European Journal of Radiology 121 (2019) 108711

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Mammography-based radiomic analysis for predicting benign BI-RADS category 4 calcifications

Chuqian Lei[a,b,1], Wei Wei[c,d,e,1], Zhenyu Liu[e], Qianqian Xiong[a,b], Ciqiu Yang[a], Mei Yang[a], Liulu Zhang[a], Teng Zhu[a], Xiaosheng Zhuang[a,f], Chunling Liu[g], Zaiyi Liu[g], Jie Tian[d,e,h,**], Kun Wang[b,a,*]

[a] The Second School of Clinical Medicine, Southern Medical University, Guangzhou, 510515, China
[b] Department of Breast Cancer, Cancer Center, Guangdong Provincial People's Hospital & Guangdong Academy of Medical Sciences, 510080, China
[c] School of Electronics and Information, Xi'an Polytechnic University, Xi'an, 710000, China
[d] Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, 710126, China
[e] CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[f] Shantou University Medical College, Shantou, 515041, Guangdong, China
[g] Department of Radiology, Guangdong Provincial People's Hospital, Guangzhou, 510080, China
[h] University of Chinese Academy of Sciences, Beijing, 100190, China

ARTICLE INFO

Keywords:
Radiomics
Breast
Calcification
Predictive value of test
Unnecessary procedures

ABSTRACT

Purpose: We developed and validated a radiomic model based on mammography and assessed its value for predicting the pathological diagnosis of Breast Imaging Reporting and Data System (BI-RADS) category 4 calcifications.

Materials and methods: Patients with a total of 212 eligible calcifications were recruited (159 cases in the primary cohort and 53 cases in the validation cohort). In total, 8286 radiomic features were extracted from the craniocaudal (CC) and mediolateral oblique (MLO) images. Machine learning was used to select features and build a radiomic signature. The clinical risk factors were selected from the independent clinical factors through logistic regression analyses. The radiomic nomogram incorporated the radiomic signature and an independent clinical risk factor. The diagnostic performance of the radiomic model and the radiologists' empirical prediction model was evaluated by the area under the receiver operating characteristic curve (AUC). The differences between the various AUCs were compared with DeLong's test.

Results: Six radiomic features and the menopausal state were included in the radiomic nomogram, which discriminated benign calcifications from malignant calcifications with an AUC of 0.80 in the validation cohort. The difference between the classification results of the radiomic nomogram and that of radiologists was significant ($p < 0.05$). Particularly for patients with calcifications that are negative on ultrasounds but can be detected by mammography (MG+/US- calcifications), the identification ability of the radiomic nomogram was very strong.

Conclusions: The mammography-based radiomic nomogram is a potential tool to distinguish benign calcifications from malignant calcifications.

Abbreviations: AUC, area under the receiver-operating characteristic curve; BI-RADS, Breast Imaging Reporting and Data System; CAD, computer aided diagnosis; CC, craniocaudal; CI, confidence interval; FNR, false negative rate; FPV, false predictive value; GLSZM, gray level size zone matrix; LASSO, least absolute shrinkage and selection operator procedure; LOOCV, leave one out cross-validation; MG, mammography; MLO, medio-lateral oblique; NCCN, National Comprehensive Cancer Network; PPV, positive predictive value; ROC, receiver-operating characteristic; ROI, region of interest; SVM, support vector machine; SZE, short zone emphasis; SZLGE, small zone low gray-level emphasis

* Corresponding author at: No. 123 Huifu West Road, Guangzhou, 510080, China.
** Corresponding author at: No. 95 Zhongguancun East Road, Beijing, 100190, China.
E-mail addresses: jie.tian@ia.ac.cn (J. Tian), gzwangkun@126.com (K. Wang).
[1] Chuqian Lei and Wei Wei contributed equally to this study.

https://doi.org/10.1016/j.ejrad.2019.108711
Received 20 April 2019; Accepted 11 October 2019
0720-048X/

1. Introduction

Calcifications account for 31 % of all breast lesions found with mammographic screening and can occur in benign or malignant breast lesions. Approximately 55 % of clinically nonpalpable breast cancer has been shown to be associated with calcifications [1,2]. According to the 5th edition of the Breast Imaging Reporting and Data System (BI-RADS) lexicon, all breast lesions are divided into 6 categories, and lesions of BI-RADS category 4 are considered to be the suspicious and of an unclear type.

Mammography and ultrasound are widely accepted as screening strategies in the early detection of breast diseases [3]. In particular, mammography is considered to be the most reliable technique for assessing the status of breast calcifications [4]. According to the BI-RADS lexicon and Breast Cancer Screening and Diagnosis Guidelines of the National Comprehensive Cancer Network (NCCN), the malignant possibility of BI-RADS category 4 diseases is 2 %–95 %. Although BI-RADS category four calcifications can be specifically classified into 4a, 4b and 4c, a biopsy is still recommended for this category to confirm the pathological properties [5,6].

The positive predictive value (PPV) of calcifications of all categories with mammography is generally under 30 %, and the PPV of BI-RADS category 4 calcifications ranges between 25.7 %–59.2 % [7–9]. These data suggested that part of mammographic calcifications yield benign pathological results [10]. False positive results reduce the positive predictive value and cause unnecessary invasive intervention in patients with benign calcifications, which aggravates the economic burden of both patients and the health care system [11]. Applying conventional image-based methods to discriminate benign calcifications from malignant calcifications remains challenging because there is currently no specific contrast agent available for component identification in breast calcifications [12]. We hypothesize that some BI-RADS category 4 calcification lesions could be downgraded and recommended for follow-up like BI-RADS category 3 lesions. It is therefore urgent to find a suitable method or tool to distinguish between benign and malignant lesions. Accordingly, benign calcifications can be followed up to reduce unnecessary invasive procedures.

Radiomic analysis based on artificial intelligence provides an opportunity for our study [13,14]. Radiomics is an emerging method for effective quantitative analysis and prediction using medical imaging big data. Radiomic analysis has been successfully employed in the field of oncology [15], including for breast carcinoma, ovarian carcinoma and rectal carcinoma [16–18]. We applied this emerging technology, which extracts quantitative features from clinical medical images that are unrecognizable to the naked eye, and used radiomics to analyze the associations between imaging features and pathological results to establish an individualized prediction model [19]. Radiomics not only enables oncologists to provide highly personalized care for tumor diagnosis and identifies phenotypic subtypes [20–22] but radiomics also noninvasively provides effective decisions at a low cost. Specifically, there have been no research studies on the predictive mammographic features of breast calcification-related lesions using radiomics.

The objective of the present research is to build and validate a radiomics-based nomogram using clinical risk factors and mammographic imaging features to predict if BI-RADS category 4 breast calcifications are malignant before clinical invasive intervention.

2. Materials and methods

2.1. Patients

This research recruited data retrospectively and was approved by the ethics committee, with the requirement for informed consent waived. The inclusion criteria were as follows: (i) breast calcification diagnosed with primary medical imaging; (ii) MG calcifications assigned to BI-RADS 4 category by radiologists; (iii) no obvious lesions found in the calcification-related area in mammography; and (iv) biopsies with available pathological results. The exclusion criteria were: (i) lack of craniocaudal (CC) or mediolateral oblique (MLO) images; (ii) did not receive pathological biopsy in our hospital; and (iii) MG images are not good enough to get measured values (i.e. over-exposure and calcification lesions are not showed on the image).

Patients with a total of three hundred twenty-seven calcifications classified as BI-RADS category 4 were screened from May 2015 to March 2018 (Fig. A.1 in supplementary material.). Among these patients, we consecutively enrolled patients with calcifications who had available pathological results in our institution. Finally, 212 calcification lesions in 207 female patients (including 5 patients with bilateral breast calcifications) were included in our research. Patients were allocated to the primary cohort and validation cohort according to the MG inspection sequence. From May 2015 to July 2017, patients with a total of 159 lesions (mean age, 47.01 ± 9.16; range, 29–86 years) were recruited to the primary cohort. From July 2017 to March 2018, patients with a total of 53 lesions (median age, 47.96 ± 8.26; range, 28–69 years) were recruited the validation population.

All mammographies were performed in the CC & MLO views at our institution with a digital mammography system (GE Medical System Senographe DS). The baseline clinicopathological data, including age, body mass index (BMI), history of contralateral breast cancer, menopausal status, and breast density on mammography and ultrasound inspection, were obtained from hospital medical records. Specimens taken by biopsies were validated by postoperation X-rays. Confirmation of the pathological results of the calcifications was provided by the pathology department of our institution.

2.2. Image segmentation and feature extraction

CC and MLO views of the mammography images of diseased breasts were obtained. This study comprised 424 images of the 207 patients, including 5 patients with bilateral lesions. The mammography images were reviewed by two radiologists (both radiologist 1 and radiologist 2 have more than 25 years of experience in breast image analysis). Neither radiologists were aware of the pathologic outcomes. The wo radiologists extracted the region of interest (ROI) by manually free-hand declination, using the open-source itk-SNAP software (version 3.4.0, www.itksnap.org) in the CC and MLO views. The ROI included a full calcification-related area (Fig. 1).

In order to evaluate interobserver and intraobserver reproducibility of image segmentation and feature extraction, we randomly select standard CC view of mammographs of 40 cases. Radiologist 1 and 2 repeatedly read each mammograph twice under the same procedure. Intra-class correlation coefficients (ICC) were used to assess the inter- and intraobserve agreement. In related research, 0 – 0.20 considered as no agreement, 0.21 – 0.40 as fair agreement, 0.41 -0.60 as moderate agreement, 0.61 - 0.80 as substantial agreement, 0.81–1.00 as perfect agreement [23].

In the current study, the radiomic features that defined the imaging characteristics of the calcifications were captured by extracting the quantitative features. A total of 8286 2D imaging features [13] were selected from the CC and MLO views, including lesion histogram (17 features), shape (2 features), textural (84 features) and wavelet (4040 features) features, and were divided into 4 groups. The histogram features included first-order statistical features, such as the maximum, minimum, and mean values of the two-dimensional image matrix. The shape features calculated the maximum 2D diameter and surface area. Textural features included conventional texture features such as GLCM and GLSZM. The wavelet features calculated the features from the Gabor filter of the original medical image. Details of the radiomic feature extraction are described in the Supplementary material.

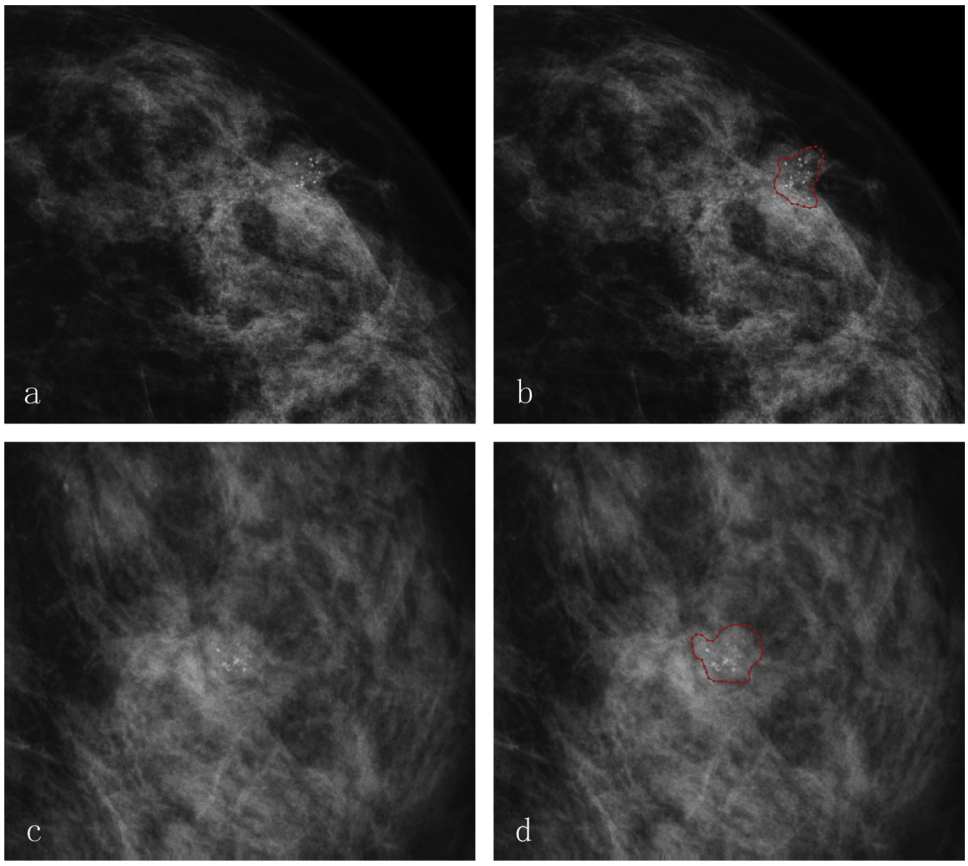

Fig. 1. Breast calcifications in mammography and segmentation of ROI with itk-SNAP software.
Mammography image in a 49-year-old woman without history of breast cancer, showing amorphous and grouped suspicious calcifications in the left outer upper quadrant breast on mammography. (a/c) Magnified craniocaudal/ medio-lateral oblique mammography; (b/d) segmentation of ROI of craniocaudal/ medio-lateral mammography image with itk-SNAP software. Both clinician 1 and clinician2 was classified this calcification as positive finding. Surgical excision was performed, and invasive ductal carcinoma (II grade) was diagnosed at histopathology, which was concordant with the diagnosis of radiomic model.

2.3. Radiomic signature building

To prevent overfitting of the model, we used the Pearson correlation coefficient to reduce the dimensions based on the training cohort [24–26]. Radiomic features most closely related to the BI-RADS category 4 calcifications were selected using the least absolute shrinkage and selection operator (LASSO) algorithm. LASSO was applied to the high-dimensional features to reduce the dimensionality and objectively select key features (Fig. A.2 in supplementary material.). The parameter λ was determined by the smallest leave-one-out cross-validation (LOOCV) error. Afterwards, the coefficients of most radiomic features were reduced to 0. Then non-zero coefficients of features built the radiomic model. The radiomic signature was a linearly weighted combination of the features of non-zero coefficients. The features extraction and selection processes were implemented in MATLAB R2016a (MathWorks, Natick, MA).

The predictive efficiency of the radiomic signatures was interpreted by the receiver operating characteristic (ROC) curve, and the performance of the signatures in differentiating benign calcifications from malignant ones in both the primary and validation cohorts was evaluated by the area under the ROC curve (AUC) and accuracy (ACC = the correct number of samples predicted / the total number of samples). The radiomic framework is shown in Fig. 3.

2.4. Development and performance of the individualized radiomic nomogram

A prediction model that combines the radiomic signature and clinical factors can usually perform better than just a single radiomic signature [27,28]. A logistic regression analysis was performed with the following clinical information: age, BMI, history of contralateral breast cancer, menopausal state, breast density and radiomic signature. To provide clinicians with a quantitative tool to predict the pathological classification of breast calcifications, we built a radiomic nomogram based on the univariable and multivariable logistic analyses of previous clinical parameters and radiomic signatures in the primary cohort. The calibration curve of the radiomic nomogram evaluated the deviation of the model's predicted value distribution. A nonsignificance test statistic implied that the model has perfect calibration [29]. We used a decision curve analysis (DCA) to show how patients can benefit from the radiomic nomogram and evaluated the clinical usefulness and robustness of the prediction model by quantifying the net benefits at different threshold probabilities.

2.5. Imaging diagnosis by the radiologists

The mammography and ultrasound data of all calcification cases are available. To further validate the performance of the radiomic nomogram, we compared the diagnostic ability of the radiomic nomogram with a visual assessment of breast screening inspections for a mammography group and a combination group with mammography and ultrasound. The same two radiologists who accomplished the image segmentation independently reviewed the segmented areas of all 424 calcification-related mammographic images (including the CC and MLO views) and made pathological predictions for each calcification. In the combination group, the two radiologists combined the mammography and ultrasound data to obtain the predicted results. The two radiologists repeated the above process within a week. ICCs were utilized to evaluate the agreement. We compared the ACC, positive predictive value (PPV), and negative predictive value (NPV) of the radiomic nomogram prediction results with those of the radiologists. All personal information was deidentified, and the radiologists were uninformed about the histopathological results.

2.6. Data and statistical analysis

Each calcification on mammography was categorized according to the 5th edition of the BI-RADS atlas. The differences in the characteristic variables (history of contralateral breast cancer, menopausal status, breast density) between the two cohorts were compared by chi-square tests or Fisher's exact tests, whereas a two-sample test was conducted to identify the differences in age and BMI. These statistical analyses were calculated with Statistical Product and Service Solutions (SPSS) (version 22). The ROC curves were implemented with "pROC" package. The "rms" package was used to construct the nomogram and to plot the calibration curve. The Hosmer-Lemeshow test was implemented with the "Resource Selection" package. The DCA was implemented with the "Decision Curve" package. The calibration curve was tested using the Hosmer-Lemeshow test for the predicted values. The differences between different AUCs were compares by DeLong's test. A p value less than 0.05 was indicated statistically significant.

3. Results

3.1. Patients

The characteristics of the patients in the primary and validation cohorts are listed in Table 1. The difference between the 2 cohorts in terms of malignant ratio was not significant (49.06 % and 52.83 % in the primary and validation cohorts, respectively, P = 0.63). The differences in the other clinical characteristics between the primary and validation cohorts in terms of age, history of contralateral breast cancer and breast density were not significant. In contrast, a few clinical characteristics differed significantly between the benign and malignant groups, including menopausal status in the primary cohort and BMI in the validation cohort.

3.2. Feature extraction, radiomic signature construction and validation

With the coarse-to-fine feature selection method, we finally extracted six features from mammography that were most strongly associated with benign and malignant calcifications to develop the radiomic signature based on the primary cohort; these included the original Gray-level size zone matrix (GLSZM) short zone emphasis (SZE) from the CC view, Gabor GLSZM_ small zone low gray-level emphasis (SZLGE) from the CC view, Gabor GLSZM SZLGE from the CC view, original GLSZM SZE from the MLO view, Gabor GLSZM zone-size variance from MLO, and Gabor GLCM cluster shade from the MLO view. The indispensable parameters for the calculations are listed in Table A.2 in the Supplementary material.

An SVM model was developed as a radiomic signature and consisted of the six selected features. The difference in radiomic scores between the benign and malignant calcifications in the primary cohort was significant (P < 0.01); a similar result was observed in the validation cohort (P < 0.01). The predictive performance of the radiomic signature is represented by the ROC curve (Fig. 2a). The radiomic signature yielded a performance with a discrimination accuracy of 75.47 % (95 % CI, 69.50 %–87.44 %) and an AUC of 0.83 (95 % CI, 0.77–0.90) in the primary cohort and a discrimination accuracy of 76.10 % (95 % CI, 63.40 %–82.80 %) and an AUC of 0.78 (95 % CI, 0.66–0.91) in the validation cohort.

3.3. Development, performance, and evaluation of the individualized radiomic nomogram

The radiomic signature and the menopausal status were identified as independent factors by the logistic regression analysis (Table A.3 in the Supplementary material). The model that combined the above predictors was constructed as the radiomic nomogram (Fig. 4a). The AUC was 0.83 (95 % CI, 0.77–0.90) in the primary cohort and 0.80 (95 % CI, 0.68-0.92) in the validation cohort based on ROC analysis (Fig. 2b). Compared with the radiomic signature, the radiomic nomogram did not significantly improve the predictive performance (P = 0.81 > 0.05).

The calibration curve demonstrated a good consistency between the diagnostic performance of the radiomic nomogram and the actual pathological results (Fig. 4b) with nonsignificant statistics (Hosmer-Lemeshow test P value = 0.90).

The decision curve analysis result of the radiomic nomogram is shown in Fig. 4**c**. The DCA demonstrated that utilizing this proposed radiomic nomogram to predict benign BI-RADS category 4 calcifications adds more benefit than treating all lesions as malignancies.

3.4. Imaging assessment by the radiologists and comparison of the prediction results between radiologists and the radiomic nomogram

Good agreement was also achieved in the imaging assessment by radiologist 1 and radiologist 2 (ICCs > 0.6). The prediction results of radiologist 1 were used for the next analysis. Imaging analysis of the mammography and combination of mammography plus ultrasound yielded a performance with an AUC of 0.61 [95 % CI, 0.55-0.68] and 0.64 [95 % CI, 0.59-0.70], respectively. The accuracies were 61.32 % (mammography) and 64.15 % (combined modality). The ACC, PPV and NPV of the visual assessment in the two groups and the radiomic nomogram are presented in Table 2. The comparison of the primary cohort and validation cohort for the visual prediction assessment between

Table 1
Characteristics of patients in the primary and validation cohorts.

Characteristic	Primary cohorts		P	Validation cohorts		P
	Benign	Malignant		Benign	Malignant	
Age, mean ± SD, years	45.93 ± 8.38	48.14 ± 9.83	0.129	49.24 ± 8.40	46.82 ± 8.11	0.292
BMI	22.16 ± 2.64	22.83 ± 3.02	0.138	21.07 ± 2.37	23.13 ± 2.88	0.007*
History of contralateral breast cancer			0.117			0.218
No	75 (92.59 %)	77 (98.71 %)		23 (92 %)	28 (100 %)	
Yes	6 (7.07 %)	1 (1.28 %)		2 (8 %)	0 (0 %)	
Menopausal status			0.002*			0.420
Premenopausal	75 (92.59 %)	58(74.36 %)		21 (84 %)	21 (75 %)	
Postmenopausal	6 (7.41 %)	20(25.64 %)		4 (16 %)	7 (25 %)	
Breast density			0.716			0.232
No-dense	12 (14.81 %)	10 (12.82 %)		8 (32 %)	5 (17.86 %)	
dense	69 (85.19 %)	68 (87.18 %)		17 (68 %)	23 (82.14 %)	

BMI body mass index; The differences in characteristic variables (History of breast cancer, Menopausal status, Density of mammography) between the two cohorts were compared by Chi-square tests or Fisher exact tests, whereas two-sample t test was conducted to identify differences in age.
* p < 0.05.

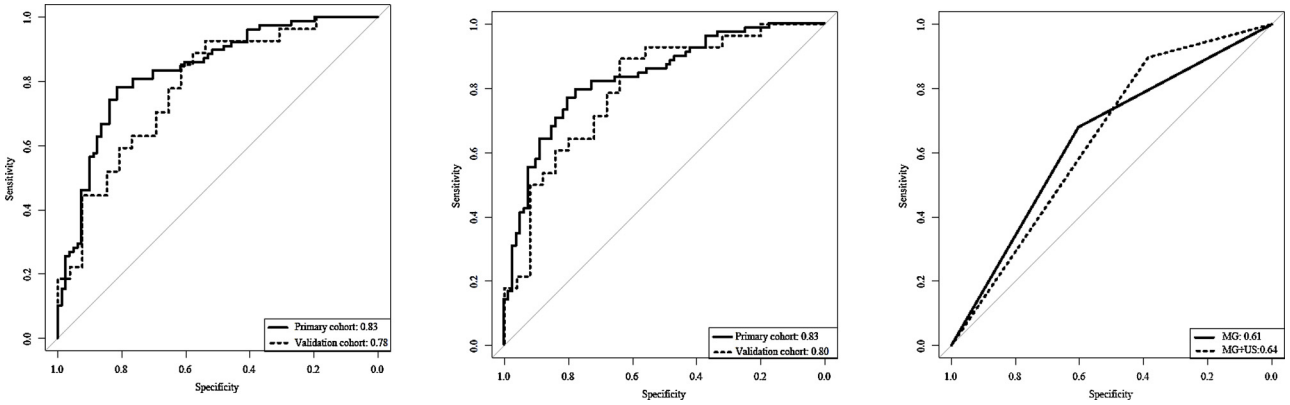

Fig. 2. The ROC curves of radiomic signature, radiomic nomogram and radiologists visual assessment.
(a) The ROC curves for differentiation of benign calcifications and malignant ones using the radiomic signature. The areas under ROC curve (AUC) are 0.83 in the primary cohort (solid line) and 0.78 in the validation cohort (dotted line). (b) The ROC curves of the radiomic signature. The areas under ROC curve (AUC) are 0.83 in the primary cohort (solid line) and 0.80 in the validation cohort (dotted line). (c) The ROC curves of the radiologist's empirical prediction with all data. The areas under ROC curve (AUC) are 0.61 in the mammography (MG) group (solid line) and 0.64 in the MG plus ultrasound (US) group (dotted line).

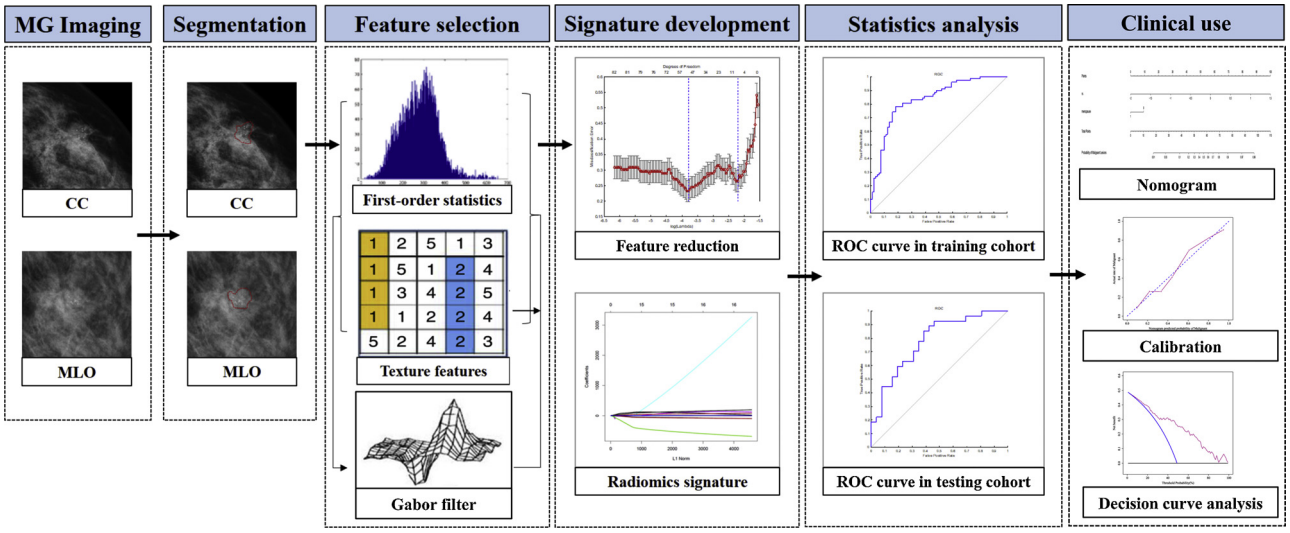

Fig. 3. The radiomics flowchart.
Segmentation is performed on mammography images to define the calcifications region. From this region, the radiomic features are extracted, including first-order statistics and texture features with and without Gabor filter. Using the Pearson correlation coefficient and LASSO method, six features were selected to build the radiomic signature. Finally, the radiomic signature and clinical factors were incorporated into a nomogram for individual evaluation.

the two groups is shown in Table A.4 in the Supplementary material.

DeLong's test suggested that the AUC value in the radiomic nomogram was higher than that of both the mammography and combination of mammography plus ultrasound in the radiologist assessments ($P < 0.05$). Additionally, based on the chi-square test, the accuracy of the radiomic nomogram was higher than that of both the mammography and combination of mammography plus ultrasound in the radiologist assessments ($P < 0.05$). These results show that the predicted power of the radiomic nomogram is better than that of the conventional visual prediction assessments for both mammography and the combination of mammography plus ultrasound.

With the combination of mammography plus ultrasound, we found that only 166 lesions (78.30 %) with visible calcifications on mammography could be visualized on ultrasound. Among the mammography-detected ultrasound-negative calcifications, 34 calcification lesions (73.91 %) were pathologically benign. The chi-square test suggested that the difference between the negative predictive value of the radiomic nomogram and that of the radiologist assessment of mammography was significant (90 % vs 82.61 %, $p = 0.003 < 0.05$).

4. Discussion

In the current analysis, we employed mammographic images to investigate the differences in breast calcification characteristics among different pathological types. A mammography-based radiomic nomogram was proposed with a better ability to predict benign BI-RADS category 4 calcifications than the radiologist empirical prediction model.

Mammography detects calcifications with high sensitivity, but the specificity for discriminating benign calcifications from malignant calcifications is low [12]. Rosenberg RD and Van Luijt PA reported that 70–80 % of suspicious mammographic findings were confirmed to result in false-positive biopsies [30,31]. Overtreatment leads to risk and complications such as hematoma, infections and increased medical expenses [32,33]. The advantage of breast ultrasound is the discovery of mass rather than calcification. Although Magnetic resonance imaging (MRI) is considered a method of breast examination with a high sensitivity, there is no consensus that MRI is superior to mammography in evaluating breast calcifications and MRI is not recommended as a routine examination for breast calcifications [10]. Therefore, we intended to use a new method to identify the benign diseases in BI-RADS

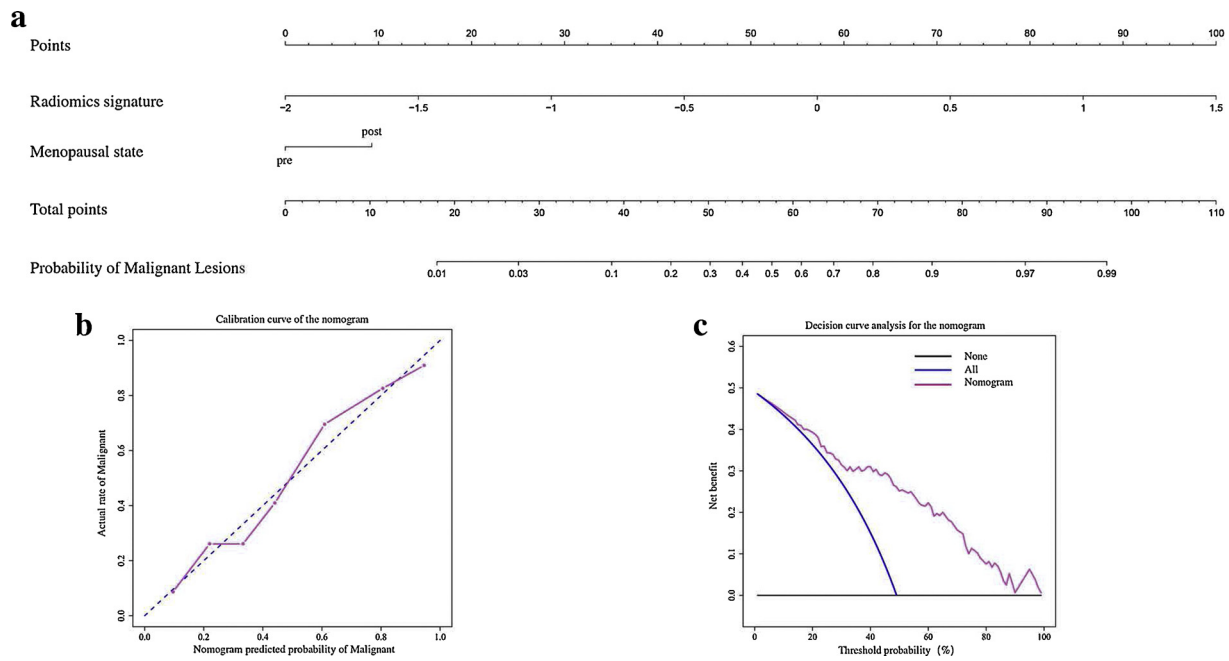

Fig. 4. Developed radiomic nomogram, calibration curves and the decision curve analysis derived from the radiomic nomogram.
(a) The radiomic nomogram was developed in the primary cohort, with the radiomic signature, and menopausal status incorporated. (b) Calibration curves of the radiomic nomogram. Calibration curves depict the calibration of each model in terms of the agreement between the predicted probability of malignancy and actual outcomes. The y axis represents the actual rate of malignancy. The x axis represents the predicted probability of malignancy. The diagonal blue line represents a perfect prediction by an ideal model. The **pink** line represents the performance of the radiomic nomogram, of which a closer fit to the diagonal blue line represents a better prediction. (c) Decision curve analysis for the radiomic nomogram. The y axis measures the net benefit. The **pink** line represents the radiomic nomogram. The blue line represents the assumption that all lesions considered malignancy. The **black** line represents the assumption that no lesions considered malignancy.

Table 2
Comparison of results of visual prediction assessment and radiomic nomogram.

Metrics	Radiomic nomogram		Visual prediction assessment	
	Primary cohort	Validation cohort	MG	MG + US
ACC	76.10	77.36	61.32	64.15
(95 %CI)	(69.40-82.80	(65.71-89.00)	(54.71-67.93)	(57.64-70.66)
AUC	0.83	0.80	0.61	0.64
(95 %CI)	(0.77-0.90)	(0.68-0.99)	(0.55-0.68)	(0.59-0.70)
PPV	78.57	73.53	60.53	59.38
(95 %CI)	(68.72-88.43)	(57.90-89.15)	(51.42-69.64)	(51.68-67.07)
NPV	74.16	84.21	62.24	78.85
(95 %CI)	(64.88-83.43)	(66.15-100)	(52.48-72.01)	(67.37-90.33)

Abbreviation: AUC area under the receiver-operating characteristic curve; ACC accuracy; PPV positive predictive value; NPV negative predictive value; MG mammography; US ultrasound.

category 4 calcifications with mammography.

We extracted features from easily obtained mammographic data to constructe a radiomic signature in the present study. The results showed that the radiomic signature had a definite predictive value with an AUC of 0.78 in the validation cohort. Previously, clinical factors were shown to be associated with different pathological classifications in breast calcifications [34]. We selected the menopausal state by multivariate logistic regression analysis and then combined this feature with the radiomic signature to develop a radiomic nomogram. However, compared with that of the radiomic signature, the AUC value was not significantly different for the radiomic nomogram. We believe that this is probably because there was only one clinical factor included in the radiomic nomogram. The calibration curves indicated a good consistency between the radiomic nomogram predictive results and the actual outcomes, and the clinical utility of the radiomic nomogram was demonstrated by the decision curve analysis. In addition, the AUC values of the primary cohort and the validation cohort of our radiomic nomogram are almost equivalent, indicating that the nomogram has good stability.

The predictive ability of the radiomic nomogram was also compared with that of the visual assessment by radiologists. The AUC and ACC of the radiomic nomogram were significantly higher than those of the radiologist prediction model ($P < 0.05$). A better performance was achieved in the machine learning classifier, which means that the high-dimensional radiomic features explored more detailed information about the breast calcifications than the naked eye [35,36]. In short, the above results suggest that the radiomic nomogram was helpful in differentiating benign calcifications from malignant calcifications.

Currently, ultrasound and mammography are major screening tools for breast lesions. Since mammography is more sensitive to screening for calcifications, after the inclusion of ultrasound data, we found that 46 cases (21.70 %) of BI-RADS category 4 mammographic calcifications could not be visualized with ultrasound. According to the previous literature, breast calcifications that are negative on ultrasound but can be detected by mammography (MG+/US-) are usually nonmalignant calcifications, with approximately 65–90 % of the lesions being benign [11,37]. The benign ratio of MG+/US- calcifications in our study was 73.91 % (34/46), which was consistent with the ratio reported in other studies and significantly higher than the benign ratio of all calcification data in this paper (50 %, $p < 0.05$). In the data of MG+/US- calcifications, we also found that the negative predictive value of the radiomics nomogram was much higher than that of mammography (90 % vs 82.61 %, $P < 0.05$). It has been indicated that, especially for this type of calcification, radiomics nomograms have the ability to screen for benign calcifications.

There are still several limitations in this study. First, the imaging data of our research were retrospectively selected from a single institution with a limited number of patients, which potentially limits the generalizability of our radiomics nomogram to be applied in other studies discriminating BI-RADS category 4 calcifications for patients throughout multiple centers. Second, the diagnostic ability of our

radiomics model has been validated in the present research, but further validation is necessary in a large and prospective patient cohort before the use of the radiomic model is generalized. Third, for MG + /US-calcifications, more data are needed to verify the diagnostic performance of the radiomic nomogram.

Our study has unique strengths. First, the present study only focuses on BI-RADS category 4 calcifications. The data in the majority of previous studies include calcifications of all BI-RADS categories [38,39]. Calcifications of categories 1–3 and category 5 are unnecessary to analyze since they have obvious benign and malignant mammographic imaging features, respectively, and thus are nearly meaningless to explore. This study aimed to investigate the most difficult and suspicious category. Second, our study concentrates on the calcification-related ROIs of mammographic images instead of full-view abnormal imaging changes, which has been the focus of most previous studies. Third, we applied a radiomic nomogram, a new approach, to solve a previously unresolved clinical problem and compared the nomogram with a frequently used empirical assessment. Lastly, most of the previous literature only focused on differences in imaging features; in contrast, our research considered both imaging features and clinical risk factors. This noninvasive model has been proven to be discriminative and can be used to support clinical decisions.

In conclusion, results of the present study showed that radiomic nomogram could be served as a potential tool in predicting the benign diseases of BI-RADS category 4 calcifications, particularly in patients with MG + /US- breast calcifications. Radiomic nomogram can identify the possible benign calcifications according to the relevant characteristics of calcifications on mammographic images. Radiomics is an emerging tool for disease diagnosis, and more studies are expected to be conducted to prove its unique value in clinical applications.

Funding sources

This study was funded by National Key Research and Development Program of China (2017YFC1309100, 2017YFA0205200); Natural Science Foundation of Guangdong Province, China (2017A030313882); National Natural Science Foundation of China (81871513, 81772012, 81922040); the Beijing Natural Science Foundation (7182109); Youth Innovation Promotion Association CAS (2019136) and CSC0-constant Rui Tumor Research Fund, China (Y-HR2016-067).

The author(s) indicated no potential conflicts of interest.

Declaration of Competing Interest

The author(s) indicated no potential conflicts of interest.

Acknowledgements

We thank all of the patients for their participation and all of their physicians and radiologists for their hard work.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2019.108711.

References

[1] C. Gajdos, P.I. Tartter, I.J. Bleiweiss, et al., Mammographic appearance of non-palpable breast cancer reflects pathologic characteristics, Ann. Surg. 235 (2002) 246–251.

[2] A.M. Bluekens, R. Holland, N. Karssemeijer, M.J. Broeders, G.J. den Heeten, Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study, Radiology 265 (2012) 707–714, https://doi.org/10.1148/radiol.12111461.

[3] L. Tabar, C.J. Fagerberg, A. Gad, et al., Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare, Lancet (London, England) 1 (1985) 829–832.

[4] G.M. Tse, P.H. Tan, H.S. Cheung, W.C. Chu, W.W. Lam, Intermediate to highly suspicious calcification in breast lesions: a radio-pathologic correlation, Breast Cancer Res. Treat. 110 (2008) 1–7, https://doi.org/10.1007/s10549-007-9695-4.

[5] C.J. D'Orsi, E.A. Sickles, E.B. Mendelson, ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System, 5th ed., American College of Radiology, Reston, 2013.

[6] M.P. Goetz, W.J. Gradishar, B.O. Anderson, et al., NCCN guidelines insights: breast cancer, version 3.2018, J. Natl. Compr. Cancer Netw. 17 (2019) 118–126, https://doi.org/10.6004/jnccn.2019.0009.

[7] E.S. Burnside, J.E. Ochsner, K.J. Fowler, et al., Use of microcalcification descriptors in BI-RADS 4th edition to stratify risk of malignancy, Radiology 242 (2007) 388–395, https://doi.org/10.1148/radiol.2422052130.

[8] C.K. Bent, L.W. Bassett, C.J. D'Orsi, J.W. Sayre, The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories, Am. J. Roentgenol. 194 (2010) 1378–1383, https://doi.org/10.2214/ajr.09.3423.

[9] U. Kettritz, G. Morack, T. Decker, Stereotactic vacuum-assisted breast biopsies in 500 women with microcalcifications: radiological and pathological correlations, Eur. J. Radiol. 55 (2005) 270–276, https://doi.org/10.1016/j.ejrad.2004.10.014.

[10] B. -Bennani-Baiti, P.A. Baltzer, MR imaging for diagnosis of malignancy in mammographic microcalcifications: a systematic review and meta-analysis, Radiology 283 (2017) 692–701, https://doi.org/10.1148/radiol.2016161106.

[11] S. -Aminololama-Shakeri, C.K. Abbey, P. Gazi, et al., Differentiation of ductal carcinoma in-situ from benign micro-calcifications by dedicated breast computed tomography, Eur. J. Radiol. 85 (2016) 297–303, https://doi.org/10.1016/j.ejrad.2015.09.020.

[12] M.H. -Park, W. Lim, D. Jo, et al., Rapid differential diagnosis of breast microcalcification using targeted near-infrared fluorophores, Adv. Healthc. Mater. 7 (2018) e1701062, , https://doi.org/10.1002/adhm.201701062.

[13] H.J.W.L. Aerts, E.R. Velazquez, R.T.H. Leijenaar, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat. Commun. 5 (2014), https://doi.org/10.1038/ncomms5006.

[14] Z. Liu, Y. Wang, X. Liu, et al., Radiomics analysis allows for precise prediction of epilepsy in patients with low-grade gliomas, Neuroimage Clin. 19 (2018) 271–278, https://doi.org/10.1016/j.nicl.2018.04.024.

[15] P. Lambin, E. Rios-Velazquez, R. Leijenaar, et al., Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer 48 (2012) 441–446, https://doi.org/10.1016/j.ejca.2011.11.036.

[16] L.J. Grimm, Breast MRI radiogenomics: current status and research implications, J. Magn. Resonance Imaging 43 (2016) 1269–1278, https://doi.org/10.1002/jmri.25116.

[17] W. Wei, Z. Liu, Y. Rong, et al., A computed tomography-based radiomic prognostic marker of advanced high-grade serous ovarian cancer recurrence: a multicenter study, Front. Oncol. 9 (2019), https://doi.org/10.3389/fonc.2019.00255.

[18] Z. Liu, X.Y. Zhang, Y.J. Shi, et al., Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer, Clin. Cancer Res. 23 (2017) 7253–7262, https://doi.org/10.1158/1078-0432.Ccr-17-1038.

[19] H.J.W.L. Aerts, The potential of radiomic-based phenotyping in precisionmedicine a review, JAMA Oncol. 2 (2016) 1636–1642, https://doi.org/10.1001/jamaoncol.2016.2631.

[20] Y.Q. Huang, C.H. Liang, L. He, et al., Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer, J. Clin. Oncol. 34 (2016), https://doi.org/10.1200/Jco.2015.65.9128 2157-+..

[21] P. Lambin, R.T.H. Leijenaar, T.M. Deist, et al., Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (2017) 749–762, https://doi.org/10.1038/nrclinonc.2017.141.

[22] Z.Y. Liu, X.Y. Zhang, Y.J. Shi, et al., Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal Cancer, Clin. Cancer Res. 23 (2017) 7253–7262, https://doi.org/10.1158/1078-0432.CCR-17-1038.

[23] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159–174.

[24] R. Tibshirani, J. Bien, J. Friedman, et al., Strong rules for discarding predictors in lasso-type problems, J. R. Stat. Soc. Ser. B 74 (2012) 245–266, https://doi.org/10.1111/j.1467-9868.2011.01004.x.

[25] N. Simon, J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for cox's proportional hazards model via coordinate descent, J. Stat. Softw. 39 (2011) 1–13.

[26] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (2010) 1–22.

[27] Y.Q. Huang, C.H. Liang, L. He, et al., Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer, J. Clin. Oncol. 34 (2016) 2157–2164, https://doi.org/10.1200/jco.2015.65.9128.

[28] N. Dinapoli, B. Barbaro, R. Gatta, et al., Magnetic resonance, vendor-independent, intensity histogram analysis predicting pathologic complete response after radiochemotherapy of rectal cancer, Int. J. Radiat. Oncol. Biol. Phys. (2018), https://doi.org/10.1016/j.ijrobp.2018.04.065.

[29] M.P. van der Paardt, M.B. Zagers, R.G. Beets-Tan, J. Stoker, S. Bipat, Patients who undergo preoperative chemoradiotherapy for locally advanced rectal cancer restaged by using diagnostic MR imaging: a systematic review and meta-analysis, Radiology 269 (2013) 101–112, https://doi.org/10.1148/radiol.13122833.

[30] R.D. Rosenberg, B.C. Yankaskas, L.A. Abraham, et al., Performance benchmarks for screening mammography, Radiology 241 (2006) 55–66, https://doi.org/10.1148/radiol.2411051504.

[31] P.A. van Luijt, J. Fracheboud, E.A. Heijnsdijk, G.J. den Heeten, H.J. de Koning,

Nation-wide data on screening performance during the transition to digital mammography: observations in 6 million screens, Eur. J. Cancer (Oxf., Eng.: 1990) 49 (2013) 3517–3525, https://doi.org/10.1016/j.ejca.2013.06.020.
[32] E. Morris, S.A. Feig, M. Drexler, C. Lehman, Implications of overdiagnosis: impact on screening mammography practices, Popul. Health Manag. 18 (Suppl 1) (2015) S3–11, https://doi.org/10.1089/pop.2015.29023.mor.
[33] S. Zackrisson, I. Andersson, L. Janzon, J. Manjer, J.P. Garne, Rate of over-diagnosis of breast cancer 15 years after end of Malmo mammographic screening trial: follow-up study, BMJ (Clinical research ed) 332 (2006) 689–692, https://doi.org/10.1136/bmj.38764.572569.7C.
[34] T. Li, L. Tang, Z. Gandomkar, et al., Mammographic density and other risk factors for breast cancer among women in China, Breast J. 24 (2018) 426–428, https://doi.org/10.1111/tbj.12967.
[35] P. Lambin, R.T.H. Leijenaar, T.M. Deist, et al., Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (2017) 749–762, https://doi.org/10.1038/nrclinonc.2017.141.
[36] H.J. Aerts, The potential of radiomic-based phenotyping in precision medicine: a review, JAMA Oncol. 2 (2016) 1636–1642, https://doi.org/10.1001/jamaoncol.2016.2631.
[37] V.S. Pankratz, A.C. Degnim, R.D. Frank, et al., Model for individualized prediction of breast cancer risk after a benign breast biopsy, J. Clin. Oncol. 33 (2015) 923–929, https://doi.org/10.1200/jco.2014.55.4865.
[38] A. Papadopoulos, D.I. Fotiadis, A. Likas, Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines, Artif. Intell. Med. 34 (2005) 141–150, https://doi.org/10.1016/j.artmed.2004.10.001.
[39] L. Wei, Y. Yang, R.M. Nishikawa, Y. Jiang, A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications, IEEE Trans. Med. Imaging 24 (2005) 371–380.