European Radiology (2022) 32:4090–4100
https://doi.org/10.1007/s00330-021-08510-8

BREAST

MRI-based radiomics analysis for differentiating phyllodes tumors of the breast from fibroadenomas

Mitsuteru Tsuchiya[1] · Takayuki Masui[1] · Kazuma Terauchi[1] · Takahiro Yamada[1] · Motoyuki Katyayama[1] · Shintaro Ichikawa[2] · Yoshifumi Noda[3] · Satoshi Goshima[2]

Received: 12 June 2021 / Revised: 27 November 2021 / Accepted: 6 December 2021 / Published online: 19 January 2022

Abstract

Objectives To evaluate the diagnostic performance of MRI-based radiomics model for differentiating phyllodes tumors of the breast from fibroadenomas.

Methods This retrospective study included 88 patients (32 with phyllodes tumors and 56 with fibroadenomas) who underwent MRI. Radiomic features were extracted from T2-weighted image, pre-contrast T1-weighted image, and the first-phase and late-phase dynamic contrast-enhanced MRIs. To create stable machine learning models and balanced classes, data augmentation was performed. A least absolute shrinkage and selection operator (LASSO) regression was performed to select features and build the radiomics model. A radiological model was constructed from conventional MRI features evaluated by radiologists. A combined model was constructed using both radiomics features and radiological features. Machine learning classifications were done using support vector machine, extreme gradient boosting, and random forest. The area under the receiver operating characteristic (ROC) curve (AUC) was computed to assess the performance of each model.

Results Among 1070 features, the LASSO logistic regression selected 35 features. Among three machine learning classifiers, support vector machine had the best performance. Compared to the radiological model (AUC: 0.77 ± 0.11), the radiomics model (AUC: 0.96 ± 0.04) and combined model (0.97 ± 0.03) had significantly improved AUC values (both $p < 0.01$) in the validation set. The combined model had a relatively higher AUC than that of the radiomics model in the validation set, but this was not significantly different ($p = 0.391$).

Conclusions Radiomics analysis based on MRI showed promise for discriminating phyllodes tumors from fibroadenomas.

Key Points

- *The radiomics model and the combined model were superior to the radiological model for differentiating phyllodes tumors from fibroadenomas.*
- *The SVM classifier performed best in the current study.*
- *MRI-based radiomics model could help accurately differentiate phyllodes tumors from fibroadenomas.*

Keywords Magnetic resonance imaging · Machine learning · Breast · Phyllodes tumor · Fibroadenoma

Abbreviations

ACC	Accuracy
AUC	Area under the curve
CI	Confidence interval
DCE	Dynamic contrast enhanced
FOV	Field of view
Gd	Gadolinium
ICC	Interobserver correlation coefficient
LASSO	Least absolute shrinkage and selection operator
ML	Machine learning
MRI	Magnetic resonance imaging
RF	Random forest
ROC	Receiver operating characteristic

✉ Mitsuteru Tsuchiya
tsuchi8@hama-med.ac.jp

[1] Department of Radiology, Seirei Hamamatsu General Hospital, 2-12-12, Sumiyoshi, Naka-ku, Hamamatsu city, Shizuoka 430-8558, Japan

[2] Department of Radiology, Hamamatsu University School of Medicine, 1-20-1, Handayama, Higashi-ku, Hamamatsu City, Shizuoka 431-3192, Japan

[3] Department of Radiology, Gifu University, 1-1, Yanagido, Gifu City, Gifu 501-1194, Japan

ROI	Region of interest
SVM	Support vector machine
T1WI	T1-weighted imaging
T2WI	T2-weighted imaging
TE	Echo time
TIC	Time-intensity curve
TR	Repetition time
XGB	Extreme gradient boosting

Introduction

Phyllodes tumor is a rare neoplasm composed of epithelial and stromal components, with a reported incidence of less than 1.0% of all female breast tumors [1–3]. Whereas all fibroadenomas are considered benign, phyllodes tumors can be subclassified histologically as benign, borderline, or malignant [4]. Sonography cannot be used to reliably differentiate phyllodes tumors from fibroadenomas [5]. Core needle biopsy often fails to distinguish them [6]. Nevertheless, a preoperative diagnosis is crucial because they require different surgical procedures. Fibroadenomas require only enucleation. In the treatment of phyllodes tumors, complete surgical excision with wide margins of at least 1 cm is required for all grades of phyllodes tumor owing to the high recurrence rate [7, 8].

Phyllodes tumors and fibroadenomas cannot be precisely differentiated on breast MRI [9], but the presence of a cystic component, strong lobulation, heterogeneity on contrast-enhanced T1-weighted images may be helpful for differentiation [10].

Radiomics analysis is an emerging translational field of research aiming to find associations between quantitative information extracted from clinical images and other clinical data. Radiomics analysis refers to various mathematical methods that allow the evaluation of the gray-level intensity and position of the pixels on medical images [11]. It was developed to detect subtle changes that may not be visible to the radiologist's eye and is believed to reflect tissue microstructure organization [12]. The heterogeneity of a tumor is a feature characterized by areas of hemorrhage, cystic changes, high cell density, necrosis, and myxoid changes [13, 14]. Radiomics analysis of breast MRI has been applied to help differentiate benign from malignant lesions [15, 16], preoperative prediction of sentinel lymph node metastasis [17], treatment response to neoadjuvant chemotherapy [18], recurrence-free survival [19], and breast cancer subtypes [20, 21].

Few studies have investigated the value of quantitative MRI texture features for differentiating phyllodes tumors from fibroadenomas to show their usefulness [22, 23]. Recently, different machine learning classifiers have been compared to determine an optimal machine learning method [24]. Thus, in our study, we investigate three machine-learning classifiers to assess the capability of the radiomics model.

The aim of this study was to develop MRI-based radiomics machine learning model for differentiating phyllodes tumors from fibroadenomas.

Materials and methods

This retrospective study was approved by our institutional ethics committee, which waived the need for informed consent.

Patients

The enrolled patients had histologically confirmed phyllodes tumor or fibroadenoma. Patients were consecutively identified by searching the pathology database from our institution for the period January 2004 to February 2020. Eighty-eight patients with a diagnosis of phyllodes tumor ($n = 32$; mean age, 47.6 ± 10.8 years) or fibroadenoma ($n = 56$; mean age, 40.2 ± 12.5 years) were enrolled. The inclusion criteria were as follows: (i) patients with pathologically confirmed phyllodes tumors or fibroadenomas after surgical operation or core needle biopsy; (ii) patients who had undergone breast dynamic contrast-enhanced (DCE) MRI prior to any surgical operation. The exclusion criteria were (i) patients with a maximum tumor diameter less than 10.0 mm; (ii) patients with an MRI performed outside our institution; (iii) poor image quality such as significant motion artifact. The exclusion criteria for a maximum tumor diameter at 10.0 mm was set to minimize the influence of partial volume effects, which might distort the true tissue-specific image texture [25].

Data augmentation

Data augmentation has been proven a powerful method for avoiding overfitting when there is a small amount of data [26–29]. The small number of patients might lead to potential overfitting, and thus, we naturally augmented the labeled data in our study by obtaining samples from different levels of the tumors [30]. Also, after we considered imbalanced data and the potential consequences of such data with regard to the machine learning schemes [31], the phyllodes tumors were segmented with more samples [30]. On average, four to five levels of data were obtained from the phyllodes tumors, and two to three levels were obtained from the fibroadenomas. The augmentation resulted in 300 labeled segmentation data (150 phyllodes tumors and 150 fibroadenomas) from 88 breast tumors (32 phyllodes tumors and 56 fibroadenomas). The 300 labeled segmentation data were divided into 200 training sets (100 phyllodes tumors and 100 fibroadenomas) and 100 validation sets (50 phyllodes tumors and 50 fibroadenomas).

The workflow of this study is shown in Fig. 1.

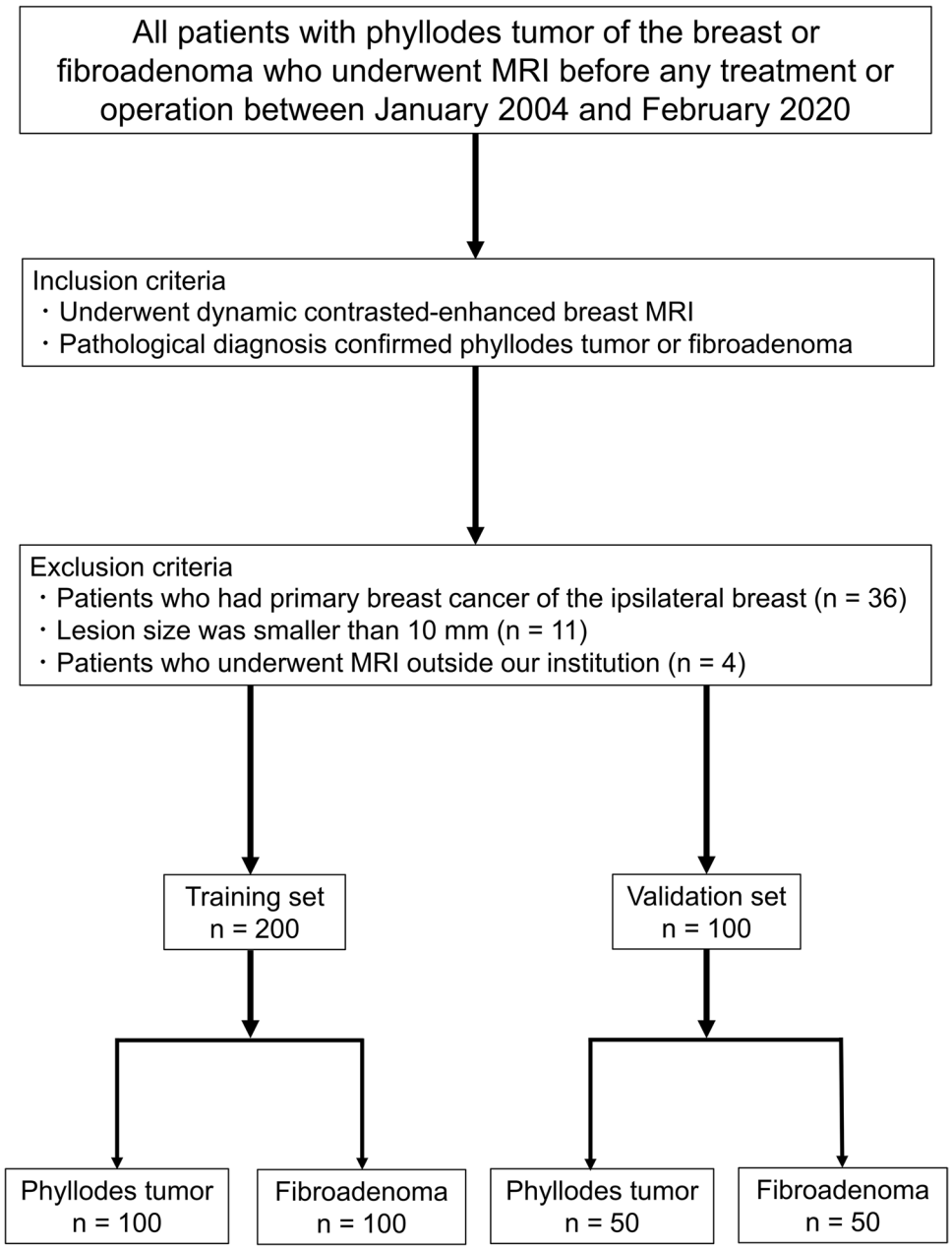

Fig. 1 Workflow of this study

MRI data acquisition

All examinations were performed with the patient in a prone position. The MR images were acquired using three MRI scanners: 60 patients with a 1.5-T system (Signa 1.5 T, GE Healthcare) with a dedicated eight-channel breast phased-array coil, 20 patients with a 3.0-T system (Discovery MR750, GE Healthcare) with a dedicated eight-channel breast phased-array coil, and 8 patients with a 3.0-T system (SIGNA Pioneer, GE Healthcare) with a dedicated eight-channel breast phased-array coil.

DCE-MR images were acquired using a 3D fat-suppressed T1-weighted volume imaging breast assessment (VIBRANT) sequence composed of one pre-contrast and three post-contrast phases labeled as pre-contrast enhanced, post-CE1, post-CE2, and post-CE3. The gadolinium-based contrast agent was administered at a concentration of 0.1 mmol gadobutrol per kg body weight (Gadavist, Bayer Healthcare Pharmaceuticals) at a rate of 2 mL/s, followed by a 20-mL saline flush at the same rate. The acquisition conditions were as follows: VIBRANT, repetition time (TR) = 7.9 ms, echo time (TE) = 4.3 ms, flip angle = 12°, field of view (FOV) = 34 cm, acquired matrix = 300 × 300, in-plane spatial resolution = 1.1 × 1.1 mm, thickness = 1.1 mm, temporal resolution = ~ 120 s, axial orientation.

In addition, axial two-dimensional fat-saturated T2-weighted fast spin-echo or iterative decomposition of

water and fat with echo asymmetry and least-squares estimation (IDEAL) images were performed under the following conditions: TR = 6680 ms, TE = 68 ms, slice thickness = 3.0 mm, matrix = 320 × 192, FOV = 36 cm.

Reference standard, region of interest segmentation, and radiomics feature extraction

Radiomic feature extraction was performed using the MaZda software (version 4.6, Technical University of Lodz) [32]. For each sequence, six feature categories, histogram, co-occurrence matrix, run length matrix, absolute gradient, autoregressive model, and wavelet transform were extracted using MaZda, as shown in Table 1. The details of the reference standard, region of interest (ROI) segmentation, radiomics feature extraction, and reproducibility analysis are described in Supplemental Material.

In this study, we proposed and developed a study scheme by using a machine learning classifier. Figure 2 shows the flowchart of the proposed scheme.

Visual assessment by radiologists

Two radiologists (radiologist 1 with 31 years and radiologist 2 with 16 years of experience in breast MRI) independently reviewed MRI features. The interpretation of MRI features was based on the following characteristics as per the American College of Radiology Breast Imaging Reporting and Data System MR imaging criteria (version 5) [22, 33, 34]. The details of the image analysis are described in Supplemental Material.

Figure 3 shows representative cases of phyllodes tumor and fibroadenoma on T2WI, pre-T1WI, the first phase of DCE-MRI, and the late phase of DCE-MRI.

ComBat harmonization

Radiomics feature values are affected by the different magnetic fields, protocols, and technical settings of the MR scanners. In this study, all radiomics features extracted from images acquired from different MR scanners were harmonized to remove the scanner effect using a ComBat harmonization procedure [35, 36].

The least absolute shrinkage and selection operator logistic regression

Feature selection was required to reduce overfitting, redundancy, or any other type of bias in our radiomics analysis. The least absolute shrinkage and selection operator (LASSO) algorithm was used in the training set for dimensionality reduction and feature selection by performing variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model produced [37, 38]. The 1- standard error of the minimum criteria (the 1-SE criteria) was used to tune the regularization parameter (λ) and for feature selection using tenfold cross-validation. The LASSO analysis was performed using the "glmnet" package

Table 1 Radiomics features extracted by MaZda

Feature category	Radiomics features
Histogram[a]	Mean, variance, skewness, kurtosis, percentiles (1%, 10%, 50%, 90%, and 99%)
Co-occurrence matrix[b] (computed for four directions [(a,0), (0,a), (a,a), (0,-a)] at five interpixel distances (a = 1–5)	Angular second moment, contrast, correlation, entropy, sum entropy, sum of squares, sum average, sum variance, inverse different moment, difference entropy, difference variance
Run-length matrix[c] (computed for four angles [vertical, horizontal, 0°, and 135°])	Run-length nonuniformity, gray-level nonuniformity, long-run emphasis, short-run emphasis, fraction of image in runs
Absolute gradient[d]	Gradient mean, variance, skewness, kurtosis, and nonzeros
Autoregressive model[e]	Teta 1 to 4, sigma
Wavelet transform[f] (calculated for four subsampling factors [n = 1–4])	Energy of wavelet coefficients in low-frequency subbands, horizontal high-frequency subbands, vertical high-frequency subbands, and diagonal high-frequency subbands

[a] Total number of histogram-based features: 9

[b] Features are computed for five between-pixel distances (1, 2, 3, 4, 5) and 4 directions (2D images). Total number of co-occurrence matrix-based features: 220

[c] Features are computed for 4 (2D images) directions. Total number of run-length matrix-based features: 20

[d] Total number of gradient-based features: 5

[e] Total number of autoregressive model–based features: 5

[f] Features were computed at five scales within four frequency bands (L, low; H, high): LL, LH, HL, and HH. Total number of Haar wavelet-based features: 20

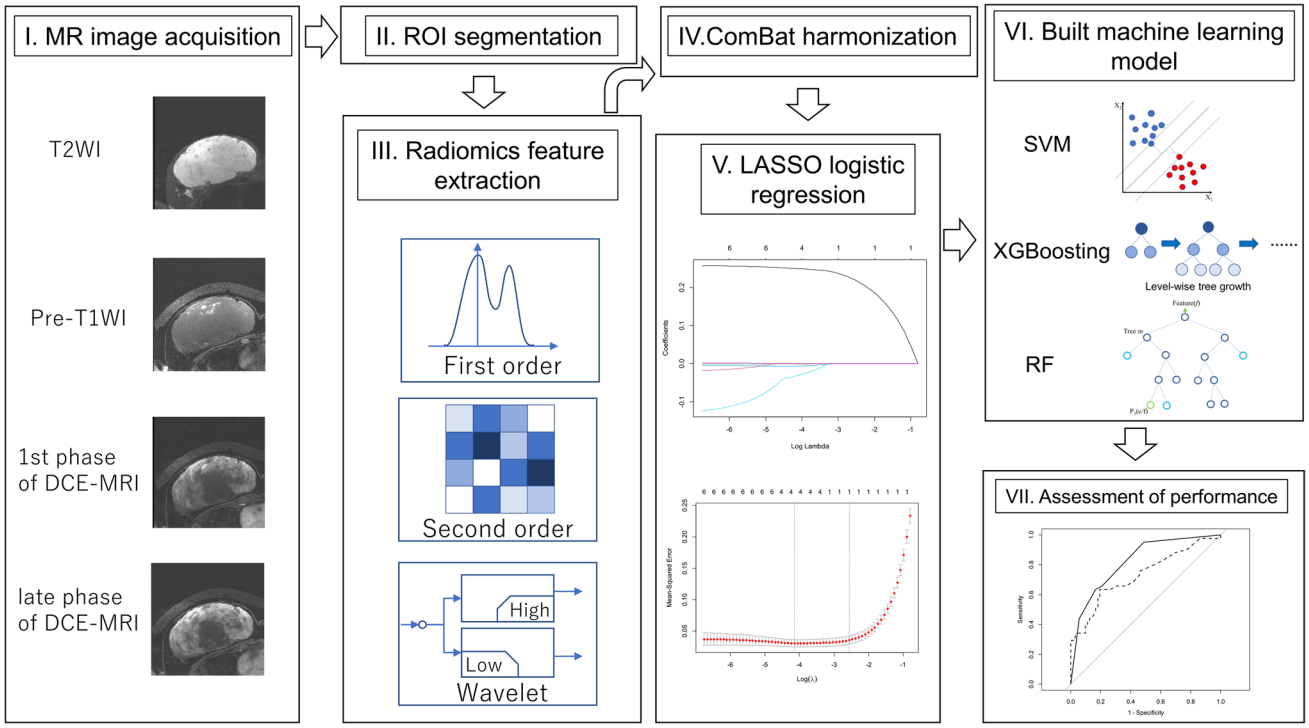

Fig. 2 Flowchart of the radiomics analysis

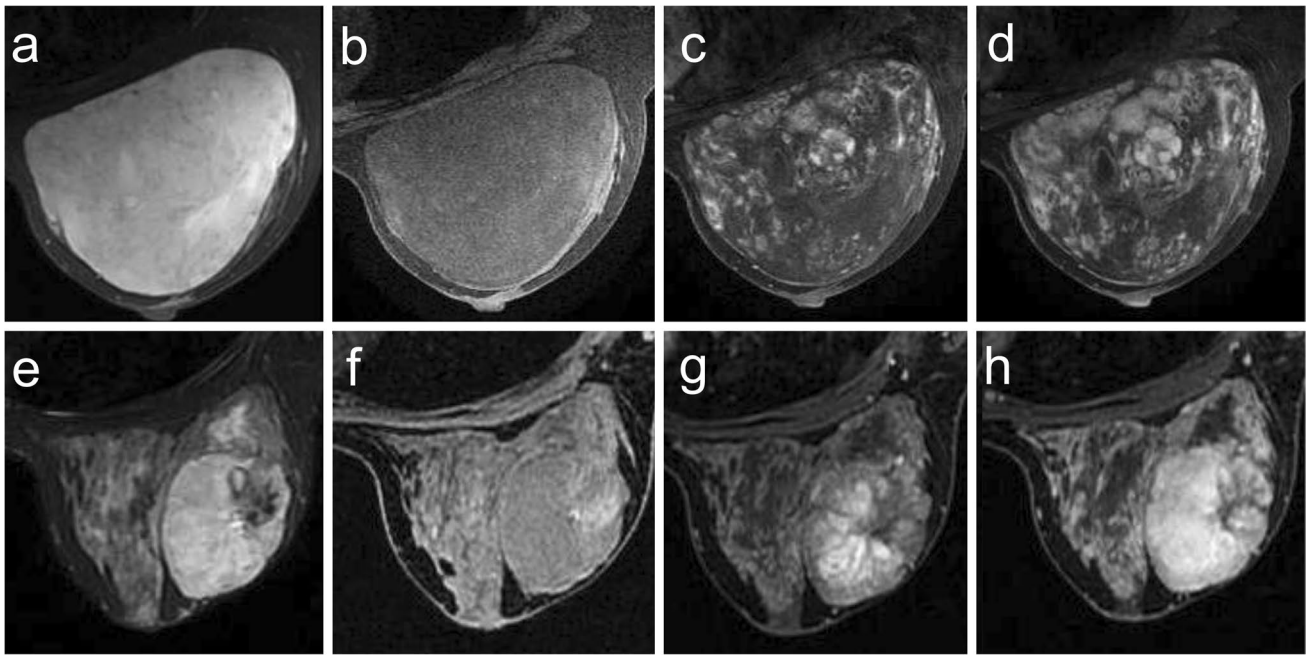

Fig. 3 Magnetic resonance images of phyllodes tumor and fibroadenoma. A 37-year-old woman with a pathologically confirmed borderline malignant phyllodes tumor. **a** Axial T2WI, (**b**) pre-T1WI, (**c**) first phase of DCE-MRI, and (**d**) late phase of DCE-MRI. A 25-year-old woman with a pathologically confirmed fibroadenoma. **e** Axial T2WI, (**f**) pre-T1WI, (**g**) first phase of DCE-MRI, and (**h**) late phase of DCE-MRI. DCE, dynamic contrast-enhanced; MRI, magnetic resonance imaging; T1WI, T1-weighted imaging; T2WI, T2-weighted imaging

in the R software (R Foundation for Statistical Computing) [39]. The features with non-zero coefficients were selected from the candidate features and formed a radiomic signature for machine-learning classification analysis.

Machine learning classifier

In order to achieve a high and robust performance of classification, three machine learning classifiers, support vector machine (SVM) [40], extreme gradient boosting (XGB) [41], and random forest (RF) [42] were implemented. Model validation was tried in our study with tenfold cross-validation. The classification algorithms were implemented using the 'caret' package in R software. The performance of classifiers was evaluated on the basis of the AUC value.

The workflow for imaging data processing and radiomics analysis is presented in Fig. 3.

Statistical analysis

The statistical analysis was conducted with R software (R Foundation for Statistical Computing) [39]. Three machine learning algorithms were applied with the R packages "e1071," "kernlab," "xgboost," and "randomForest." The performance of classifiers was evaluated using the area under the curve (AUC). AUC, accuracy, sensitivity, and specificity of each classifier were evaluated by the package "MLeval." The DeLong test was evaluated by package "pROC." A p value less than 0.05 indicated a statistically significant difference. The group differences were assessed using a Mann–Whitney U test for continuous variables. Univariate and bivariate analyses were performed with SPSS software 26 (IBM Corp.).

Results

Patient demographics and visual assessment of conventional MRI

The details of the patients are shown in Table 2. All of the patients were female. A significant difference was found between the phyllodes tumor groups and the fibroadenoma group regarding age, size, internal septation, strong lobulation, and cyst ($p < 0.05$). No significant difference was found between the two groups in terms of the heterogeneity of T2WI, the heterogeneity of the late-phase DCE-MRI, and TIC pattern. Older age, larger size, the presence of internal septation, the presence of strong lobulation, and the presence of cyst demonstrated the strongest statistical association with phyllodes tumors.

Table 2 Demographic characteristics of patients and visual assessment of conventional magnetic resonance imaging

Characteristics	Phyllodes tumor $N=32$	Fibroadenoma $N=56$	p value
Mean age ± SD, years	47.6 ± 10.8	40.2 ± 12.5	0.012
Maximum diameter (mm)	51.2 ± 29.9	26.8 ± 17.9	< 0.000
Internal septation			0.002
Present	31 (96.9%)	38 (67.9%)	
Absent	1 (3%)	18 (32.1%)	
Strong lobulation			0.005
Present	19 (59.4%)	16 (28.6%)	
Absent	13 (41.0%)	40 (71.4%)	
T2WI			0.968
Heterogeneous	13 (41.0%)	23 (41.0%)	
Homogeneous	19 (59.4%)	33 (58.9%)	
Cyst			0.001
Present	16 (50.0%)	9 (16.1%)	
Absent	16 (50.0%)	47 (83.9%)	
Late-phase DCE-MRI			0.056
Heterogeneous	13 (40.1%)	12 (21.4%)	
Homogeneous	20 (62.5%)	44 (78.6%)	
TIC pattern			0.610
Persistent pattern	21 (65.6%)	34 (60.7%)	
Plateau pattern	11 (34.4%)	21 (37.5%)	
Washout pattern	0 (0%)	1 (1.8%)	
Phyllodes tumor grade			
Benign	16 (50.0%)		
Borderline	11 (34.4%)		
Malignant	5 (15.6%)		

SD, standard deviation; *T2WI*, T2-weighted image; *DCE*, dynamic contrast-enhanced; *TIC*, time-intensity curve; *MRI*, magnetic resonance imaging

Intra- and interobserver reproducibility of radiomics features

A total of 1100 radiomic features for each slice were extracted from T2WI, pre-T1WI, the first-phase of DCE-MRI, and the late-phase of DCE-MRI. Among the 1100 extracted features, 30 features with an interobserver correlation coefficient (ICC) value less than 0.81 for intra- and inter-reader reproducibility were excluded. On the other hand, 1070 features with an ICC equal or higher than 0.81 were included in the subsequent feature selection process.

LASSO logistic regression

To identify the relevant predictors, all explanatory features extracted from MR images of the training set were included in the LASSO logistic regression. Features with regression coefficients of zero were eliminated (Fig. 4a and b).

Finally, 35 features (8 histogram features, 8 co-occurrence matrix features, 5 run-length matrix features, 8 autoregressive model features, and 8 wavelet transform features) out of 1070 radiomics features were selected to build the radiomics model (Table 3). The heat map of the selected features is presented in Fig. 5 and shows the distribution differences of normalized radiomics feature values.

Diagnostic performance of the radiomics model with machine learning classifiers

The diagnostic performance for the differentiation of phyllodes tumor from fibroadenoma of the radiomics model was evaluated using receiver operating characteristic (ROC) curves of the training and validation sets. Of the three machine learning classifiers, SVM yielded the highest AUC of 0.99 (95% confidence interval [CI]: 0.98–1.00), with a sensitivity 97.0% (95% CI: 92.0–99.0), specificity 98.0% (95% CI: 93.0–99.0), and accuracy 97.5% (95% CI: 96.0–98.8) in the training set. XGB yielded an AUC of 0.98 (95% CI: 0.96–1.00), and RF yielded an AUC of 0.98 (95% CI: 0.96–1.00) in the training set.

The radiomics model with SVM yielded an AUC of 0.96 (95% CI: 0.92–1.00), with a sensitivity 90.0% (95% CI: 79.0–96.0), specificity 92.0% (95% CI: 81.0–97.0), and accuracy 91.0% (95% CI: 87.9–93.7) in the validation set. XGB yielded an AUC of 0.93 (95% CI: 0.88–0.98) and RF yielded an AUC of 0.91 (95% CI: 0.85–0.97) in the validation set.

The radiological model

To evaluate the diagnostic performance of the visual assessment, a radiological model was built. For the radiological model, the SVM classifier yielded an AUC of 0.77 (95% CI: 0.66–0.88). The sensitivity, specificity, and accuracy of the radiological model were 65.6%, 85.7%, and 78.0%, respectively.

The combined radiomics and visual assessment model

For the combined radiomics and radiological model, the SVM classifier yielded an AUC of 0.97 (95% CI: 0.94–1.00), with a sensitivity of 92.0%, specificity of 94.0%, and accuracy of 93.0%.

The diagnostic performance of the radiomics model, the radiological model, and the combined model are shown in Table 4. ROC curves of the radiomics model, radiological model, and the combined model are shown in Fig. 6.

The combined model and the radiomics model also had a significantly higher AUC than that of the radiological model ($p < 0.001$ and $p < 0.001$, respectively). The combined model had a comparatively higher AUC than that of the radiomics model in the validation set, but this was not significantly different ($p = 0.391$).

Discussion

In the current study, we developed and validated an MRI-based radiomics model for differentiating phyllodes tumor from fibroadenoma. The proposed radiomics model with SVM showed good diagnostic performance for differentiating phyllodes tumor from fibroadenoma, with AUC values of 0.99 and 0.96 when applied to the training and validation sets, respectively. The radiomics model could significantly improve diagnostic performance compared to the radiological model, with an AUC of 0.96 and 0.76 in the validation set. The combined model achieved the highest performance

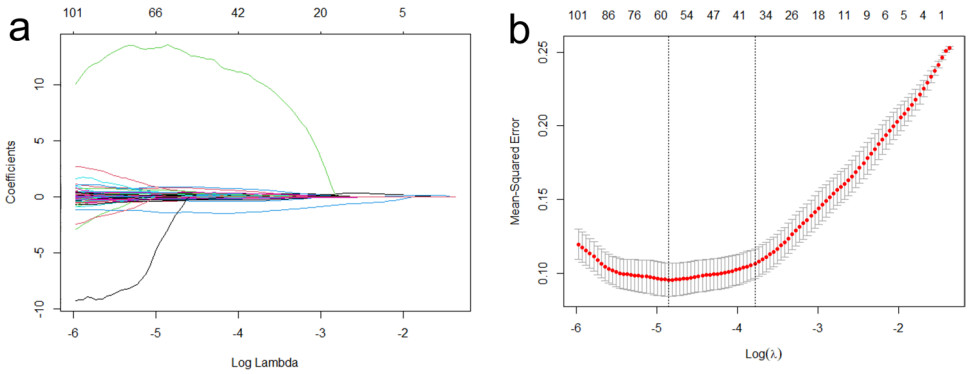

Fig. 4 Radiomics feature selection using the least absolute shrinkage and selection operator (LASSO) regression model. **a** Tuning parameter (λ) selection in the LASSO model used tenfold cross-validation via the minimum criterion. The optimal values of the LASSO tuning parameter (λ) are indicated by dotted vertical lines, and a value λ of 0.0078 was chosen. **b** LASSO coefficient profiles of 1,070 radiomics features. A coefficient profile plot was generated versus the selected log λ values using tenfold cross-validation. Thirty-five radiomics features with nonzero coefficients were selected

Table 3 Radiomics features selected by the least absolute shrinkage and selection operator regression

Feature category	Feature name	MRI sequence
Histogram	Skewness	T2WI
Histogram	Perc.10%	T2WI
Run-length matrix	Horzl_ShrtREmp	T2WI
Autoregressive model	Teta4	T2WI
Autoregressive model	Sigma	T2WI
Wavelet transform	WavEnLL_s-1	T2WI
Wavelet transform	WavEnLH_s-3	T2WI
Wavelet transform	WavEnLH_s-4	T2WI
Wavelet transform	WavEnHL_s-4	T2WI
Histogram	Kurtosis	T1WI
Histogram	Perc.01%	T1WI
Co-occurrence matrix	S(0,1)Correlat	T1WI
Co-occurrence matrix	S(1,1)Correlat	T1WI
Co-occurrence matrix	S(2,0)Correlat	T1WI
Co-occurrence matrix	S(5,-5)AngScMom	T1WI
Co-occurrence matrix	S(5,-5)SumEntrp	T1WI
Autoregressive model	Teta4	T1WI
Wavelet transform	WavEnHH_s-4	T1WI
Histogram	Skewness	First-phase of DCE-MRI
Histogram	Perc.01%	First-phase of DCE-MRI
Co-occurrence matrix	S(1,1)Correlat	First-phase of DCE-MRI
Run-length matrix	Horzl_GLevNonU	First-phase of DCE-MRI
Run-length matrix	Horzl_ShrtREmp	First-phase of DCE-MRI
Run-length matrix	Vertl_LngREmph	First-phase of DCE-MRI
Autoregressive model	Teta4	First-phase of DCE-MRI
Wavelet transform	WavEnHL_s-4	First-phase of DCE-MRI
Histogram	Kurtosis	Late-phase of DCE-MRI
Histogram	Perc.01%	Late-phase of DCE-MRI
Co-occurrence matrix	S(0,1)Correlat	Late-phase of DCE-MRI
Co-occurrence matrix	S(1,-1)Correlat	Late-phase of DCE-MRI
Run-length matrix	Vertl_ShrtREmp	Late-phase of DCE-MRI
Autoregressive model	Teta1	Late-phase of DCE-MRI
Autoregressive model	Teta4	Late-phase of DCE-MRI
Wavelet transform	WavEnHL_s-1	Late-phase of DCE-MRI
Wavelet transform	WavEnHH_s-1	Late-phase of DCE-MRI

LASSO, least absolute shrinkage and selection operator; *T1WI*, T1-weighted imaging; *T2WI*, T2-weighted imaging; *DCE*, dynamic contrast-enhanced; *MRI*, magnetic resonance imaging

with an AUC value of 0.97. In comparison with the radiological model, our combined model and radiomics model yielded higher performance.

Mai et al. were the first to investigate machine learning–based MRI texture analysis for differentiating phyllodes tumor from fibroadenoma [22]. They reported that a combination of clinical and conventional MRI features with texture features from T2W- short tau inversion recovery sequences yielded the highest AUC of 0.95. Similarly, in the present study, when combined with the radiological model, the discriminative performance of the radiomics model (0.96) can be improved to an AUC value of 0.97. This indicates that radiomics features and conventional MRI features provide supplementary information to differentiate phyllodes tumor from fibroadenoma. To build a robust prediction model, the visual assessment by radiologists should be combined with the radiomics features for high diagnostic ability.

Our study has a number of unique characteristics compared to the previous study.

First, the dataset in our patient cohort was imbalanced regarding the distribution of the two classes, comprising of 32 patients with phyllodes tumors and 56 patients with fibroadenomas; this was due to the low prevalence of phyllodes tumor [43]. The small patient population might have led to a risk of overfitting regarding machine learning-based classifications. We performed data augmentation in order to achieve better class balance and to avoid model overfitting before further evaluation.

Second, ComBat harmonization was used to remove the possible effects caused by different scanners and different magnetic field strengths. MR images were acquired from three different scanners with different magnetic field strengths, which can affect the extracted features. The ComBat harmonization method has been previously used in several MR radiomic studies and has confirmed the effectiveness of the harmonization for MR images, and harmonization did not alter the discriminant information conveyed by the features [36, 37, 44, 45].

Third, generally, a high number of radiomics features are extracted from images in radiomics analysis. Since some features are redundant and unstable, it is crucial to select significant and stable features in machine-learning-based radiomics analysis. Thus, LASSO regression was employed to select robust features and remove redundant features, achieving robust classification performance [37, 38, 46]. In this study, among 1070 features, 8 histogram features, 8 co-occurrence matrix features, 5 run-length matrix features, 6 autoregressive model features, and 8 wavelet transform features were selected by LASSO logistic regression. It can be seen that histogram features, co-occurrence matrix features, and wavelet transform features play important roles in discriminating phyllodes tumors from fibroadenoma.

Finally, Mai et al. applied only one machine learning classifier (k-nearest neighbor) [47] to differentiate phyllodes tumor from fibroadenoma [22]. However, when less data is available, several experiments with various machine learning classifiers could be needed to find the best machine learning scheme. A previous study suggested that the diagnostic performance of radiomics analysis is highly dependent on the choice of the machine leaning classifier [24]. In this study, we applied three machine learning classifiers, SVM [40], XGB [41], and RF [42] and obtained an AUC range of 0.91–0.96 in validation set. Our

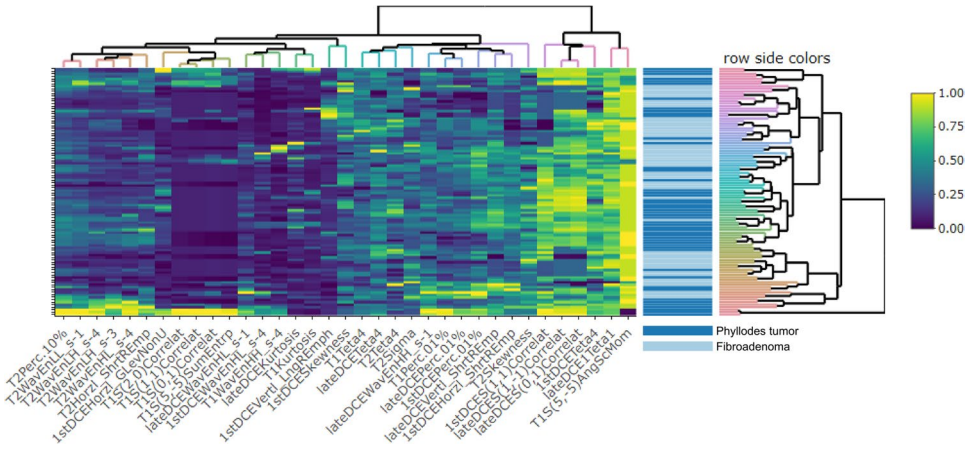

Fig. 5 Heat map of the selected features after least absolute shrinkage and selection operator (LASSO) regression. The *x* axis refers to radiomic features, and *y* axis refers to different subjects. Dendrograms regarding radiomics and subjects were displayed to facilitate the visualization of the radiomic patterns. The type of tumor for each subject was indicated by different colors (phyllodes tumor as blue/fibroadenoma as light blue)

Table 4 Diagnostic performance of the radiomics model, the radiological model, and the combined model

Data set		ML classifier	AUC	Sensitivity	Specificity	Accuracy
Training set ($n=200$)		SVM	0.99 (0.98–1.00)	97.0% (92.0–99.0)	98.0% (93.0–99.0)	97.5% (96.0–98.8)
		XGB	0.98 (0.96–1.00)	93.0% (86.0–97.0)	97.0% (92.0–99.0)	95.0% (92.6–97.1)
		RF	0.98 (0.96–1.00)	92.0% (85.0–96.0)	95.0% (89.0–98.0)	93.5% (92.0–94.8)
Validation set ($n=100$)	Radiomics model	SVM	0.96 (0.92–1.00)	90.0% (79.0–96.0)	92.0% (81.0–97.0)	91.0% (87.9–93.7)
		XBG	0.93 (0.88–0.98)	88.0% (76.0–94.0)	86.0% (74.0–93.0)	87.0% (82.9–90.6)
		RF	0.91 (0.85–0.97)	84.0% (71.0–92.0)	90.0% (79.0–96.0)	87.0% (84.3–89.5)
	Radiological model	SVM	0.77 (0.66–0.88)	65.6% (48.0–80.0)	85.7% (74.0–93.0)	78.0% (73.7–81.9)
		XBG	0.78 (0.67–0.89)	65.6% (48.0–80.0)	82.1% (70.0–90.0)	76.1% (71.3–79.7)
		RF	0.78 (0.67–0.89)	65.6% (48.0–80.0)	89.3% (79.0–95.0)	80.7% (70.8–88.7)
	Combined model	SVM	0.97 (0.94–1.00)	92.0% (81.0–97.0)	94.0% (84.0–98.0)	93.0% (90.4–95.2)
		XBG	0.95 (0.91–0.99)	92.0% (81.0–97.0)	92.0% (81.0–97.0)	92.0% (90.7–93.2)
		RF	0.94 (0.89–0.99)	90.0% (79.0–96.0)	90.0% (79.0–96.0)	90.0% (84.5–94.4)

95% confidence intervals are in parenthesis

ML, machine learning; *SVM*, support vector machine; *XBG*, extreme gradient boosting; *RF*, random forest; *AUC*, area under the curve

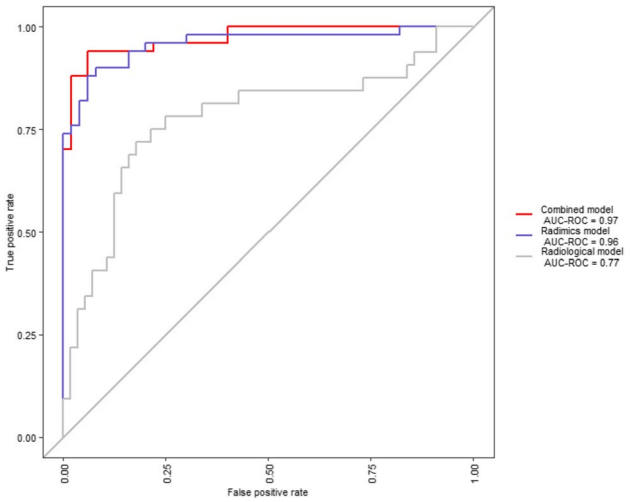

Fig. 6 Receiver operating characteristic (ROC) curves of the radiomics, radiological, and combined models

study found that the radiomics model with an SVM classifier had the highest AUC value among the three machine learning classifiers.

Our study has several limitations. First, any retrospective single-center study may have a selection bias. Therefore, future studies should have a large sample size, multivendor images, and an external test set [48]. Second, our sample size was small, with a tumor class imbalance. Unfortunately, the low occurrence rate of phyllodes tumors determines that a large sample size will be hard to achieve [43]. Small sample size and class imbalance can cause overfitting in machine learning classifications, which we tried to address by applying data augmentation, which has been shown to be successful [27, 29, 30]. Third, the 2D ROIs were manually drawn. Although 3D volumetric analysis had a better performance than 2D analysis [49], we think that 3D VOI would not be clinically practical due

to excessive segmentation duration. 2D ROIs are easier to calculate with less time consumption [50]. Fourth, MR image data were collected from three different scanners with different magnetic field strengths, which can affect the extracted features [26, 27]. In the present study, a ComBat harmonization method was used to remove scanner-specific effects from features.

Conclusions

In conclusion, our study demonstrates that an MRI-based radiomics model can differentiate phyllodes tumor from fibroadenoma accurately and robustly and can serve as a valuable clinical tool for the clinical decision-making process.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-021-08510-8.

Acknowledgements We thank Dr. Yoshiro Otsuki from the Department of Pathology, Seirei Hamamatsu General Hospital, for the pathological data.

Funding The authors state that this work has not received any funding.

Declarations

Guarantor The scientific guarantor of this publication is Takayuki Masui, M.D., Ph.D.

Conflict of interest MT, TM, KT, TY, MK, SI, YN, and SG declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohorts overlap A part of our study population has been previously presented in ECR2019.

Methodology
- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. Rowell MD, Perry RR, Hsiu JG, Barranco SC (1993) Phyllodes tumors. Am J Surg 165:376–379
2. Reinfuss M, Mituś J, Duda K, Stelmach A, Ryś J, Smolak K (1996) The treatment and prognosis of patients with phyllodes tumor of the breast: an analysis of 170 cases. Cancer 77:910–916
3. Geisler DP, Boyle MJ, Malnar KF et al (2000) Phyllodes tumors of the breast: a review of 32 cases. Am Surg 66:360–366
4. O'Malley FP, Pinder SE, Goldblum JR (2006) Fibroepithelial lesions, including fibroadenoma and phyllodes tumor. Breast pathology, 1st edn. Churchill Livingstone/ Elsevier, Philadelphia, PA, pp 109–115
5. Chao TC, Lo YF, Chen SC, Chen MF (2002) Sonographic features of phyllodes tumors of the breast. Ultrasound Obstet Gynecol 20:64–71
6. Foxcroft LM, Evans EB, Porter AJ (2007) Difficulties in the preoperative diagnosis of phyllodes tumours of the breast: a study of 84 cases. Breast 16:27–37
7. Chaney AW, Pollack A, McNeese MD et al (2000) Primary treatment of cystosarcoma phyllodes of the breast. Cancer 89:1502–1511
8. Yabuuchi H, Soeda H, Matsuo Y et al (2006) Phyllodes tumor of the breast: correlation between MR findings and histologic grade. Radiology 241:702–709
9. Wurdinger S, Herzog AB, Fischer DR et al (2005) Differentiation of phyllodes breast tumors from fibroadenomas on MRI. AJR Am J Roentgenol 185:1317–1321
10. Kamitani T, Matsuo Y, Yabuuchi H et al (2014) Differentiation between benign phyllodes tumors and fibroadenomas of the breast on MR imaging. Eur J Radiol 83:1344–1349
11. Davnall F, Yip CS, Ljungqvist G et al (2012) Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging 3:573–589
12. Chamming's F, Ueno Y, Ferré R et al (2018) Features from computerized texture analysis of breast cancers at pretreatment MR imaging are associated with response to neoadjuvant chemotherapy. Radiology 286:412–420
13. Choi N, Kim K, Shin KH et al (2018) Malignant and borderline phyllodes tumors of the breast: a multicenter study of 362 patients (KROG 16–08). Breast Cancer Res Treat 171:335–344
14. Balaji R, Ramachandran KN (2009) Magnetic resonance imaging of a benign phyllodes tumor of the breast. Breast Care (Basel) 4:189–191
15. Lo Gullo R, Daimiel I, Rossi Saccarelli C et al (2020) Improved characterization of sub-centimeter enhancing breast masses on MRI with radiomics and machine learning in BRCA mutation carriers. Eur Radiol 30:6721–6731
16. Li X, Jiang N, Zhang C, Luo X, Zhong P, Fang J (2021) Value of conventional magnetic resonance imaging texture analysis in the differential diagnosis of benign and borderline/malignant phyllodes tumors of the breast. Cancer Imaging 21:29
17. Dong Y, Feng Q, Yang W et al (2018) Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI. Eur Radiol 28:582–591
18. Parikh J, Selmi M, Charles-Edwards G et al (2014) Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. Radiology 272:100–112
19. Kim JH, Ko ES, Lim Y et al (2017) Breast cancer heterogeneity: MR imaging texture analysis and survival outcomes. Radiology 282:665–675
20. Xie T, Zhao Q, Fu C et al (2019) Differentiation of triple-negative breast cancer from other subtypes through whole-tumor histogram analysis on multiparametric MR imaging. Eur Radiol 29:2535–2544
21. Fan M, Zhang P, Wang Y et al (2019) Radiomic analysis of imaging heterogeneity in tumours and the surrounding parenchyma based on unsupervised decomposition of DCE-MRI for predicting molecular subtypes of breast cancer. Eur Radiol 29:4456–4467

22. Mai H, Mao Y, Dong T et al (2019) The utility of texture analysis based on breast magnetic resonance imaging in differentiating phyllodes tumors from fibroadenomas. Front Oncol 9:1021
23. Jiang N, Zhong L, Zhang C, Luo X, Zhong P, Li X (2021) Value of conventional MRI texture analysis in the differential diagnosis of phyllodes tumors and fibroadenomas of the breast. Breast Care (Basel) 16:283–290
24. Nakagawa M, Nakaura T, Namimoto T et al (2019) Machine learning to differentiate T2-weighted hyperintense uterine leiomyomas from uterine sarcomas by utilizing multiparametric magnetic resonance quantitative imaging features. Acad Radiol 26:1390–1399
25. Fruehwald-Pallamar J, Czerny C, Holzer-Fruehwald L et al (2013) Texture-based and diffusion-weighted discrimination of parotid gland lesions on MR images at 3.0 Tesla. NMR Biomed 26:1372–1379
26. Jaitly N, Hinton GE (2011) Learning a better representation of speech soundwaves using restricted boltzmann machines. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP):5884–5887. https://doi.org/10.1109/ICASSP.2011.5947700
27. Zhu JCN, Perkins H, Zhang B (2014) Gibbs max-margin topic models with data augmentation. J Mach Learn Res 15:1073–1110
28. Girish GN, Thakur B, Chowdhury SR, Kothari AR, Rajan J (2019) Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model. IEEE J Biomed Health Inform 23:296–304
29. Wong SC, Gatt A, Stamatescu V, McDonnell MD (2016) Understanding data augmentation for classification:when to warp? 2016 International Conference on Digital Image Computing (DICTA) 1–6. https://doi.org/10.1109/DICTA.2016.7797091
30. Kocak B, Durmaz ES, Ates E, Ulusan MB (2019) Radiogenomics in clear cell renal cell carcinoma: machine learning-based high-dimensional quantitative CT texture analysis in predicting PBRM1 mutation status. AJR Am J Roentgenol 212:W55–W63
31. He H, Garcia EA (2009) Learning from imbalanced data. IEE Trans Knowl Data Eng 21:1263–1284
32. Szczypiński PM, Strzelecki M, Materka A, Klepaczko A (2009) MaZda–a software package for image texture analysis. Comput Methods Programs Biomed 94:66–76
33. Morris E, Comstock C, Lee C (2013) ACR BI-RADS® Atlas. Breast Imaging Reporting and Data System, Reston, VA, American College of Radiology
34 Rao AA, Feneis J, Lalonde C, Ojeda-Fournier H (2016) A pictorial review of changes in the BI-RADS fifth edition. Radiographics 36:623–639
35. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8:118–127
36. Orlhac F, Lecler A, Savatovski J et al (2020) How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. Eur Radiol. https://doi.org/10.1007/s00330-020-07284-9
37. Hu J, Zhao Y, Li M et al (2020) Machine-learning-based computed tomography radiomic analysis for histologic subtype classification of thymic epithelial tumours. Eur J Radiol 126:108929
38. Zheng YM, Li J, Liu S et al (2020) MRI-Based radiomics nomogram for differentiation of benign and malignant lesions of the parotid gland. Eur Radiol. https://doi.org/10.1007/s00330-020-07483-4
39. Foundation R (2018) R: a language and environment for statistical computing. Austria, Vienna
40. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20:273–297. https://doi.org/10.1007/BF00994018
41. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. Association for Computing Machinery, San Francisco, California, pp 785–794
42. Breiman L (2001) Random forests. Machine Learning 45:5–32. https://doi.org/10.1023/A:1010933404324
43. Liberman L, Bonaccio E, Hamele-Bena D, Abramson AF, Cohen MA, Dershaw DD (1996) Benign and malignant phyllodes tumors: mammographic and sonographic findings. Radiology 198:121–124
44. Lucia F, Visvikis D, Vallières M et al (2019) External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging 46:864–877
45. Wang H, Zhang J, Bao S et al (2020) Preoperative MRI-based radiomic machine-learning nomogram may accurately distinguish between benign and malignant soft-tissue lesions: a two-center study. J Magn Reson Imaging 52:873–882
46. Yin P, Mao N, Zhao C et al (2019) Comparison of radiomics machine-learning classifiers and feature selection for differentiation of sacral chordoma and sacral giant cell tumour based on 3D computed tomography features. Eur Radiol 29:1841–1847
47. Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician 46:175–185
48. Bluemke DA, Moy L, Bredella MA et al (2020) Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers-from the Radiology Editorial Board. Radiology 294:487–489
49. Yang L, Yang J, Zhou X et al (2019) Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. Eur Radiol 29:2196–2206
50. Lubner MG, Stabo N, Lubner SJ et al (2015) CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. Abdom Imaging 40:2331–2337