka M, Kanao S, Iima M, Onishi N, Ohashi A, Sakaguchi R, Toi M, Togashi K. The Value of Lesion Size as an Adjunct to the BI-RADS-MRI 2013 Descriptors in the Diagnosis of Solitary Breast Masses. *Magn Reson Med Sci* 2018;17:203-10.
56. Liu D, Ba Z, Ni X, Wang L, Yu D, Ma X. Apparent Diffusion Coefficient to Subdivide Breast Imaging Reporting and Data System Magnetic Resonance Imaging (BI-RADS-MRI) Category 4 Lesions. *Med Sci Monit* 2018;24:2180-8.
57. Asada T, Yamada T, Kanemaki Y, Fujiwara K, Okamoto S, Nakajima Y. Grading system to categorize breast MRI using BI-RADS 5th edition: a statistical study of non-mass enhancement descriptors in terms of probability of malignancy. *Jpn J Radiol* 2018;36:200-8.
58. Ellmann S, Wenkel E, Dietzel M, Bielowski C, Vesal S, Maier A, Hammon M, Janka R, Fasching PA, Beckmann MW, Schulz Wendtland R, Uder M, Bäuerle T. Implementation of machine learning into clinical breast MRI: Potential for objective and accurate decision-making in suspicious breast masses. *PLoS One* 2020;15:e0228446.
59. Sun SY, Ding Y, Li Z, Nie L, Liao C, Liu Y, Zhang J, Zhang D. Multiparameter MRI Model With DCE-MRI, DWI, and Synthetic MRI Improves the Diagnostic Performance of BI-RADS 4 Lesions. *Front Oncol* 2021;11:699127.
60. Vassiou K, Fanariotis M, Tsougos I, Fezoulidis I. Incorporating diffusion-weighted imaging in a diagnostic algorithm for multiparametric MR mammography. *Acta Radiol* 2022;63:1332-43.
61. Dietzel M, Baltzer PAT. How to use the Kaiser score as a clinical decision rule for diagnosis in multiparametric breast MRI: a pictorial essay. *Insights Imaging* 2018;9:325-35.