



# An MRI-Based Radiomics Model for Predicting the Benignity and Malignancy of BI-RADS 4 Breast Lesions

Renzhi Zhang<sup>1†</sup>, Wei Wei<sup>2†</sup>, Rang Li<sup>3,4</sup>, Jing Li<sup>1</sup>, Zhuhuang Zhou<sup>4</sup>, Menghang Ma<sup>2</sup>, Rui Zhao<sup>2</sup> and Xinming Zhao<sup>1\*</sup>

<sup>1</sup> Department of Diagnostic Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, <sup>2</sup> School of Electronics and Information, Xi'an Polytechnic University, Xi'an, China, <sup>3</sup> College of Engineering, Boston University, Boston, MA, United States, <sup>4</sup> Faculty of Environment and Life, Beijing University of Technology, Beijing, China

## OPEN ACCESS

### Edited by:

Natalie Julie Serkova,  
University of Colorado, United States

### Reviewed by:

Kun Zheng,  
Peking Union Medical College Hospital  
(CAMS), China  
Stefano Marrone,  
University of Naples Federico II, Italy  
Vishwa S. Parekh,  
Johns Hopkins University,  
United States

### \*Correspondence:

Xinming Zhao  
zhaoxinming@cicams.ac.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Cancer Imaging and  
Image-directed Interventions,  
a section of the journal  
Frontiers in Oncology

Received: 30 June 2021

Accepted: 08 December 2021

Published: 28 January 2022

### Citation:

Zhang R, Wei W, Li R, Li J,  
Zhou Z, Ma M, Zhao R and  
Zhao X (2022) An MRI-Based  
Radiomics Model for Predicting  
the Benignity and Malignancy  
of BI-RADS 4 Breast Lesions.  
*Front. Oncol.* 11:733260.  
doi: 10.3389/fonc.2021.733260

**Objectives:** The probability of Breast Imaging Reporting and Data Systems (BI-RADS) 4 lesions being malignant is 2%–95%, which shows the difficulty to make a diagnosis. Radiomics models based on magnetic resonance imaging (MRI) can replace clinicopathological diagnosis with high performance. In the present study, we developed and tested a radiomics model based on MRI images that can predict the malignancy of BI-RADS 4 breast lesions.

**Methods:** We retrospective enrolled a total of 216 BI-RADS 4 patients MRI and clinical information. We extracted 3,474 radiomics features from dynamic contrast-enhanced (DCE), T<sub>2</sub>-weighted images (T<sub>2</sub>WI), and diffusion-weighted imaging (DWI) MRI images. Least absolute shrinkage and selection operator (LASSO) and logistic regression were used to select features and build radiomics models based on different sequence combinations. We built eight radiomics models which were based on DCE, DWI, T<sub>2</sub>WI, DCE+DWI, DCE+T<sub>2</sub>WI, DWI+T<sub>2</sub>WI, and DCE+DWI+T<sub>2</sub>WI and a clinical predictive model built based on the visual assessment of radiologists. A nomogram was constructed with the best radiomics signature combined with patient characteristics. The calibration curves for the radiomics signature and nomogram were conducted, combined with the Hosmer-Lemeshow test.

**Results:** Pearson's correlation was used to eliminate 3,329 irrelevant features, and then LASSO and logistic regression were used to screen the remaining feature coefficients for each model we built. Finally, 12 related features were obtained in the model which had the best performance. These 12 features were used to build a radiomics model in combination with the actual clinical diagnosis of benign or malignant lesion labels we have obtained. The best model built by 12 features from the 3 sequences has an AUC value of 0.939 (95% CI, 0.884–0.994) and an accuracy of 0.931 in the testing cohort. The sensitivity, specificity, precision and Matthews correlation coefficient (MCC) of testing

cohort are 0.932, 0.923, 0.982, and 0.791, respectively. The nomogram has also been verified to have calibration curves with good overlap.

**Conclusions:** Radiomics is beneficial in the malignancy prediction of BI-RADS 4 breast lesions. The radiomics predictive model built by the combination of DCE, DWI, and T<sub>2</sub>WI sequences has great application potential.

**Keywords:** LASSO, BI-RADS 4, breast lesion, magnetic resonance imaging, radiomics

## INTRODUCTION

The 2020 Global Cancer Report released by the International Agency for Research on Cancer (IARC) shows that female breast cancer has replaced lung cancer as the most common cancer in the world with an increase of approximately 2.3 million new cases (11.7%) throughout the year (1). Breast cancer is also estimated to top the list of new morbidity and mortality among all types of cancers in women. During treatment, the diagnosis of benign or malignant breast lesions has become the most basic and important step in the treatment of breast diseases.

According to the breast cancer screening guidelines of the National Comprehensive Cancer Network (NCCN) and BI-RADS, suspicious lesions can be classified into 6 categories (2). The fourth category of breast disease is defined as a type of breast lesions with suspicious malignancy and uncertain pathological types. The probability of being malignant is 2%–95%, although this type of breast disease has a further classification of 4a, 4b, and 4c, because of the large range of the possibility of the existence of malignant lesion, all patients with BI-RADS 4 breast diseases are recommended to undergo biopsy of suspicious areas to clarify their pathological properties (3). Because of the blurry qualitative characteristics of BI-RADS 4 breast diseases, we can see that most patients with BI-RADS 4 of breast diseases are overdiagnosed and treated with the puncture case analysis, which requires a certain degree of trauma to the body. Moreover, the clinical diagnosis inevitably has a certain false-positive rate and missed diagnosis rate (4, 5). Therefore, we propose a hypothesis to establish a predictive radiomics model based on the patients' preoperative imaging information, thereby avoiding patients with benign breast lesions from undergoing invasive pathological testing.

Computer-aided diagnosis methods based on medical imaging have been increasing in clinical application value in recent years. This method only needs to mark the abnormal signs, and then perform common image processing on this basis to get the diagnosis result. Therefore, the concept of radiomics came into being (6). Radiomics is a research method that extracts high-throughput image features from medical images and conducts quantitative research. DCE, DWI, and T<sub>2</sub>WI are three routine methods for the diagnosis and observation of breast diseases. Breast DCE imaging has high sensitivity in breast cancer screening for women who have accumulated breast cancer risk for more than 20%–25% (7). However, its specificity depends on a variety of external factors, such as the professional skills of the reader or the method of using quantitative techniques. DWI can characterize

the three-dimensional fluidity of water in the body and indirectly detect and visualize the microstructure (8). DWI and apparent diffusion coefficient (ADC) have been successfully applied to the clinical diagnosis and screening of breast cancer. Compared with the average specificity of 80% in DCE for breast cancer diagnosis, the average specificity of the combined diagnosis of DCE and DWI can reach 89.2% (9). Therefore, radiologists mainly use the combination of these two technologies in the diagnosis of breast cancer. T<sub>2</sub>WI is usually used to exclude cysts, intramammary lymph nodes, and other benign breast lesions (10). One of the most advantageous characteristics of T<sub>2</sub>WI is that signal strength is directly related to the underlying disease state for most breast cancer lesions mainly showing uneven or slightly high signal on T<sub>2</sub>WI MRI, while the surrounding tissues show low or medium signal (11). As existing research shows that the T<sub>2</sub>WI MRI image has a strong ability to interpret the pathology and diagnosis of breast diseases (12). Research combining T<sub>2</sub>WI MRI, DCE MRI, and DWI MRI to increase the actual diagnosis efficiency also picture a wider application of combining multiple MRI features in the field of breast diagnosis (13, 14).

It can be seen from the above results that the combined application of different MRI images can play a greater practical role in the diagnosis of breast diseases. This study aims to establish an auxiliary diagnosis prediction model that can be used to predict benign and malignant breast lesions by combining MRI images with three sequences of DCE, DWI, and T<sub>2</sub>WI.

## MATERIALS AND METHODS

This study had been reviewed and approved by the Ethics Committee of Cancer Hospital, Chinese Academy of Medical Sciences, and had been in line with the Declaration of Helsinki. All patients participating in this study waived the requirement of informed consent. The image processing methods had met the terms and conditions mentioned in the Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis (TRIPOD), Image Biomarker Standardization Initiative (IBSI), Checklist for Artificial Intelligence in Medical Imaging (CLAIM).

### Patients

The patient information for this research study comes from the picture archiving and communication system (PACS) of the Cancer Hospital of the Chinese Academy of Medical Sciences. Due to the following criteria, we continuously enrolled a total of

216 research subjects from 230 patients who received diagnosis at our hospital from September 2018 to December 2019. Each of the patients was classified with the fifth version of BI-RADS guidelines. The criteria for inclusion in the group were as follows: (i) Patients have been diagnosed with a BI-RADS type IV breast lesion. (ii) The pathological diagnoses results were confirmed by puncture pathology diagnoses. (iii) Patients have complete MRI images in PACS with axial DCE, DWI, and T<sub>2</sub>WI sequences obtained before patients underwent biopsy.

All the in-group patients were separated into a training cohort of 144 patients and a testing cohort of 72 patients divided by scan time from early to late according to the ratio of 2:1. We use leave-one-out cross-validation (LOOCV), and the training cohort is used to train and validate the model. Testing cohort is used to test the performance of the model.

## Semiautomatic Image Segmentation and Feature Extraction

We collected image information of the 216 enrolled patients, which were composed of three kinds of MRI images: DCE, DWI, and T<sub>2</sub>WI. All patients' MRI scans were completed before they underwent biopsy. Patients in training and testing cohorts were scanned using the same equipment. The DCE sequences were obtained by a higher axial resolution T1-weighted DCE imaging with a temporal resolution of 90 s. The DWI sequences were obtained by a DWI sequence with two *b*-values (0–1,000 s/mm<sup>2</sup>). The T<sub>2</sub>WI sequences were obtained from a higher axial resolution T<sub>2</sub>WI turbo spin-echo sequence. All scans were done at a magnetic field strength of 3.0 T.

These MRI images were reviewed by 2 radiologists who were not aware of the real pathological diagnoses. They calibrated and refined the segmentation results from Radiomics ([www.radiomics.net.cn](http://www.radiomics.net.cn)) of the breast lesion regions of interest (ROI) of the patients' pretreatment images in the DCE, DWI, and T<sub>2</sub>WI views. This software is a computerized semiautomatic image segmentation software with high accuracy trained by using a deep learning model. It uses the most recognized nnU-Net model framework, which is a generalized U-net framework and can obtain better training results through a careful preprocessing process and diverse network training schemes. An example of Radiomics software combined with manual fine-tuning for semiautomatic image segmentation is shown in **Figure 1**.

The specific operations of feature extraction were all done through Python (3.7.9), where the "pyradiomics" package ([pyradiomics.readthedocs.io](http://pyradiomics.readthedocs.io)) was used in conjunction. The hyperparameters of feature extractor set as follows: Laplacian of Gaussian filter: sigma: [2.0, 3.0, 4.0, 5.0]; normalize: True; normalize scale: 100; resampled pixel spacing (1, 1, 1); interpolator: sitkBSpline; binWidth: 5; and voxelArrayShift: 300. The other settings are default. A total of 3,474 quantitative original image features including first-order statistics features (18 features), 3D shape-based features (14 features), gray-level cooccurrence matrix features (24 features), gray-level run length matrix features (16 features), gray-level size zone matrix features (16 features), and gray-level dependence matrix features (14 features) were extracted from the

corresponding regions of interest in the original images in the DCE, DWI, and T<sub>2</sub>WI views (15). The original images were later added with wavelet filtering and Laplace of Gaussian (LoG) filtering respectively to extract the above features. Wavelet filtering was aimed to yield 8 decompositions per level of the original images. LoG filtering was used as an edge enhancement filter which emphasized areas of gray level changes, where sigma defines how the emphasized textures are supposed to be (16, 17).

## Radiomics Signature Building

Due to the same coarse-to-fine strategy, we filtered the features to prevent the model from overfitting (18–20). First, a mono-factor analysis was performed on all features, and all features were ranked in order from smallest to largest *p*-value, and the top 5% of features were filtered out. Second, the LASSO algorithm made the image features most relevant to the BI-RADS 4 breast lesions to be filtered out by compressing the correlation coefficients of some of the features and zeroing out another part of the coefficients. Finally, the LOOCV was performed to select a model. After the above steps are completed, the coefficients of most radiomics features were compressed to zero. Then, a radiomics signature was established based on a linearly weighted combination of features with nonzero coefficients (21). The above content is the entire process of our training model. The feature extraction and selection process were implemented in Python (3.7.9).

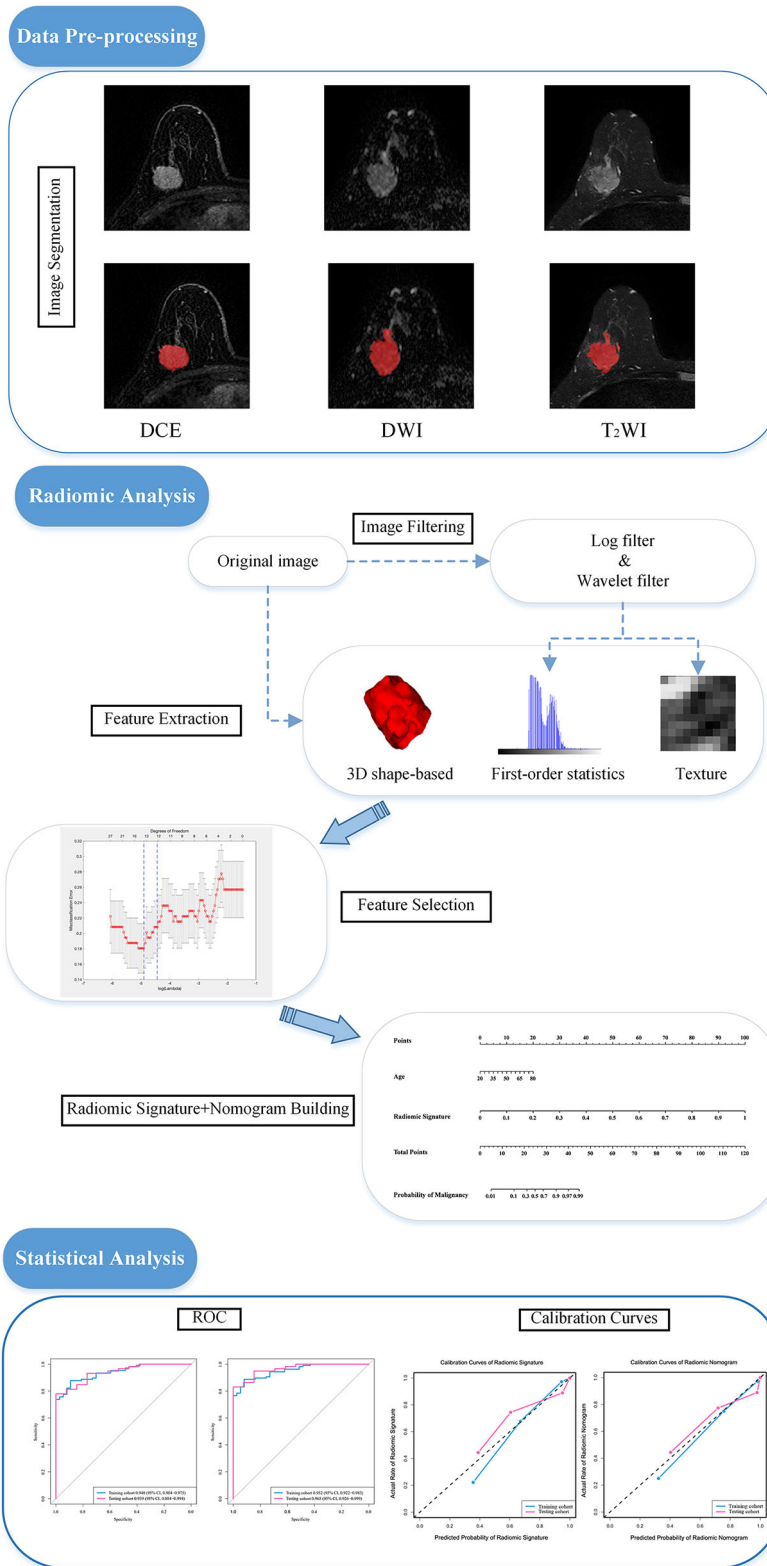
The ability of the model to recognize benign and malignant lesions is evaluated by the drawn ROC. ROC can directly show the sensitivity (true-positive rate) and false-positive rate of the predicted results of the model. Meanwhile, the AUC value and the accuracy value are two main indicators that can be indirectly obtained from the curve. In addition, we also calculated precision and MCC metrics to evaluate the model.

## Development and Performance of the Models Built With Sequences Combinations

Based on the above modeling steps, we performed prediction model building using different combinations of the three sequences, respectively. First, we modeled the three sequences individually based on the eigenvalue data extracted from each sequence. Second, the feature data of each two sequences were combined and modeled based on this data. Finally, we modeled the data using the combined data of the three sequences. These seven different models were used to compare performance in the same computational way.

## Development and Performance of the Radiomics Nomogram

To verify that radiomics signature combined with clinical factor has a greater predictive ability, we built the radiomics nomogram (21). Through the progress of organizing the patients' information and univariate logistic regression analysis, we selected age as the clinical factor used for radiomics nomogram building. We constructed a calibration curve for the nomogram to demonstrate its predictive efficacy.



**FIGURE 1** | Image segmentation and the procedure of developing a predictive model. The experiment is divided into three main parts: image pre-processing, image radiomic analysis and statistical analysis. Radiomic analysis includes image feature extraction and feature filtering. DCE, dynamic contrast enhanced imaging; DWI, diffusion weighted imaging; T<sub>2</sub>WI, 2 weighted imaging; ROC, receiver operating characteristic curve.

## Imaging Diagnosis by the Radiologists

In order to further verify the application value of the prediction model, we invited two radiologists to participate in our model verification step. Both radiologists come from the Cancer Hospital of the Chinese Academy of Medical Sciences and have more than 10 years of clinical diagnosis experience. The experiment requires the two experts to combine DCE, DWI, and T<sub>2</sub>WI MRI images of each patient without contacting the patient and completely ignorant of the clinical information and actual diagnosis results to jointly make their pathological diagnosis of each case of whether benign or malignant breast lesion judgments. During the actual experiment, two radiologists conducted pathological identification of all patients through visual observation of images. By comparing the prediction results of our model with the diagnosis results of these two experts, we can evaluate whether this model has value in clinical application.

After obtaining the diagnosis results of the two radiologists, we used the same calculation method to calculate the accuracy, sensitivity, specificity, precision, and MCC of the radiologist's diagnoses and compare them with the corresponding indicators of the radiomics model.

## Data and Statistical Analysis

To further evaluate the results, we plotted nomograms from the clinical information and plotted calibration curves combined with the Hosmer-Lemeshow test for the nomograms and the ROC curves of the radiomics model with the strongest performance. The p-value was calculated by a two samples *t*-test to evaluate the degree of group differentiation of the data. The above statistical evaluation work was done through R version 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria).

## RESULTS

### Clinical Characteristics of the Patients

The clinical characteristics of the patients in the training cohort and the testing cohort are shown in **Table 1**. The benign rate of

breast lesion in the training cohort is 25.69% and that of the testing cohort is 18.05%. The differences in imaging devices, age, and other clinical characteristics were not statistically significant between the two cohorts.

### Feature Selection, Radiomics Signature Development, and Validation

Twelve features were derived from DCE, DWI, and T<sub>2</sub>WI MRI images with the principle of coarse-to-fine. The selected features and corresponding coefficients are listed in **Table 2**.

### Development, Performance, and Evaluation of the Prediction Models

By comparing the ROC curves obtained after building radiomics prediction models for single sequences separately and for every two sequences combined with the ROC curve of radiomics prediction models built by combining the three sequences, the radiomics models built by combining the DCE, DWI, and T<sub>2</sub>WI sequences obtained a significant performance improvement. Also, the AUC value of the model based on three sequences combined is the highest among all AUC values we have got. The performance of our joint prediction model is significantly improved, especially in terms of the specificity of the model prediction. The results are shown in **Table 3**. The ROC curves are shown in **Figure 2**.

### Assessment of the Radiomics Nomogram

We conducted the radiomics nomogram by combining the patients' ages and the radiomics signature. The nomogram has achieved an AUC value of 0.965 (95% CI, 0.926–0.999) and an accuracy of 0.912 in the testing cohort. The radiomics nomogram is shown in **Figure 3**. The calibration curves of the radiomics nomogram and the radiomics model based on DCE, DWI, and T<sub>2</sub>WI are shown in **Figure 4**.

### Comparison of the Prediction Results Between Radiologists and the Radiomics Prediction Model

Two radiologists who have reviewed the in-group patients' images reached a good consensus of their visual diagnosis

**TABLE 1** | Basic clinical information of enrolled patients.

Characteristics	Training (N = 144)	Testing (N = 72)	Total (N = 216)	P-value
Age at surgery (years), median (range)	45 (22-72)	45 (23-78)	45 (22-78)	0.052 <sup>a</sup>
Benign (%)	37 (25.69)	13 (18.06)	50 (23.15)	0.867 <sup>b</sup>
Adenomatosis	9 (6.25)	3 (4.17)	12 (5.56)	
Phyllodes tumor	28 (19.44)	10 (13.89)	38 (17.59)	
Malignant (%)	107 (74.31)	59 (81.94)	166 (76.85)	0.066 <sup>b</sup>
Invasive ductal carcinoma	67 (46.53)	28 (38.89)	95 (43.98)	
Colloid carcinoma	20 (13.89)	11 (15.28)	31 (14.35)	
Medullary carcinoma	18 (12.60)	19 (25.39)	37 (17.13)	
Neuroendocrine carcinoma	1 (0.69)	1 (1.39)	2 (0.93)	
Solid papillary carcinoma	1 (0.69)	0 (0.00)	1 (0.46)	

The differentiation in the characteristics (age when diagnosed, benignity and malignancy, pathological diagnosis) in the training cohort and the Testing cohort were evaluated. P-value less than 0.05 proves that the groups are significantly different. The above P-values show the training and the Testing cohorts are non-significantly different.

<sup>a</sup>P-value was calculated by two sample *t*-test.

<sup>b</sup>P-values were calculated by Fisher exact test.

**TABLE 2** | Results of the feature selection for the model based on DCE, DWI and T<sub>2</sub>WI.

Sequences	Features	Coefficients
T <sub>2</sub> WI	Wavelet LHH glrlm long run low gray level emphasis	-0.19765
	Original glszm Gray Level Non-Uniformity Normalized	-0.52642
	Wavelet LLL glszm Small Area Low Gray Level Emphasis	-1.53857
	Wavelet LHL glszm Low Gray Level Zone Emphasis	-1.34546
	Wavelet LLH glszm Small Area Low Gray Level Emphasis	-6.49947
	Log sigma 4-0-mm-3D glrlm Long Run Low Gray Level Emphasis	-0.00238
	Log sigma 4-0-mm-3D glrlm Long Run Emphasis	-0.00004
DCE	Wavelet LHH glszm Gray Level Non-Uniformity Normalized	-1.11352
	Original glszm Small Area Emphasis	4.57208
	Wavelet LHH glcm Correlation	12.60474
DWI	Wavelet LLL glcm Inverse Variance	-4.61787
	Wavelet HHH glrlm Long Run High Gray Level Emphasis	-0.02102

The features were selected by LASSO algorithm. These coefficients show the magnitude of the weight of their corresponding characteristics in the regression model.

**TABLE 3** | Results of the radiomic models and the models based on the nomogram and visual assessment.

	AUC		Accuracy		Sensitivity/Recall		Specificity		Precision		MCC	
	Training (95% CI)	Testing (95% CI)	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
DCE	0.901 (0.853-0.949)	0.844 (0.741-0.946)	0.806	0.819	0.804	0.864	0.811	0.615	0.925	0.911	0.561	0.444
DWI	0.871 (0.812-0.930)	0.798 (0.689-0.907)	0.847	0.801	0.846	0.814	0.865	0.769	0.947	0.941	0.651	0.493
T <sub>2</sub> WI	0.877 (0.822-0.932)	0.838 (0.731-0.940)	0.868	0.777	0.879	0.746	0.838	0.846	0.940	0.957	0.679	0.492
DCE+DWI	0.932 (0.894-0.970)	0.821 (0.727-0.916)	0.861	0.708	0.860	0.712	0.865	0.692	0.948	0.913	0.675	0.324
DCE+T <sub>2</sub> WI	0.924 (0.880-0.968)	0.853 (0.751-0.957)	0.889	0.820	0.907	0.813	0.838	0.846	0.942	0.96	0.721	0.551
DWI+T <sub>2</sub> WI	0.889 (0.837-0.941)	0.834 (0.731-0.938)	0.882	0.777	0.869	0.780	0.919	0.769	0.969	0.939	0.730	0.453
DCE+DWI+T <sub>2</sub> WI	0.940 (0.904-0.975)	0.939 (0.884-0.994)	0.924	0.931	0.935	0.932	0.892	0.923	0.961	0.982	0.806	0.791
Nomogram	0.952 (0.922-0.983)	0.965 (0.926-0.999)	0.896	0.912	0.887	0.932	0.919	0.846	0.969	0.965	0.756	0.737
Visual Assessment	0.613 (0.528-0.712)	0.563 (0.470-0.772)	0.632	0.611	0.644	0.627	0.594	0.538	0.821	0.860	0.21	0.130

The 9 models contain models based on DCE, DWI, T<sub>2</sub>WI, DCE+DWI, DCE+T<sub>2</sub>WI, DWI+T<sub>2</sub>WI, DCE+DWI+T<sub>2</sub>WI, nomogram and visual assessment.

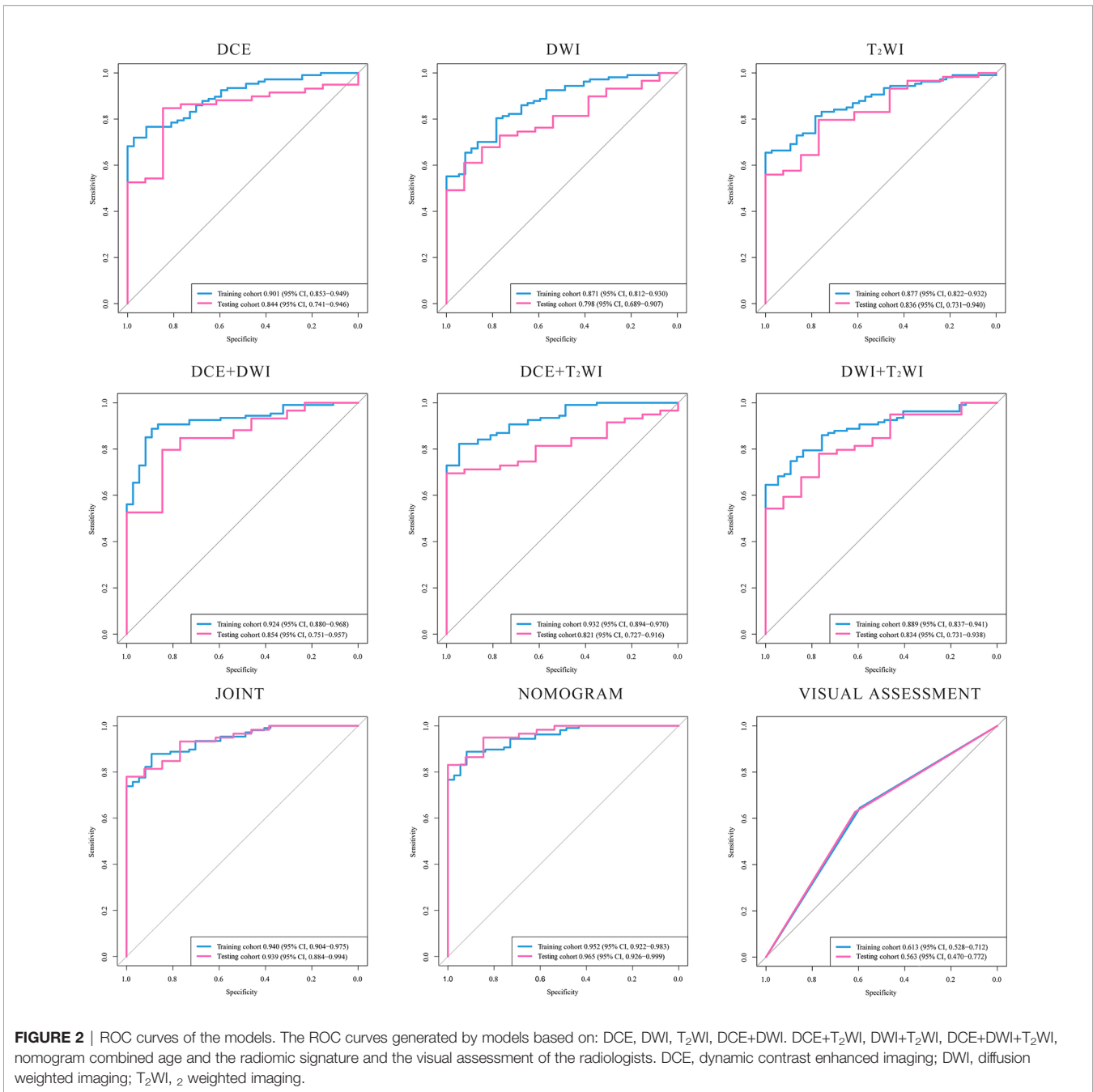
upon the DCE, DWI, and T<sub>2</sub>WI sequences' combination. The results from the radiologists have been used to make a further assessment. According to the model built with the visual assessment achieved from the experts, the AUC value is 0.563 (95% CI, 0.470–0.772), and the accuracy is 0.611 in the testing cohort. The results of the DeLong test showed that the predictive AUC values and accuracy of the radiomics prediction model with the combination of the three sequences were significantly higher than those of the prediction model built by the experience of radiologists.

## DISCUSSION

In this study, we conducted a study on the relationship between the malignancy of BI-RADS 4 breast lesions and the imaging features in DCE, DWI, and T<sub>2</sub>WI MRI, and developed a radiomics prediction model based on MRI. It has been proved

that the model has a stronger predictive ability than radiologists' empirical predictions and can accurately identify the benign and malignant BI-RADS 4 breast lesions, which has application value (22).

Breast MRI has the advantages of good soft-tissue resolution and no radiation and is significantly better than mammography and ultrasonography for early diagnosis and local staging of breast cancer (23, 24). Because MRI is insensitive to microcalcifications and requires a high degree of magnetic field homogeneity, it is easy to cause a false-negative diagnosis (25). Inaccuracies in visual assessment can also lead to overdiagnosis of patients. As a definition from the NCCN guidelines points out, the possibility of the occurrence of BI-RADS 4 breast lesion malignancy range from 2% to 95%, but the actual PPV of breast lesion ranges from 25.7% to 59.2% (26–28). Also, the radiomics model we developed combines multimode MRI images to provide diagnostic doctors and oncologists with a quantitative evaluation tool with greater reliability.

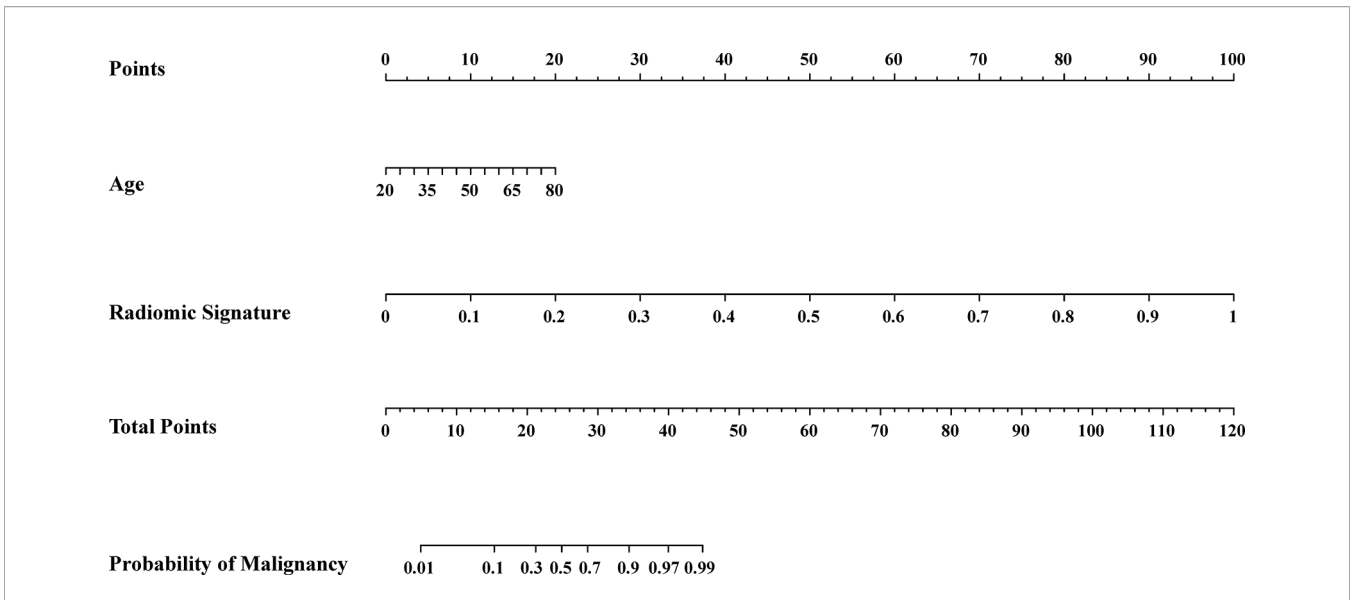


In summary, our study focuses the concentration on developing a way to predict malignant breast lesions through imaging data without causing trauma to the patients.

Among the finally obtained features after extraction and compression by the LASSO logistic regression, the number of T<sub>2</sub>WI image features is the largest, the coefficient value of DCE image features is the largest, and only one DWI feature is selected and the correlation coefficient of this feature is relatively small, which shows that the image features of T<sub>2</sub>WI and DCE have a greater influence on the benignity and malignancy of breast lesions. Compared with other parametric images, T<sub>2</sub>WI MRI images are more likely to reflect the cyst, the margins of the

lesion, and the surrounding lymph nodes (29). Therefore, T<sub>2</sub>WI images are often used to detect benign lesions (12). DCE scanning is the most common scan used to diagnose breast disease in clinical practice; it provides higher sensitivity, however, its specificity is variable (30). It can be concluded that T<sub>2</sub>WI scanning and DCE scanning have some complementary properties, and therefore a prediction model combining both sequences can yield substantial performance improvements.

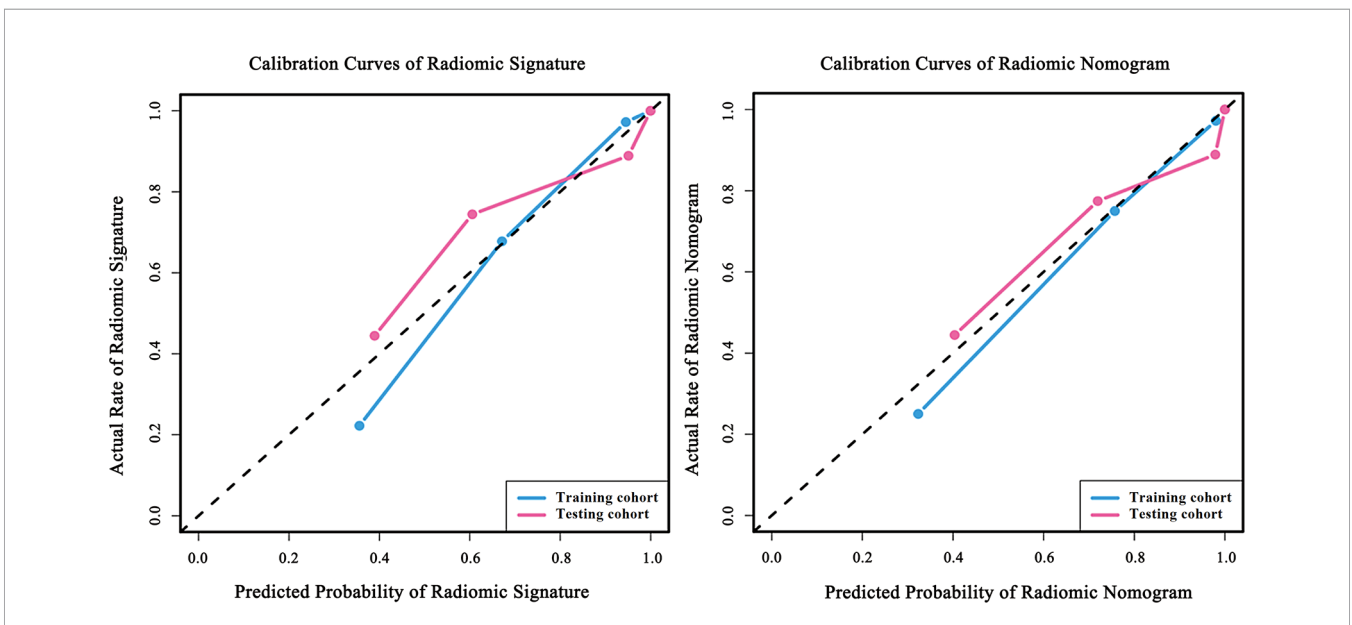
From the features we screened out, wavelet-filtered small area low gray-level emphasis texture feature in the T<sub>2</sub>WI sequence and wavelet-filtered correlation texture feature in the DCE



**FIGURE 3 |** Radiomic nomogram. The radiomic nomogram was conducted based on the patients' ages from the clinical information and the radiomic signature obtained from the best radiomic model which was based on DCE, DWI and T<sub>2</sub>WI.

sequence was the most descriptive for breast BI-RADS4-like suspicious lesions. Small area low gray-level emphasis is one of the gray-level size zone matrix (GLSZM) features, which describe the amount of homogeneous connected areas within the volume of a certain size and intensity, thereby describing lesion heterogeneity at a regional scale (26). The grayscale area size matrix is the primary form of the Thibault matrix, which is an advanced statistical matrix of texture features and a powerful tool for medical image analysis. The more homogeneous the image

texture is, the larger and flatter the matrix width is. Unlike the stroke and cooccurrence matrices, the GLSZM does not require multiple directional calculations. Specifically, GLSZM is effective in characterizing texture consistency, nonperiodic or speckled textures, and has better performance than granularity, stroke matrix, and cooccurrence matrix for cell nuclei, and dermis (27). Wavelet filtered correlation texture feature is one of the gray level cooccurrence matrix (GLCM) features, which describes the joint distribution of two pixels that have some spatial location



**FIGURE 4 |** Calibration curves of the radiomic model and the radiomic nomogram. Calibration curves of radiomic signature were built by the radiomic model based on DCE, DWI and T<sub>2</sub>WI. Calibration curves of radiomic nomogram were built by the nomogram. The diagonal line represents the perfect prediction of the ideal model. The blue and pink lines represent the performance of the training and testing cohort in the models, where the models closer to the diagonal line represent better predictions. The calibration curves have gone through the Hosmer-Lemeshow test and have achieved a favorite result.



relationship (28). The correlation features to measure the degree of similarity of the elements of the spatial grayscale cooccurrence matrix in the row direction. Thus, the local grayscale correlation in the image can be seen from the correlation value magnitude. When the values of matrix elements are uniformly equal, the correlation value is large; contrarily, the correlation value is small when the matrix pixel values differ significantly. If there is a horizontal texture in the image, the correlation value of the horizontal matrix is larger than that of the rest of the matrixes.

Another noteworthy strength of this study is our image segmentation method. The images used in this study were automatically segmented by computer using an optimized deep learning model and then corrected and refined by 2 professional radiologists, so our regions of interest segmentation have a high degree of accuracy and precision. This shows that the image feature values we extracted in this research-based learning are also more convincing. As for image feature values are the basis for the establishment of our prediction model, so this is one of the reasons that can prove that the prediction denseness has application value.

In addition, to demonstrate that this prediction model can reach a higher level than physicians' diagnosis, we invited two radiologists who have more than 10 years of experience in breast cancer. The two doctors had no prior information about the patient's personal information and real diagnosis; they were able to determine the benignity or malignancy of the patient's breast lesions by combining only three sequences of MRI images. The results confirmed that the AUC value, accuracy, sensitivity, specificity precision, and MCC of our developed radiomics prediction model were much higher than those of the radiomics experts.

Medical imaging technology is able to capture a vast amount of information, but most of information was reported in a qualitative and quantitative way. Prospective studies indicated that the computer aided detection (CAD) system constructed using extracted and selected features can effectively distinguish benign and malignant breast lesions (31, 32). Chen et al. (33) extracted a large number of dynamic features from the temporal enhancement pattern of a tumor. Zheng et al. (34) extracted dynamic enhancement and architectural features and spatial variations of pixelwise temporal enhancements from MRI images. Although these studies had also achieved good results, due to the development of medical imaging technique, these features cannot contain rich information on tumors. In 2012, Lambin et al. (17) formally proposed radiomics which attracted the attention of many computer scientists, radiologists, and oncologists. Liu et al. (22) used deep learning to extract features from mammography-based in predicting malignancy. Karen et al. (35) used datasets that contained 64 lesions DCE-MRI images and extracted 38 radiomics features from each image to build SVM models which is able to distinguish the malignant and benign lesions. However, the datasets in these reports are all single modality. Extracting more features can discover the connection between deeper features and the training task. The higher dimensional radiomics features can more fully express tumor heterogeneity, and these features can in part

describe the characteristics of breast cancer based on their usefulness, predictive power and uniqueness. In our research, we extracted 3,474 radiomics features from a combination of three sequences (DCE, DWI, T2WI) MRI images to build the model. Thus, we have achieved a satisfactory result.

Our research results confirmed that the multimodal fusion models can complement each other. When one modality cannot obtain obvious information of a single modality, another modality can provide weak supervision information for it. Among the single-sequence models, the model constructed based on DCE scanning technology has the highest AUC. DCE sequence cannot only clearly show morphologic and hemodynamics features of the lesion (36) but also have more significance for the sharpening of the shape and scope of the lesion according to various manifestations such as enhancement mode, blood supply, and cell composition of the lesion. In the dual-sequence combined model, the DCE+T2WI model has higher model evaluation performance than the single-sequence model in terms of accuracy, specificity, and sensitivity. The signal intensity of T2WI is directly related to the shape of the underlying lesion and is usually used to exclude cysts, intramammary lymph nodes, and other benign breast lesions to improve the specificity of diagnosis (37). Among all the models in this study, the best model effect is DCE+T2WI+DWI. Each model of a single sequence has advantages and disadvantages, but the fusion of the three sequences can assist each other (37, 38). By comparing the single-sequence model with the multisequence model, it was found that the multimodal fusion models had better performance in predicting the benign and malignant of BI-RADS 4 breast lesions.

Upon reflection, although this study provides significant benefits, it also has some limitations worth discussing. Firstly, this is a single-center study. Therefore, the lack of multicenter data fusion analysis might affect the generalization ability of the model to a certain extent. Secondly, this prediction model is based on three sequences of images from breast MRI and can combine the characteristics of different images for prediction. However, breast ultrasonography and mammography are also two major means of detecting breast disease (39, 40). Because of the higher cost of MRI, the more restrictive population, and the limitations of an incomplete examination, mammography, and ultrasound scans are even more routinely used as clinical tests (41, 42). So, the absence of these two medical images in this study could potentially cause our model to be less representative, as our prediction model could not be applied to these two sequences of images.

In conclusion, the model based on DCE, DWI, and T<sub>2</sub>WI combined is the most effective in predicting the benignity and malignancy of BI-RADS 4 suspicious lesion.

## DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because "Due to the privacy of patients, the MRI data and clinical information related to patients cannot be available for public

access". Requests to access the datasets should be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Cancer Hospital, Chinese Academy of Medical Sciences. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

ReZ and WW: study concept, acquisition, analysis and interpretation of data, statistical analysis, and drafting of the article. RL and JL: acquisition, analysis and interpretation of data,

statistical analysis, and drafting of the article. ZZ: study concept, acquisition, analysis and interpretation, and critical revision of article. MM and RuZ: acquisition and analysis of data. XZ: study concept and design, interpretation of data, and drafting and critical revision of the article. All authors revised the manuscript critically for important intellectual content and approved the final submitted version.

## FUNDING

This work was sponsored by the National Natural Science Foundation of China (81901827), Key project of Beijing Hope Marathon Special Fund from China Cancer Foundation (LC2018A20), Natural Science Basic Research Program of Shaanxi (2020JQ-836, 2020JQ-837), Xi'an Polytechnic University (BS201987), and National Key Research and Development Plan of China (2017YFC1309102).

## REFERENCES

- Wild C, Weiderpass E, Stewart B. *World Cancer Report: Cancer Research for Cancer Prevention*. Lyon: International Agency for Research on Cancer (2020) p. 23–33.
- Goetz MP, Gradishar WJ, Anderson BO, Gradishar WJ, Anderson BO, Abraham J, et al. NCCN Guidelines Insights: Breast Cancer, Version 3.2018: Featured Updates to the NCCN Guidelines. *J Natl Compr Cancer Net* (2019) 17(2):118–26. doi: 10.6004/jnccn.2019.0009
- Bennani-Baiti B, Dietzel M, Baltzer PA. MRI for the Assessment of Malignancy in BI-RADS 4 Mammographic Microcalcifications. *PLoS One* (2017) 12(11):e0188679. doi: 10.1371/journal.pone.0188679
- Gradishar W, Anderson B, Balassanian R, Blair SL, Burstein H, Cyr A, et al. NCCN Guidelines Insights: Breast Cancer. *Version* (2017) 1:15. doi: 10.6004/jnccn.2017.0044
- Kuhl CK. Abbreviated Magnetic Resonance Imaging (MRI) for Breast Cancer Screening: Rationale, Concept, and Transfer to Clinical Practice. *Annu Rev Med* (2019) 70:501–19. doi: 10.1146/annurev-med-121417-100403
- D'Orsi C, Bassett L, Feig S. Breast Imaging Reporting and Data System (BI-RADS). New York: Oxford University Press. *Breast Imaging Atlas* (2018). doi: 10.1093/med/9780190270261.003.0005
- Kumar V, Gu Y, Basu S, Berglund A, Eschrich S, Schabath M, et al. Radiomics: The Process and the Challenges. *Magn Reson Imaging* (2012) 30(9):1234–48. doi: 10.1016/j.mri.2012.06.010
- Yankeelov TE, Lepage M, Chakravarthy A, Broome EE, Niermann KJ, Kelley MC, et al. Integration of Quantitative DCE-MRI and ADC Mapping to Monitor Treatment Response in Human Breast Cancer: Initial Results. *Magn Reson Imaging* (2007) 25(1):1–13. doi: 10.1016/j.mri.2006.09.006
- Partridge SC, McDonald ES. Diffusion Weighted Magnetic Resonance Imaging of the Breast: Protocol Optimization, Interpretation, and Clinical Applications. *Magn Reson Imaging Clin* (2013) 21(3):601–24. doi: 10.1016/j.mric.2013.04.007
- Fan M, Yuan W, Zhao W, Xu M, Wang S, Gao X, et al. Joint Prediction of Breast Cancer Histological Grade and Ki-67 Expression Level Based on DCE-MRI and DWI Radiomics. *IEEE J Biomed Health Inf* (2019) 24(6):1632–42. doi: 10.1109/JBHI.2019.2956351
- Liu Z, Meng X, Zhang H, Li Z, Liu J, Sun K, et al. Predicting Distant Metastasis and Chemotherapy Benefit in Locally Advanced Rectal Cancer. *Nat Commun* (2020) 11(1):1–11. doi: 10.1038/s41467-020-18162-9
- Dong Y, Feng Q, Yang W, Lu Z, Deng C, Zhang L, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer Based on Radiomics of T2-Weighted Fat-Suppression and Diffusion-Weighted MRI. *Eur Radiol* (2018) 28(2):582–91. doi: 10.1007/s00330-017-5005-7
- Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. *Clin Cancer Res* (2019) 25(12):3538–47. doi: 10.1158/1078-0432.CCR-18-3190
- Li W, Newitt DC, Gibbs J, Wilmes L, Jones E, Arasu V, et al. Predicting Breast Cancer Response to Neoadjuvant Treatment Using Multi-Feature MRI: Results From the I-SPY 2 TRIAL. *NPJ Breast Cancer* (2020) 6(1):1–6. doi: 10.1038/s41523-020-00203-7
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5(1):1–9. doi: 10.1038/ncomms5006
- Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
- Tibshirani R, Bien J, Friedman J, Hastie T, Simon N, Taylor J, et al. Strong Rules for Discarding Predictors in Lasso-Type Problems. *J R Stat Soc: Ser B (Statistical Methodology)* (2012) 74(2):245–66. doi: 10.1111/j.1467-9868.2011.01004.x
- Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Software* (2011) 39(5):1. doi: 10.18637/jss.v039.i05
- Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* (2010) 33(1):1. doi: 10.18637/jss.v033.i01
- Wei W, Wang K, Liu Z, Tian K, Wang L, Du J, et al. Radiomic Signature: A Novel Magnetic Resonance Imaging-Based Prognostic Biomarker in Patients With Skull Base Chordoma. *Radiother Oncol* (2019) 141:239–46. doi: 10.1016/j.radonc.2019.10.002
- Liu H, Chen Y, Zhang Y, Wang L, Luo R, Wu H, et al. A Deep Learning Model Integrating Mammography and Clinical Factors Facilitates the Malignancy Prediction of BI-RADS 4 Microcalcifications in Breast Cancer Screening. *Eur Radiol* (2021) 31(8):5902–12. doi: 10.1007/s00330-020-07659-y
- Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PHM, Monnikhof EM, et al. Supplemental MRI Screening for Women With Extremely Dense Breast Tissue. *N Engl J Med* (2019) 381(22):2091–102. doi: 10.1056/NEJMoa1903986

24. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* (2019) 9(5):1303. doi: 10.7150/thno.30309
25. Jiang Y, Edwards AV, Newstead GM. Artificial Intelligence Applied to Breast MRI for Improved Diagnosis. *Radiology* (2021) 298(1):38–46. doi: 10.1148/radiol.2020200292
26. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Shape and Texture Indexes Application to Cell Nuclei Classification. *Int J Pattern Recognit Artif Intell* (2013) 27(01):1357002. doi: 10.1142/S0218001413570024
27. Thibault G, Angulo J, Meyer F. Advanced Statistical Matrices for Texture Characterization: Application to Cell Classification. *IEEE Trans Biomed Eng* (2013) 61(3):630–7. doi: 10.1109/TBME.2013.2284600
28. Hall-Beyer M. *GLCM Texture: A Tutorial V. 3.0 March 2017*. (2017) Calgary: University of Calgary Press. doi: 10.11575/PRISM/33280.
29. Nketiah G, Elschot M, Kim E, Teruel J, Scheenen T, Bathen T, et al. T2-Weighted MRI-Derived Textural Features Reflect Prostate Cancer Aggressiveness: Preliminary Results. *Eur Radiol* (2017) 27(7):3050–9. doi: 10.1007/s00330-016-4663-1
30. Gubern-Mérida A, Martí R, Melendez J, Hauth JL, Mann RM, Karssemeijer N, et al. Automated Localization of Breast Cancer in DCE-MRI. *Med Image Anal* (2015) 20(1):265–74. doi: 10.1016/j.media.2014.12.001
31. Arbash Meinel L, Stolpen AH, Berbaum KS, Fajardo LL, Reinhardt JM. Breast MRI Lesion Classification: Improved Performance of Human Readers With a Backpropagation Neural Network Computer-Aided Diagnosis (CAD) System. *J Magn Reson Imaging* (2007) 25(1):89–95. doi: 10.1002/jmri.20794
32. Vadivel A, Surendiran B. A Fuzzy Rule-Based Approach for Characterization of Mammogram Masses Into BI-RADS Shape Categories. *Comput Biol Med* (2013) 43(4):259–67. doi: 10.1016/j.compbiomed.2013.01.004
33. Chen WJ, Giger ML, Bick U, Newstead GM. Automatic Identification and Classification of Characteristic Kinetic Curves of Breast Lesions on DCE-MRI. *Med Phys* (2006) 33(8):2878–87. doi: 10.1118/1.2210568
34. Zheng YJ, Englander S, Baloch S, Zacharaki E, Fan Y, Schnell M, et al. STEP: Spatiotemporal Enhancement Pattern for MR-Based Breast Tumor Diagnosis. *Med Phys* (2009) 36(7):3192–204. doi: 10.1118/1.3151811
35. Drukker K, Anderson R, Edwards A, Papaioannou J, Pineda F, Abe H, et al. Radiomics for Ultrafast Dynamic Contrast-Enhanced Breast MRI in the Diagnosis of Breast Cancer: A Pilot Study. In: *Conference on Medical Imaging - Computer-Aided Diagnosis*, vol. Volume 10575. Houston, TX: Proceedings of SPIE (2018).
36. Zhang Q, Peng Y, Liu W, Bai J, Zheng J, Yang X, et al. Radiomics Based on Multimodal MRI for the Differential Diagnosis of Benign and Malignant Breast Lesions. *J Magn Reson Imaging* (2020) 52(2):596–607. doi: 10.1002/jmri.27098
37. Dalmis MU, Gubern-Merida A, Vreemann S, Bult P, Karssemeijer N, Mann R, et al. Artificial Intelligence-Based Classification of Breast Lesions Imaged With a Multiparametric Breast MRI Protocol With Ultrafast DCE-MRI, T2, and DWI. *Invest Radiol* (2019) 54(6):325–32. doi: 10.1097/RLI.0000000000000544
38. Dijkstra H, Dorrius MD, Wielemans M, Pijnappel RM, Oudkerk M, Sijens PE. Quantitative DWI Implemented After DCE-MRI Yields Increased Specificity for BI-RADS 3 and 4 Breast Lesions. *J Magn Reson Imaging* (2016) 44(6):1642–9. doi: 10.1002/jmri.25331
39. Gary MT, Tan P-H, Cheung HS, Chu WC, Lam WW. Intermediate to Highly Suspicious Calcification in Breast Lesions: A Radio-Pathologic Correlation. *Breast Cancer Res Treat* (2008) 110(1):1–7. doi: 10.1007/s10549-007-9695-4
40. Liu Z, Zhang X-Y, Shi Y-J, Wang L, Zhu H-T, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Cancer Res* (2017) 23(23):7253–62. doi: 10.1158/1078-0432.CCR-17-1038
41. Fischer U. X-Ray Mammography: Calcifications, Masses and Architectural Distortions. *Der Radiol* (2008) 48(8):785–97, quiz 798. doi: 10.1007/s00117-008-1711-0
42. Lei C, Wei W, Liu Z, Xiong Q, Yang C, Yang M, et al. Mammography-Based Radiomic Analysis for Predicting Benign BI-RADS Category 4 Calcifications. *Eur J Radiol* (2019) 121:108711. doi: 10.1016/j.ejrad.2019.108711

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer KZ declared a shared parent affiliation with several of the authors, ReZ, JL, XZ, to the handling editor at time of review.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Wei, Li, Li, Zhou, Ma, Zhao and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.