European Journal of Radiology Open 13 (2024) 100608

Contents lists available at ScienceDirect

European Journal of Radiology Open

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

MRI-based radiomics machine learning model to differentiate non-clear cell renal cell carcinoma from benign renal tumors

Ruiting Wang [a,b,1], Lianting Zhong [c,1], Pingyi Zhu [a,b], Xianpan Pan [d], Lei Chen [d], Jianjun Zhou [c,e,f,*,2], Yuqin Ding [a,b,**,2]

[a] Department of Radiology, Zhongshan Hospital, Fudan University, Shanghai, China
[b] Shanghai Institute of Medical Imaging, Shanghai, China
[c] Department of Radiology, Zhongshan Hospital (Xiamen), Fudan University, Xiamen, Fujian, China
[d] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China
[e] Xiamen Municipal Clinical Research Center for Medical Imaging, Xiamen, Fujian, China
[f] Fujian Province Key Clinical Specialty for Medical Imaging, Xiamen, Fujian, China

ARTICLE INFO

Keywords:
Renal cell carcinoma
Benign renal tumors
Machine learning
Radiomics
Magnetic resonance imaging

ABSTRACT

Purpose: We aim to develop an MRI-based radiomics model to improve the accuracy of differentiating non-ccRCC from benign renal tumors preoperatively.
Methods: The retrospective study included 195 patients with pathologically confirmed renal tumors (134 non-ccRCCs and 61 benign renal tumors) who underwent preoperative renal mass protocol MRI examinations. The patients were divided into a training set (n = 136) and test set (n = 59). Simple t-test and the Least Absolute Shrink and Selection Operator (LASSO) were used to select the most valuable features and the rad-scores of them were calculated. The clinicoradiologic models, single-sequence radiomics models, multi-sequence radiomics models and combined models for differentiation were constructed with 2 classifiers (support vector machine (SVM), logistic regression (LR)) in the training set and used for differentiation in the test set. Ten-fold cross validation was applied to obtain the optimal hyperparameters of the models. The performances of the models were evaluated by the area under the receiver operating characteristic (ROC) curve (AUC). Delong's test was performed to compare the performances of models.
Results: After univariate and multivariate logistic regression analysis, the independent risk factors to differentiate non-ccRCC from benign renal tumors were selected as follows: age, tumor region, hemorrhage, pseudocapsule and enhancement degree. Among the 14 machine learning classification models constructed, the combined model with LR has the highest efficiency in differentiating non-ccRCC from benign renal tumors. The AUC in the training set is 0.964, and the accuracy is 0.919. The AUC in the test set is 0.936, and the accuracy is 0.864.
Conclusion: The MRI-based radiomics machine learning is feasible to differentiate non-ccRCC from benign renal tumors, which could improve the accuracy of clinical diagnosis.

1. Introduction

As the use of various imaging methods continues increasing, the incidental detection of renal masses has also continued rising [1–3]. Among them, renal cell carcinomas (RCCs) are the most common. Clear cell renal cell carcinoma (ccRCC), papillary renal cell carcinoma (pRCC) and chromophobe renal cell carcinoma (chRCC) are the three most common subtypes of RCC. Angiomyolipoma (AML) and renal oncocytoma (RO) are the most common solid benign renal tumors [4]. Percutaneous renal biopsy could provide preoperative pathology diagnosis, whose accuracy is reported ranging from 70 % to 90 % [5]. However, the histologic and molecular heterogeneity limit the accuracy of biopsy

* Correspondence to: Department of Radiology, Zhongshan Hospital (Xiamen), Fudan University, 668 Jinhu Road, Huli District, Xiamen 361015, China.
** Correspondence to: Department of Radiology, Zhongshan Hospital, Shanghai Institute of Medical Imaging, 180 Fenglin Road, Shanghai 200032, PR China
E-mail addresses: zhoujianjunzs@126.com (J. Zhou), nancydingding@126.com (Y. Ding).
[1] Ruiting Wang and Lianting Zhong are listed as co-first authors of this article, who have the same contribution to the data collection and manuscript writing.
[2] Yuqin Ding and Jianjun Zhou are listed as co-corresponding authors of this article, who have the same contribution to the design and guidance of this manuscript.

https://doi.org/10.1016/j.ejro.2024.100608
Received 4 August 2024; Received in revised form 9 October 2024; Accepted 20 October 2024

results [6,7]. As an invasive method, biopsy can lead to some complications. The differential diagnosis on benign and malignant renal tumors by computed tomography (CT) and magnetic resonance imaging (MRI), which can provide qualitative and comprehensive assessment non-invasively, has always been a research hot spot. Accurate preoperative differential diagnosis has a great impact on clinical management as benign renal tumors could be treated conservatively to avoid physical and economic burden caused by surgery, while RCCs need to be treated by partial or radical nephrectomy due to the potential to metastasis and death [8]. Because of the high morbidity, there have been many imaging studies on ccRCCs. In addition to the typical imaging manifestations, they are easier to be differentiated from other renal tumors. Owing to the lower incidence and uncertainty in preoperative diagnosis, it is difficult to differentiate non-ccRCC from benign renal tumors. In the most cases, classic AMLs can be diagnosed by identifying the intratumoral macroscopic fat component in the images [9]. However, AML without visible fat (AMLwvf) is typically shows homogeneous enhancement and prolonged enhancement pattern, which are similar to pRCC and chRCC [10]. Due to the similar origination from the collecting duct system, RO and chromophobe renal cell carcinoma (chRCC) overlap in morphological and radiologic manifestations [11]. Although central scar and segmental enhancement inversion are the typical imaging features of oncocytoma, they are lack of specificity because some chRCCs also have the similar features [12]. Due to the high preoperative misdiagnosis rate, many benign renal tumors tend to be over treated [13,14].

As an emerging field in image analysis, radiomics can extract huge amounts of high-throughput quantitative features from images beyond what naked eyes are capable of detecting and characterize more global intratumoral heterogeneity non-invasively [15–17]. These features are able to be applied in machine learning (ML) algorithms and contribute to lesion detection, diagnosis, assessment of prognosis and prediction of therapeutic effects [18]. In general, ML uses a training set to perform tasks such as feature selection and tuning hyperparameter, while a validation or testing set to evaluate the performance of the model [5]. Many studies have shown that MRI-based radiomics had desirable performance in distinguishing renal tumors [19–21]. In this study, we aimed at developing an MRI-based radiomics ML model to differentiate the most common non-ccRCCs (pRCC and chRCC) from benign renal tumors (AMLwvf and RO).

2. Methods

2.1. Patient cohort

We screened the clinical records of patients with postoperative pathologically confirmed with non-ccRCCs and renal benign tumors between January 2009 and January 2021, which were retrieved from electronic medical record (EMR) system of Zhongshan hospital, Fudan University. The hospital ethic committee approved this retrospective study and waived patient informed consent. The inclusion criteria were as follows: (1) patients had been pathologically proven as non-ccRCCs (including pRCC and chRCC) and benign renal tumors (including AMLwvf and RO); (2) patients had undergone contrast-enhanced MRI renal mass scans, including T2-weighted imaging (T2), T1-weighted imaging at unenhanced phase (UP), corticomedullary phase (CMP) and nephrographic phase (NP); (3) No visible fat component could be found in the renal tumors on all MRI scans; (4) patients underwent surgery within 1 months after MRI examinations. The exclusion criteria were as follows: (1) patients had renal surgery or treatment before MRI scans; (2) poor image quality such as prominent artifacts; (3) tumors smaller than 1.0 cm in diameter. The 1.0 cm lower threshold was selected to avoid potential confounding from partial volume averaging in smaller renal lesions. In total, 195 patients including 134 cases of non-ccRCCs (79 pRCCs and 55 chRCCs) and 61 cases of benign renal tumors without visible fat component (46 AMLwvfs and 15 ROs) were enrolled in this study according to the inclusion and exclusion criteria. All the patients were randomly assigned to the training set (n = 136) and test set (n = 59) by a ratio of 7:3.

2.2. MRI acquisition protocols

All patients were examined with the 3.0 T MRI scanners (Magnetom Aera; Siemens Healthineers) in our hospital. The routine MRI scan sequences included axial three-dimensional chemical shift in- and opposed-phase T1-weighted interpolated breath-hold examination, diffusion weighted imaging (DWI) with two *b* values (0, 500 sec/mm^2) or three *b* values (0, 50, 500 sec/mm^2), T2-weighted turbo spin-echo pulse sequence with fat suppression (T2), dynamic three-dimensional T1-weighted volumetric interpolated breath-hold examination at unenhanced phase (UP), corticomedullary phase (30–35 s) (CMP), nephrographic phase (80–90 s) (NP), and excretory phase (180 s) after injection of 0.1 mmol/kg gadopentetate dimeglumine (Magnevist; Bayer Schering Pharma AG) at a rate of 2 ML/s.

2.3. Clinicoradiologic features

We obtained the clinical data of the patients (age, gender, pathological reports, surgery records, etc) from the EMR system. The radiologic features were respectively reviewed by two radiologists (R.T.W. and Y.Q.D., 2 years and 10 years of abdominal MRI experience, respectively) who were blind to the pathological results. In case of any discrepancies, a consensus was reached after discussion. All the continuous variables were converted to categorical variables. The following qualitative radiologic features were evaluated: (1) tumor diameter ($\leq$ 4 cm, >4 cm and $\leq$ 7, >7 cm); (2) tumor region (left, right; upper, middle, lower; extra, mixed, intra); (3) tumor shape (round, irregular); (4) hemorrhage (yes, no); (5) necrosis (no, <50 %, $\geq$50 %); (6) cystic degeneration (no, <50 %, $\geq$50 %); (7) scar (yes, no); (8) Angular interface with renal parenchyma (yes, no); (9) capsule (yes, no); (10) boundary (well-defined, ill-defined) and (11) Signal intensity on T2WI image (high, not high). The dynamic enhancement radiologic features were as follows: (1) enhancement degree (obvious, moderate, mild); (2) enhancement pattern (wash-in and wash-out, persistent, delayed) and (3) uniformity on NP image (heterogeneous, homogeneous).

2.4. Radiomics analysis

Radiomics workflow (Fig. 2) comprised manual tumor segmentation, feature extraction and selection, ML model construction and evaluation. All the MRI images were anonymized and stored in DICOM format. The tumor boundaries of each slice were manually delineated on the four sequences to obtain the tailored volumetric area of interests (VOIs) by R.T.W. (2 years of abdominal MRI experience) with 3D Slicer software v. 4.11 (https://www.slicer.org/). Subsequently, the VOIs were reviewed and approved by Y.Q.D. (10 years of abdominal MRI experience). To extract robust features and assure reproducibility, 20 patients were randomly selected and the VOIs of them were delineated again by R.T.W. after 1 month to assess the inter-observer intraclass correlation coefficients (ICC).

Radiomics features were extracted with PyRadiomics (https://pyradiomics.readthedocs.io/en/v3.0.1/) embedded in the uAI Research Portal (Version: 20210730). A set of 2600 features were respectively extracted from each of the four sequences, mainly including 3 categories: shape, first-order statistics (histogram analysis) and second-order statistics (grey-level distribution of the image, including gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), gray level run length matrix (GLRLM), neighbouring gray tone difference matrix (NGTDM) and gray level dependence matrix (GLDM)). A total of 104 original image radiomics features were processed through 24 filters, resulting in 2496 high-dimensional image omics features. The

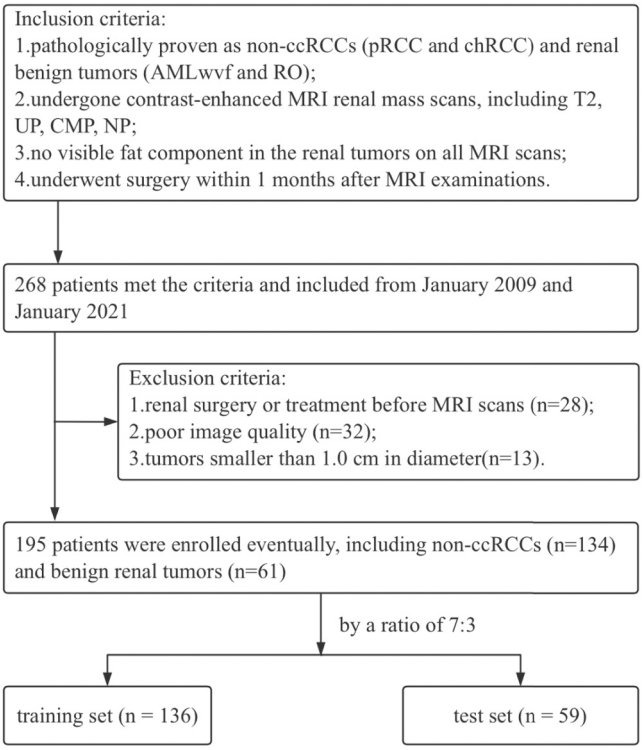

Fig. 1. The flow chart of the patient enrollment process.

24 filters adopted were as follows: Additive Gaussian noise Filter, Bilateral Filter, Binomial Blur Image Filter, Box Mean Filter, Box Sigma Image Filter, Curvature Flow Filter, Discrete Gaussian Filter, Laplacian Sharpening Filter, Mean Filter, Median Filter, Normalize Filter, Recursive Gaussian Filter, Shot Noise Filter, Smoothing Recursive Gaussian Filter, Speckle Noise Filter, LoG Filter, Wavelet Filter (LLH、LHL、LHH、HLL、HLH、HHL、HHH、LLL). Z-scores normalization was applied in these quantitative features before feature selection. Features with ICC $\geq$ 0.75 % suggested good consistency and were selected for further analysis [22]. Because feature numbers should be controlled less than 10 % of the sample size to avoid overfitting [23], these features were firstly selected by simple t-tests. Then, the Least Absolute Shrink and Selection Operator (LASSO) was adopted, a model which is most commonly used to analyze small samples with high-dimensional features and can select the features with strong relevance [24,25]. Ten-fold cross validation was applied to obtain the optimal hyperparameter λ by 10000 iterations. The LASSO method based on the optimal λ was used to select the features with non-zero coefficients. Finally, the selected features of single sequence were combined and selected by LASSO again to obtained an optimal multi-sequence radiomics model. The greater absolute value of feature correlation coefficient is, the stronger relevance between tumor and feature is. To compare these single- and multi-sequence models unbiasedly, the features of which absolute value of the coefficient ranking in the top eight were selected for each model considering the sample size and feature numbers. Finally, the radiomics score (rad--score) of the optimal sequence combination for each patient were calculated with the formula generated by a linear combination of the selected features, which were weighted by their coefficients. The Methodological Radiomics Score (METRICS) is made available by a large group of international domain experts, aiming at evaluating and improving the research quality in radiomics and machine learning [26]. Guideline suggested that a web application has been developed to calculate of the METRICS score (https://metricsscore.github.io/metrics/METRICS.html). The METRICS score of our study is 80.3 % with excellent quality category.

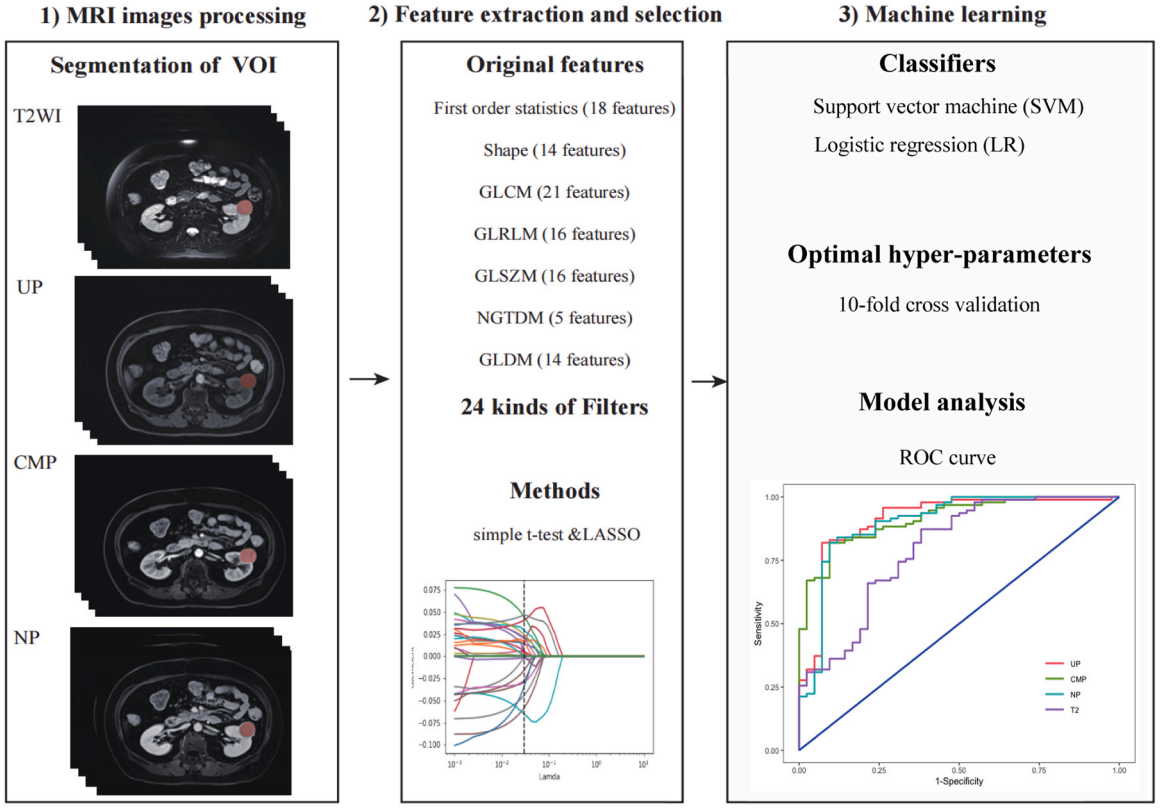

Fig. 2. Flowchart of radiomics analysis.

2.5. Model construction and evaluation

In this analysis, the models were constructed by support vector machine (SVM) and logistic regression (LR) classifiers with the selected features in the training set. The Radial Basis Function (RBF) kernel function, most widely used in SVM, was adopted in our study. There were two parameters in this function: the penalty parameter C and the kernel parameter γ. A 10-fold cross validation strategy was adopted to find the optimal combinations of the hyperparameters (Table S1) and the ROC curves were calculated in the test set to evaluated the models.

2.6. Statistical analysis

Statistical analysis was performed with IBM SPSS (v. 26) and R software (v. 4.1.3). Categorical variables were compared using chi-square test or Fisher's exact test. R software packages and Python v.

Table 1
Univariate and multivariate logistic regression analysis.

Variable	Univariate analysis			Multivariate analysis		
	OR	95 %CI	*p* value	OR	95 %CI	*p* value
Age (years)						
≤50	1.000			1.000		
>50	3.004	1.413–6.386	0.004*	6.019	1.527–23.729	0.010*
Gender						
Male	3.697	1.714–7.975	0.001*	2.901	0.869–9.686	0.083
Female	1.000			1.000		
Diameter (cm)						
≤4	1.000					
>4, ≤7	1.303	0.499–3.401	0.589			
>7	0.684	0.180–2.597	0.577			
Region1						
Left	1.250	0.603–2.590	0.548			
Right	1.000					
Region2						
Upper	1.000			1.000		
Middle	2.942	1.108–7.815	0.030*	2.825	0.659–12.104	0.162
Lower	1.471	0.632–3.425	0.371	2.154	0.488–9.501	0.311
Region3						
Extra	1.000			1.000		
Mixed	1.131	0.487–2.626	0.774	1.335	0.345–5.168	0.676
Intra	5.793	1.955–17.164	0.002*	6.232	1.291–30.083	0.023*
Shape						
Round	1.000					
Irregular	0.968	0.467–2.006	0.930			
Hemorrhage						
Yes	9.913	3.277–29.983	<0.001*	16.859	2.206–128.840	0.006*
No	1.000			1.000		
Necrosis						
No	1.000			1.000		
< 50 %	2.826	1.174–6.803	0.020*	0.507	0.107–2.406	0.392
≥50 %	1.833	0.183–18.364	0.606	0.631	0.006–63.817	0.845
Cystic degeneration						
No	1.000					
< 50 %	0.333	0.084–1.315	0.333			
≥50 %	1.250	0.241–6.497	1.250			
Scar						
No	1.000					
Yes	0.688	0.211–2.245	0.536			
Angular interface with renal parenchyma						
No	1.000					
Yes	0.437	0.163–1.171	0.100			
Capsule						
No	1.000			1.000		
Yes	3.642	1.701–7.797	0.001*	8.371	2.126–32.963	0.002*
Boundary						
Well-defined	1.000					
Ill-defined	2.381	0.498–11.378	0.277			
Signal intensity on T2WI image						
High	0.981	0.318–3.024	0.973			
Not high	1.000					
Enhancement degree						
Mild	29.469	6.216–139.695	<0.001*	74.115	6.017–912.929	0.001*
Moderate	3.129	1.326–7.382	0.009*	1.696	0.450–6.389	0.435
Obvious	1.000			1.000		
Enhancement pattern						
Wash-in and wash-out	1.000			1.000		
Persistent	2.800	0.877–8.944	0.082	2.015	0.301–13.502	0.470
Delayed	6.581	2.719–15.932	<0.001*	0.961	0.214–4.319	0.959
Uniformity on NP image						
Heterogeneous	1.581	0.759–3.293	0.221			
Homogeneous	1.000					

* Represents $p < 0.05$. T2WI: T2-weighted image; NP: Enhanced T1-weighted image in nephrographic phase; OR: odds ratio; CI: confidence interval;

3.7.7 (https://www.python.org) scikit-learn package v. 1.0.2 (https://scikit-learn.org/stable/index.html) were used to construct the models and evaluate the performances. Statistical comparisons of AUCs among different models were performed by the Delong's test. $P < 0.05$ was considered to be statistically significant.

3. Results

3.1. The performances of clinicoradiologic models

A total of 195 patients (134 non-ccRCCs (79 pRCCs and 55 chRCCs) and 61 benign renal tumors without visible fat (46 AMLwvfs, 15 ROs)) were recruited, including 136 patients in the training set and 59 in the test set. The preoperative clinicoradiologic features of the enrolled patients were shown in Table S2. As shown in Table 1, univariate and multivariate logistic regression analysis results indicated that age > 50-year-old ($p = 0.010$, OR = 6.019, 95 %CI: 1.527–23.729), intra region ($p = 0.023$, OR = 6.232, 95 %CI: 1.291–30.083), hemorrhage ($p = 0.006$, OR = 16.859, 95 %CI: 2.206–128.840), capsule ($p = 0.002$, OR = 8.371, 95 %CI: 2.126–32.963) and mild enhancement ($p = 0.001$, OR = 74.115, 95 %CI: 6.017–912.929), were the independent risk factors for non-ccRCCs. These clinicoradiologic features were used to construct classification models with SVM and LR. The performances of the clinicoradiologic models were shown in Table 3 and Fig. 3. The AUC of LR model was higher than SVM model but with no statistical difference in the test set (0.853 vs 0.802, $p = 0.153$, Delong's test).

3.2. Performances of single-sequence radiomics models for discrimination

The selected radiomics features of each single-sequence were shown in Table 2 and the performances of each model were shown in Table 3 and Fig. 3. For the AUC results of the vast models, they had good classification abilities. In the training set and test set, the AUCs of CMP were superior to other sequences and UP ranked second, while NP ranked the third and T2 performed the worst. The AUCs of CMP and UP in the test set were respectively 0.815 (95 % CI:0.706–0.923) and 0.754 (95 % CI:0.609–0.899) with SVM, and 0.797 (95 % CI:0.686–0.909) and 0.745 (95 % CI:0.600–0.890) with LR. For the two classifiers, the vast models based on SVM had higher AUCs than LR. For the SVM models, the Delong's test showed no statistical differences between CMP and UP, and between NP and T2 ($p = 0.514$, 0.119), but there were statistical differences between UP and T2, and between CMP and T2 ($p = 0.038$, 0.004).

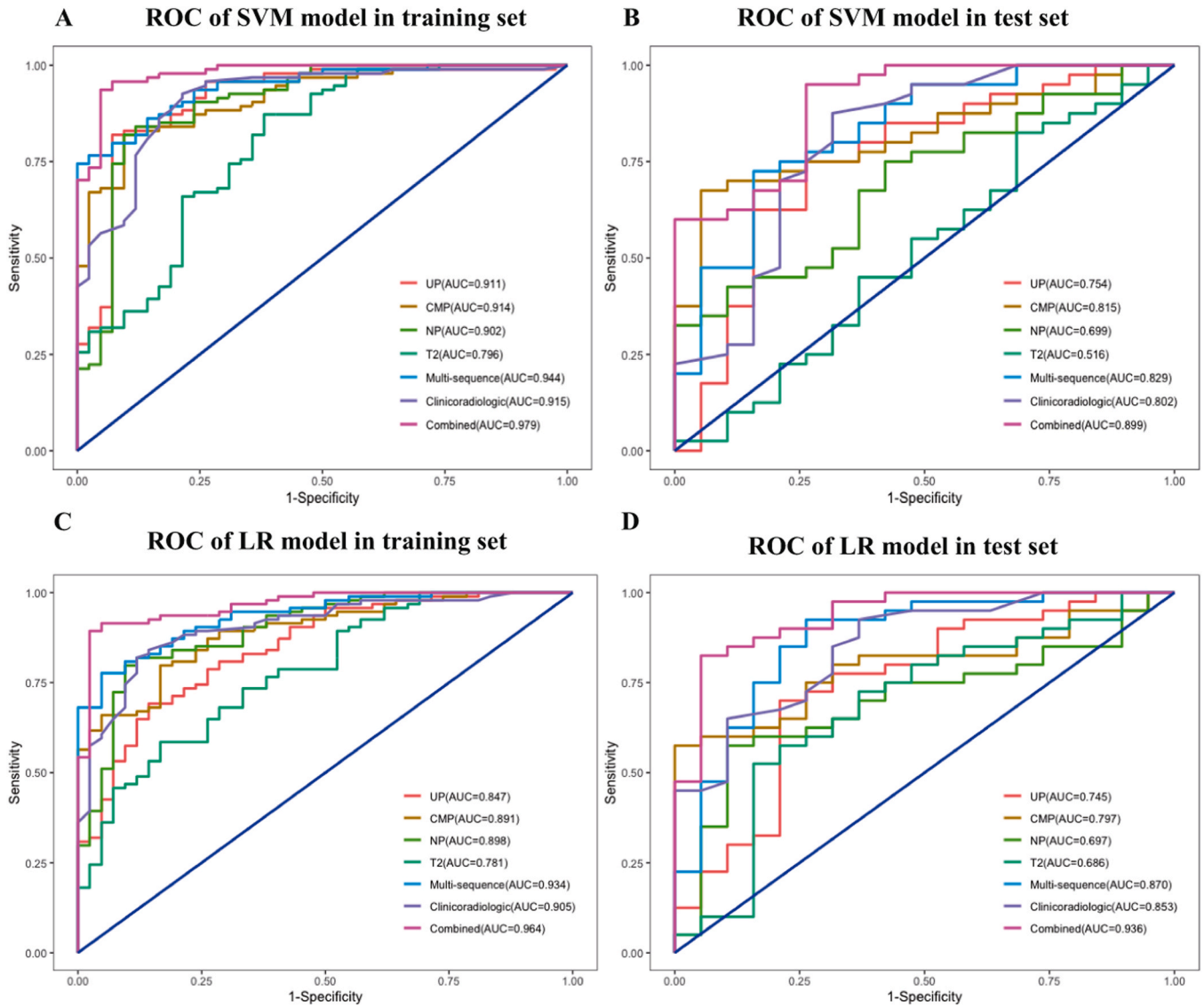

Fig. 3. The ROC curves of different models for differentiating non-ccRCC from benign renal tumors in training set and test set.

Table 2
The selected radiomics features extracted from each single-sequence.

Sequence	Features	Coefficients
UP	wavelet_firstorder_wavelet-HHH-Mean	0.084
	binomialblurimage_glszm_SizeZoneNonUniformityNormalized	−0.050
	normalize_firstorder_90Percentile	0.048
	wavelet_gldm_wavelet-LLL-DependenceVariance	−0.043
	wavelet_glszm_wavelet-LHH-SmallAreaLowGrayLevelEmphasis	0.034
	normalize_firstorder_RootMeanSquared	0.034
	normalize_firstorder_Mean	0.028
	wavelet_ngtdm_wavelet-LHL-Contrast	−0.023
CMP	discretegaussian_firstorder_Skewness	0.056
	log_firstorder_log-sigma−0–5-mm−3D-Skewness	0.052
	original_glcm_Idmn	0.047
	normalize_firstorder_RootMeanSquared	−0.044
	normalize_glszm_GrayLevelVariance	0.044
	wavelet_glcm_wavelet-HHL-ClusterShade	−0.037
	log_glszm_log-sigma−0–5-mm−3D-SmallAreaEmphasis	0.037
	wavelet_glszm_wavelet-HLH-SizeZoneNonUniformityNormalized	−0.026
NP	log_firstorder_log-sigma−0–5-mm−3D-Skewness	0.134
	wavelet_glcm_wavelet-LHH-Imc1	0.063
	wavelet_glszm_wavelet-HHL-SmallAreaLowGrayLevelEmphasis	0.062
	wavelet_firstorder_wavelet-HLH-Mean	−0.059
	wavelet_gldm_wavelet-LHH-DependenceNonUniformityNormalized	−0.048
	binomialblurimage_glcm_ClusterShade	0.040
	laplaciansharpening_firstorder_Skewness	0.037
	laplaciansharpening_gldm_LargeDependenceHighGrayLevelEmphasis	−0.027
T2	boxmean_firstorder_Skewness	−0.092
	log_glrlm_log-sigma−1–0-mm−3D-LongRunLowGrayLevelEmphasis	−0.074
	boxsigmaimage_glszm_SmallAreaEmphasis	−0.072
	log_glszm_log-sigma−1–5-mm−3D-SmallAreaLowGrayLevelEmphasis	−0.045
	wavelet_firstorder_wavelet-HLH-Skewness	0.045
	mean_glszm_SizeZoneNonUniformityNormalized	−0.044
	wavelet_firstorder_wavelet-HLL-Skewness	0.025
	wavelet_glcm_wavelet-LHH-Correlation	−0.020

UP: unenhanced T1-weighted image; CMP: enhanced T1-weighted image in corticomedullary phase; NP: enhanced T1-weighted image in nephrographic phase; T2: T2-weighted image; glcm: gray level cooccurrence matrix; gldm: gray level dependence matrix; glszm: gray level size zone matrix; ngtdm: neighbouring gray tone difference matrix; glrlm: gray level run length matrix.

Table 3
The performance of each model in training set and test set.

Features	Classifier	Training set (n=136)				Test set (n=59)			
		AUC (95 %CI)	Acc	Spe	Sen	AUC (95 %CI)	Acc	Spe	Sen
UP	SVM	0.911(0.852–0.970)	0.853	0.929	0.819	0.754(0.609–0.899)	0.746	0.737	0.750
	LR	0.847(0.778–0.915)	0.743	0.857	0.691	0.745(0.600–0.890)	0.729	0.789	0.700
CMP	SVM	0.914(0.867–0.961)	0.846	0.905	0.819	0.815(0.706–0.923)	0.763	0.947	0.675
	LR	0.891(0.838–0.944)	0.809	0.833	0.798	0.797(0.686–0.909)	0.712	1.000	0.575
NP	SVM	0.902(0.838–0.965)	0.846	0.905	0.819	0.699(0.561–0.837)	0.695	0.579	0.750
	LR	0.898(0.841–0.955)	0.831	0.905	0.798	0.697(0.557–0.838)	0.678	0.895	0.575
T2	SVM	0.796(0.711–0.881)	0.794	0.619	0.872	0.516(0.347–0.684)	0.661	0.316	0.825
	LR	0.781(0.698–0.863)	0.662	0.833	0.585	0.686(0.529–0.842)	0.627	0.842	0.525
Multi-sequence	SVM	0.944(0.910–0.979)	0.824	1.000	0.745	0.829(0.713–0.945)	0.763	0.842	0.725
	LR	0.934(0.764–0.975)	0.831	0.952	0.777	0.870(0.764–0.975)	0.864	0.737	0.925
Clinicoradiologic	SVM	0.915(0.863–0.967)	0.882	0.786	0.926	0.802(0.667–0.937)	0.814	0.684	0.875
	LR	0.905(0.854–0.957)	0.838	0.881	0.819	0.853(0.753–0.954)	0.831	0.632	0.925
Combined	SVM	0.979(0.958–1.000)	0.941	0.952	0.936	0.899(0.815–0.982)	0.881	0.737	0.950
	LR	0.964(0.935–0.993)	0.919	0.976	0.894	0.936(0.870–1.000)	0.864	0.947	0.825

SVM: support vector machine; LR: logistic regression; AUC: area under the curve; CI: confidence interval; UP: unenhanced T1-weighted image; CMP: enhanced T1-weighted image in corticomedullary phase; NP: enhanced T1-weighted image in nephrographic phase; T2: T2-weighted image; Acc: accuracy; Spe: specificity; Sen: sensitivity.

3.3. Performances of multi-sequence radiomics models for discrimination

The selected radiomics features of the multi-sequence models were shown in Table 4. The performances of the multi-sequence models were displayed in Table 3 and Fig. 3. The AUCs of multi-sequence radiomics models were higher than the AUCs of single-sequence radiomics models. In the training set, the AUC of SVM model was a bit higher than LR model (0.944 vs 0.934, $p = 0.215$, Delong's test), while in the test set, the AUC of LR model was higher than SVM model (0.870 vs 0.829, $p = 0.065$, Delong's test). The Delong's test showed no statistical differences between SVM models and LR models.

3.4. Performances of combined model for discrimination

In order to further improve the classification performances of the models, we combined the clinicoradiologic features with the rad-scores of multi-sequence radiomics features to build an optimal model. The performances of the combined models were displayed in Table 3 and Fig. 3. In the training and test set, the AUCs of the combined models were higher than the other models. The LR-combined model achieved the best performance, of which AUC is 0.936 in the test set. Its performance was significantly better than SVM-multi-sequence radiomics model (AUC = 0.829, p = 0.028, Delong's test), and its AUC was higher

Table 4
The selected radiomics features extracted from multi-sequence.

Features	Coefficients
wavelet_firstorder_wavelet-HHH-Mean_UP	0.100
boxmean_firstorder_Skewness_T2	−0.100
wavelet_glszm_wavelet-LHH-SmallAreaLowGrayLevelEmphasis_UP	0.087
discretegaussian_firstorder_Skewness_CMP	0.087
normalize_glszm_GrayLevelVariance_CMP	0.076
wavelet_ngtdm_wavelet-LHL-Contrast_UP	−0.071
boxsigmaimage_glszm_SmallAreaEmphasis_T2	−0.057
original_glcm_Idmn_CMP	0.057

UP: unenhanced T1-weighted image; CMP: enhanced T1-weighted image in corticomedullary phase; NP: enhanced T1-weighted image in nephrographic phase; T2: T2-weighted image; glcm: gray level cooccurrence matrix; glszm: gray level size zone matrix; ngtdm: neighbouring gray tone difference matrix.

than the SVM-combined model (AUC = 0.899, p = 0.234, Delong's test) with no statistically significant differences.

4. Discussion

In this study, we built 14 models for differentiating non-ccRCCs from benign renal tumors with the features of 4 categories (clinicoradiologic features, MRI radiomics features from single-sequence and multi-sequence, combined features) and 2 kinds of classifiers (SVM, LR). Our results suggested that most models had good performances. Compared with the single-sequence radiomics models, both the multi-sequence models and the combined models had performance improvement. The combined models could optimally distinguish between non-ccRCCs and renal benign tumors.

As an advanced computational technology, previous studies addressed radiomics could extracted quantitative features from images of tumors [27]. It could more comprehensively show the heterogeneity of tumors than morphological visual analysis and cast off the deficiencies of traditional imaging diagnosis methods [28]. Because of the excellent soft tissue resolution, MRI is commonly applied to tumor detection [29]. Among the four single-sequence models in our study, the result showed that CMP models achieved the highest AUCs, while T2 models achieved the lowest AUCs. Categories of all the radiomics features selected from the four sequences were first-order statistics and second-order statistics, which both reflected the heterogeneity of the mass. The best feature of CMP is discretegaussian_firstorder_Skewness, which showed the positive relation with non-ccRCCs. Skewness represents the degree of skewness in the distribution of voxel intensity values in an image. The higher the skewness is, the more unevenly the data distributes. Yuki Arita et al. [30] also reported that GLZLM_LZHGE, indicating high signal homogeneity, was the dominant feature for fpAMLs to be distinguished from non-ccRCCs. Therefore, we consider that image heterogeneity is an important characteristic of malignant tumors. As previous studies referred, pRCC and chRCC often show mild to moderate progressive enhancement on CMP image, while AMLwvf and RO often show early moderate to obvious enhancement [4,31,32]. RO often show high signal intensity on T2 image, while pRCC, chRCC often show low signal intensity [4]. Due to the abundant smooth muscle, AMLwvf often show low signal intensity on T2 image as well [4,33]. What's more, our multivariate logistic regression analysis result of the clinicoradiologic features also showed that mild enhancement was the independent risk factors of non-ccRCCs ($p = 0.001$). These evidences might explain why the radiomics models based on CMP performed better than T2 for discriminating non-ccRCCs from benign renal tumors，which was consistent with previous study [34,35].

Wang et al. [35] selected the significant radiomics features extracted from MRI to fit the multivariate logistic model for the discrimination of ccRCC, pRCC, chRCC and achieved excellent AUCs for T2, CMP, NP and the combined three sequences (0.631, 0.790, 0.959, and 0.959 between ccRCC and chRCC; 0.688, 0.854, 0.909, and 0.955 between pRCC and chRCC; 0.747, 0.810, 0.814, and 0.890 between ccRCC and pRCC). The results showed that the combination of three MRI sequences performed better than single sequence for discrimination. Our result also showed that the AUCs of the multi-sequence models were indeed improved compared with the four single-sequence models. We speculated that the combination of multi-sequences could contain more valuable tumor heterogeneity information. However, we also found that the selected features of multi-sequence models were extracted from three sequences (UP, CMP, T2) but not four. Yang et al. [36] fed fifteen combinations of four-phasic CT features into 224 classification models, which were built with 8 classifiers and 28 feature selection methods. They achieved the best AUC of 0.90 with the models of "SVM-t_score-UP" and "SVM-relief-UP+NP" on the differentiation of AMLwvf and RCC less than 4 cm. It suggested that multi-sequence combination did not mean the performance of model must be improved. Although our study suggested that "UP+CMP+T2" scanning might be sufficient for radiomics analysis for differentiating non-ccRCCs from benign renal tumors, standard MRI renal mass scanning protocol is recommended to avoid diagnosis error.

Some studies referred that the combination of imaging and clinical features was beneficial to improve the diagnosis performance [37–41]. Chong et al. [42] reported that the performances of the models built on the independent risk factors from clinic, image and rad-scores with LR and RF had been improved, which were compared with the vast models built on the single kind of features. In our study, the combined models indeed performed best because the high-dimensional information of radiomics compensated for the shortcomings of subjective diagnosis by radiologists. Michail E Klontzas et al. [43] reported that the combined SPECT/radiomics model achieved higher accuracy (95 %) with an AUC of 98.3 % than the radiomics-only model (71.67 %) with an AUC of 75 % and visual evaluation of 99mTc Sestamibi SPECT/CT alone (90.8 %) with an AUC of 90.8 % to differentiate ROs from RCCs. In addition, they also reported that the combined radiometabolomics model achieved an AUC of 86.4 %, whereas metabolomics-only and radiomics-only classifiers achieved AUC of 72.7 % and 68.2 % [44]. Consistent with our conclusion，radiomics alone did not sufficiently for tumor differentiation. Some clinical features, including proteomics and metabolomics, are worth exploring in combination with radiomics.

Our study had some innovations. We analyzed the traditional imaging and clinical features of non-ccRCCs and benign renal tumors in detail, which can provide doctors with more key points for diagnosis. Moreover, we extracted radiomics features of 3D tumors as full as possible, while some previous study analyzed the maximum cross-section of tumors, which would lose much information. Finally, we adopted the methods of multi-sequence, multi-classifier and multi-combination, while many previous studies only used single classifier or fewer sequences and combinations [5,45]. Comparing the 14 models constructed in this study, the most models built with LR performed better. Miskin et al. [46] implemented three CT texture-based classifiers (SVM, RF, LR) to distinguish benign from potentially malignant cystic renal masses and reported that LR was the best with the highest AUC value of 0.90. It was speculated that complex models such as SVM may require more samples [47]. Although we used multiple MRI sequences and classifiers to build models, other powerful AI techniques such as DP and radiomics features of other functional imaging deserve exploration in further study.

There were several limitations in our study. Firstly, our study was retrospectively implemented in single institution with a small sample size which might lead to selection bias and lack of external validation. Secondly, the data were relatively imbalanced. A larger sample size and multi-center data is needed in future research to validate the model. Thirdly, manual delineation of VOI can ensure accuracy, but the time-consuming problem cannot be avoided. In addition, it is subjective and may not guarantee complete consistency of VOIs delineated from each sequence. In the future, more accurate semi-automatic or fully automatic delineation technology is needed to improve efficiency and reproducibility.

5. Conclusions

The MRI-based radiomics machine learning model showed a favorable performance in differentiating non-ccRCCs from benign renal tumors. Before clinical practice, multi-center and prospective studies involving larger datasets should be performed to validate the model.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Zhongshan Hospital, Fudan University (approval number B2020–214R). All experiments were performed in accordance with relevant guidelines and regulations. The requirement for informed consent was waived by the Research Ethics Committee of Zhongshan Hospital, Fudan University.

Institutional Review Board Statement

This retrospective study was approved by our institutional ethics committees, with waiver of informed consent granted.

Funding

This work was supported by National Natural Science Foundation of China (82202285).

CRediT authorship contribution statement

Yuqin Ding: Writing – review & editing, Methodology. **Pingyi Zhu:** Investigation. **Xianpan Pan:** Methodology. **Lei Chen:** Software, Methodology. **Jianjun Zhou:** Writing – review & editing, Methodology. **Ruiting Wang:** Writing – original draft. **Lianting Zhong:** Resources, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejro.2024.100608.

References

1 S. Gurbani, D. Morgan, V. Jog, et al., Evaluation of radiomics and machine learning in identification of aggressive tumor features in renal cell carcinoma (RCC), Abdom. Radio. (NY) 46 (9) (2021) 4278–4288, https://doi.org/10.1007/s00261-021-03083-y.
2 C.C. Moreno, J. Hemingway, A.C. Johnson, D.R. Hughes, P.K. Mittal, R. Duszak Jr., Changing abdominal imaging utilization patterns: perspectives from medicare beneficiaries over two decades, J. Am. Coll. Radio. 13 (8) (2016) 894–903, https://doi.org/10.1016/j.jacr.2016.02.031.
3 G. Gandaglia, P. Ravi, F. Abdollah, et al., Contemporary incidence and mortality rates of kidney cancer in the United States, Can. Urol. Assoc. J. 8 (7-8) (2014) 247–252, https://doi.org/10.5489/cuaj.1760.
4 C. Lopes Vendrami, C. Parada Villavicencio, T.J. DeJulio, et al., Differentiation of Solid Renal Tumors with Multiparametric MR Imaging, Radiographics 37 (7) (2017) 2026–2042, https://doi.org/10.1148/rg.2017170039.
5 D. Said, S.J. Hectors, E. Wilck, et al., Characterization of solid renal neoplasms using MRI-based quantitative radiomics features, Abdom. Radio. (NY) 45 (9) (2020) 2840–2850, https://doi.org/10.1007/s00261-020-02540-4.
6 S. Durinck, E.W. Stawiski, A. Pavía-Jiménez, et al., Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes, Nat. Genet 47 (1) (2015) 13–21, https://doi.org/10.1038/ng.3146.
7 M. Gerlinger, A.J. Rowan, S. Horswell, et al., Intratumor heterogeneity and branched evolution revealed by multiregion sequencing, N. Engl. J. Med 366 (10) (2012) 883–892, https://doi.org/10.1056/NEJMoa1113205.
8 X. Li, P. Nie, J. Zhang, F. Hou, Q. Ma, J. Cui, Differential diagnosis of renal oncocytoma and chromophobe renal cell carcinoma using CT features: a central scar-matched retrospective study, Acta Radio. 63 (2) (2022) 253–260, https://doi.org/10.1177/0284185120988109.
9 Z. Han, Y. Zhu, J. Xu, et al., Predictive value of CT-based radiomics in distinguishing renal angiomyolipomas with minimal fat from other renal tumors, Dis. Markers 2022 (2022) 9108129, https://doi.org/10.1155/2022/9108129.
10 N. Takahashi, S. Leng, K. Kitajima, et al., Small (< 4 cm) renal masses: differentiation of angiomyolipoma without visible fat from renal cell carcinoma using unenhanced and contrast-enhanced CT, AJR Am. J. Roentgenol. 205 (6) (2015) 1194–1202, https://doi.org/10.2214/ajr.14.14183.
11 A.B. Rosenkrantz, N. Hindman, E.F. Fitzgerald, B.E. Niver, J. Melamed, J.S. Babb, MRI features of renal oncocytoma and chromophobe renal cell carcinoma. *AJR*, Am. J. Roentgenol. 195 (6) (2010) W421–W427, https://doi.org/10.2214/ajr.10.4718.
12 Y. Li, X. Huang, Y. Xia, L. Long, Value of radiomics in differential diagnosis of chromophobe renal cell carcinoma and renal oncocytoma, Abdom. Radio. (NY) 45 (10) (2020) 3193–3201, https://doi.org/10.1007/s00261-019-02269-9.
13 M. Nikpanah, Z. Xu, D. Jin, et al., A deep-learning based artificial intelligence (AI) approach for differentiation of clear cell renal cell carcinoma from oncocytoma on multi-phasic MRI, Clin. Imaging 77 (2021) 291–298, https://doi.org/10.1016/j.clinimag.2021.06.016.
14 J.R. Young, D. Margolis, S. Sauk, A.J. Pantuck, J. Sayre, S.S. Raman, Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT, Radiology 267 (2) (2013) 444–453, https://doi.org/10.1148/radiol.13112617.
15 A.K. Paschall, S.M. Mirmomen, R. Symons, et al., Differentiating papillary type I RCC from clear cell RCC and oncocytoma: application of whole-lesion volumetric ADC measurement, Abdom. Radio. (NY) 43 (9) (2018) 2424–2430, https://doi.org/10.1007/s00261-017-1453-4.
16 B. Kocak, E.A. Kus, A.H. Yardimci, C.T. Bektas, O. Kilickesmez, Machine Learning in Radiomic Renal Mass Characterization: Fundamentals, Applications, Challenges, and Future Directions. *AJR*, Am. J. Roentgenol. 215 (4) (2020) 920–928, https://doi.org/10.2214/ajr.19.22608.
17 M.W. Ball, S.M. Bezerra, M.A. Gorin, et al., Grade heterogeneity in small renal masses: potential implications for renal mass biopsy, J. Urol. 193 (1) (2015) 36–40, https://doi.org/10.1016/j.juro.2014.06.067.
18 Z. Liu, S. Wang, D. Dong, et al., The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges, Theranostics 9 (5) (2019) 1303–1322, https://doi.org/10.7150/thno.30309.
19 X.Y. Sun, Q.X. Feng, X. Xu, et al., Radiologic-radiomic machine learning models for differentiation of benign and malignant solid renal masses: comparison with expert-level radiologists. *AJR*, W44-w54, Am. J. Roentgenol. 214 (1) (2020), https://doi.org/10.2214/ajr.19.21617.
20 S. Matsumoto, Y. Arita, S. Yoshida, et al., Utility of radiomics features of diffusion-weighted magnetic resonance imaging for differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma: model development and external validation, Abdom. Radio. (NY) 47 (6) (2022) 2178–2186, https://doi.org/10.1007/s00261-022-03486-5.
21 Q. Xu, Q. Zhu, H. Liu, et al., Differentiating Benign from Malignant Renal Tumors Using T2- and Diffusion-Weighted Images: A Comparison of Deep Learning and Radiomics Models Versus Assessment from Radiologists, J. Magn. Reson Imaging 55 (4) (2022) 1251–1259, https://doi.org/10.1002/jmri.27900.
22 Y.M. Zheng, J. Li, S. Liu, et al., MRI-Based radiomics nomogram for differentiation of benign and malignant lesions of the parotid gland, Eur. Radio. 31 (6) (2021) 4042–4052, https://doi.org/10.1007/s00330-020-07483-4.
23 X. Zhao, Y. Zhou, Y. Zhang, et al., Radiomics based on contrast-enhanced MRI in differentiation between fat-poor angiomyolipoma and hepatocellular carcinoma in noncirrhotic liver: a multicenter analysis, Front Oncol. 11 (2021) 744756, https://doi.org/10.3389/fonc.2021.744756.
24 S. Mueller-Using, T. Feldt, F.S. Sarfo, K.A. Eberhardt, Factors associated with performing tuberculosis screening of HIV-positive patients in Ghana: LASSO-based predictor selection in a large public health data set, BMC Public Health 16 (2016) 563, https://doi.org/10.1186/s12889-016-3239-y.
25 L. Chen, Y. Shen, X. Huang, et al., MRI-Based Radiomics for Differentiating Orbital Cavernous Hemangioma and Orbital Schwannoma, Front Med (Lausanne) 8 (2021) 795038, https://doi.org/10.3389/fmed.2021.795038.
26 B. Kocak, T. Akinci D'Antonoli, N. Mercaldo, et al., METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII, Insights Imaging 15 (1) (2024) 8, https://doi.org/10.1186/s13244-023-01572-w.
27 B. Zhao, Y. Tan, W.Y. Tsai, et al., Reproducibility of radiomics for deciphering tumor phenotype with imaging, Sci. Rep. 6 (2016) 23428, https://doi.org/10.1038/srep23428.
28 R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: Images Are More than Pictures, They Are Data, Radiology 278 (2) (2016) 563–577, https://doi.org/10.1148/radiol.2015151169.
29 H. Wang, P. Nie, Y. Wang, et al., Radiomics nomogram for differentiating between benign and malignant soft-tissue masses of the extremities, J. Magn. Reson Imaging 51 (1) (2020) 155–163, https://doi.org/10.1002/jmri.26818.
30 Y. Arita, S. Yoshida, T.C. Kwee, et al., Diagnostic value of texture analysis of apparent diffusion coefficient maps for differentiating fat-poor angiomyolipoma from non-clear-cell renal cell carcinoma, Eur. J. Radio. 143 (2021) 109895, https://doi.org/10.1016/j.ejrad.2021.109895.
31 R.S. Lim, T.A. Flood, M.D.F. McInnes, L.T. Lavallee, N. Schieda, Renal angiomyolipoma without visible fat: Can we make the diagnosis using CT and MRI? Eur. Radio. 28 (2) (2018) 542–553, https://doi.org/10.1007/s00330-017-4988-4.
32 X.L. Li, L.X. Shi, Q.C. Du, W. Wang, L.W. Shao, Y.W. Wang, Magnetic resonance imaging features of minimal-fat angiomyolipoma and causes of preoperative

misdiagnosis, World J. Clin. Cases 8 (12) (2020) 2502–2509, https://doi.org/10.12998/wjcc.v8.i12.2502.
33 L. Jian, Y. Liu, Y. Xie, S. Jiang, M. Ye, H. Lin, MRI-Based Radiomics and Urine Creatinine for the Differentiation of Renal Angiomyolipoma With Minimal Fat From Renal Cell Carcinoma: A Preliminary Study, Front Oncol. 12 (2022) 876664, https://doi.org/10.3389/fonc.2022.876664.
34 U.N. Hoang, S. Mojdeh Mirmomen, O. Meirelles, et al., Assessment of multiphasic contrast-enhanced MR textures in differentiating small renal mass subtypes, Abdom. Radio. (NY) 43 (12) (2018) 3400–3409, https://doi.org/10.1007/s00261-018-1625-x.
35 W. Wang, K. Cao, S. Jin, X. Zhu, J. Ding, W. Peng, Differentiation of renal cell carcinoma subtypes through MRI-based radiomics analysis, Eur. Radio. 30 (10) (2020) 5738–5747, https://doi.org/10.1007/s00330-020-06896-5.
36 R. Yang, J. Wu, L. Sun, et al., Radiomics of small renal masses on multiphasic CT: accuracy of machine learning-based classification models for the differentiation of renal cell carcinoma and angiomyolipoma without visible fat, Eur. Radio. 30 (2) (2020) 1254–1263, https://doi.org/10.1007/s00330-019-06384-5.
37 W. Chen, T. Zhang, L. Xu, et al., Radiomics Analysis of Contrast-Enhanced CT for Hepatocellular Carcinoma Grading, Front Oncol. 11 (2021) 660509, https://doi.org/10.3389/fonc.2021.660509.
38 J. Peng, J. Zhang, Q. Zhang, Y. Xu, J. Zhou, L. Liu, A radiomics nomogram for preoperative prediction of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma, Diagn. Inter. Radio. 24 (3) (2018) 121–127, https://doi.org/10.5152/dir.2018.17467.
39 M. Wei, Y. Zhang, G. Bai, et al., T2-weighted MRI-based radiomics for discriminating between benign and borderline epithelial ovarian tumors: a multicenter study, Insights Imaging 13 (1) (2022) 130, https://doi.org/10.1186/s13244-022-01264-x.
40 T. Asi, M. Tuncali, M. Tuncel, et al., The role of Tc-99m MIBI scintigraphy in clinical T1 renal mass assessment: Does it have a real benefit? Urol. Oncol. 38 (12) (2020) 937.e11–937.e17, https://doi.org/10.1016/j.urolonc.2020.07.018.
41 R. Gao, J. Pang, P. Lin, et al., Identification of clear cell renal cell carcinoma subtypes by integrating radiomics and transcriptomics, Heliyon 10 (11) (2024) e31816, https://doi.org/10.1016/j.heliyon.2024.e31816.
42 H.H. Chong, L. Yang, R.F. Sheng, et al., Multi-scale and multi-parametric radiomics of gadoxetate disodium-enhanced MRI predicts microvascular invasion and outcome in patients with solitary hepatocellular carcinoma ≤ 5 cm, Eur. Radio. 31 (7) (2021) 4824–4838, https://doi.org/10.1007/s00330-020-07601-2.
43 M.E. Klontzas, E. Koltsakis, G. Kalarakis, et al., Machine Learning Integrating (99m) Tc Sestamibi SPECT/CT and Radiomics Data Achieves Optimal Characterization of Renal Oncocytic Tumors, Cancers (Basel) 15 (14) (2023), https://doi.org/10.3390/cancers15143553.
44 M.E. Klontzas, E. Koltsakis, G. Kalarakis, et al., A pilot radiometabolomics integration study for the characterization of renal oncocytic neoplasia, Sci. Rep. 13 (1) (2023) 12594, https://doi.org/10.1038/s41598-023-39809-9.
45 I.L. Xi, Y. Zhao, R. Wang, et al., Deep Learning to Distinguish Benign from Malignant Renal Lesions Based on Routine MR Imaging, Clin. Cancer Res 26 (8) (2020) 1944–1952, https://doi.org/10.1158/1078-0432.Ccr-19-0374.
46 N. Miskin, L. Qin, S.A. Matalon, et al., Stratification of cystic renal masses into benign and potentially malignant: applying machine learning to the bosniak classification, Abdom. Radio. (NY) 46 (1) (2021) 311–318, https://doi.org/10.1007/s00261-020-02629-w.
47 B. Mao, J. Ma, S. Duan, Y. Xia, Y. Tao, L. Zhang, Preoperative classification of primary and metastatic liver cancer via machine learning-based ultrasound radiomics, Eur. Radio. 31 (7) (2021) 4576–4586, https://doi.org/10.1007/s00330-020-07562-6.