



Fuhrman nuclear grade prediction of clear cell renal cell carcinoma: influence of volume of interest delineation strategies on machine learning-based dynamic enhanced CT radiomics analysis

Shiwei Luo¹ · Ruili Wei¹ · Songlin Lu² · Shengsheng Lai³ · Jialiang Wu⁴ · Zhe Wu¹ · Xinrui Pang¹ · Xinhua Wei¹ · Xinqing Jiang¹ · Xin Zhen² · Ruimeng Yang¹

Received: 28 June 2021 / Revised: 18 August 2021 / Accepted: 9 September 2021 / Published online: 12 October 2021

© European Society of Radiology 2021

Abstract

Objective To investigate the influence of different volume of interest (VOI) delineation strategies on machine learning-based predictive models for discrimination between low and high nuclear grade clear cell renal cell carcinoma (ccRCC) on dynamic contrast-enhanced CT.

Methods This study retrospectively collected 177 patients with pathologically proven ccRCC (124 low-grade; 53 high-grade). Tumor VOI was manually segmented, followed by artificially introducing uncertainties as: (i) contour-focused VOI, (ii) margin erosion of 2 or 4 mm, and (iii) margin dilation (2, 4, or 6 mm) inclusive of perirenal fat, peritumoral renal parenchyma, or both. Radiomics features were extracted from four-phase CT images (unenhanced phase (UP), corticomedullary phase (CMP), nephrographic phase (NP), excretory phase (EP)). Different combinations of four-phasic features for each VOI delineation strategy were used to build 176 classification models. The best VOI delineation strategy and superior CT phase were identified and the top-ranked features were analyzed.

Results Features extracted from UP and EP outperformed features from other single/combined phase(s). Shape features and first-order statistics features exhibited superior discrimination. The best performance (ACC 81%, SEN 67%, SPE 87%, AUC 0.87) was achieved with radiomics features extracted from UP and EP based on contour-focused VOI.

Conclusion Shape and first-order features extracted from UP + EP images are better feature representations. Contour-focused VOI erosion by 2 mm or dilation by 4 mm within peritumor renal parenchyma exerts limited impact on discriminative performance. It provides a reference for segmentation tolerance in radiomics-based modeling for ccRCC nuclear grading.

Key Points

- Lesion delineation uncertainties are tolerated within a VOI erosion range of 2 mm or dilation range of 4 mm within peritumor renal parenchyma for radiomics-based ccRCC nuclear grading.
- Radiomics features extracted from unenhanced phase and excretory phase are superior to other single/combined phase(s) at differentiating high vs low nuclear grade ccRCC.
- Shape features and first-order statistics features showed superior discriminative capability compared to texture features.

Keywords Carcinoma, renal cell · Tomography, X-ray computed · Pathology · Neoplasm grading · Machine learning

Abbreviations

ACC	Accuracy
AUC	Area under the curve

Shiwei Luo and Ruili Wei contributed equally.

✉ Xin Zhen
xinzhen@smu.edu.cn

✉ Ruumeng Yang
eyruimengyang@scut.edu.cn

¹ Department of Radiology, Guangzhou First People's Hospital, South China University of Technology, Guangzhou, Guangdong 510180, China

² School of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong 510515, China

³ School of Medical Equipment, Guangdong Food and Drug Vocational College, Guangzhou, Guangdong 510520, China

⁴ Department of Radiology, University of Hong Kong-Shenzhen Hospital, Shenzhen, Guangdong 518000, China

ccRCC	Clear cell renal cell carcinoma
CECT	Contrast-enhanced computed tomography
CMP	Corticomedullary phase
CT	Computed tomography
EP	Excretory phase
NP	Nephrographic phase
PACS	Picture archiving and communication system
RCC	Renal cell carcinoma
ROC	Receiver operating characteristic
SEN	Sensitivity
SMOTE	Synthetic minority oversampling technique
SPE	Specificity
UP	Unenhanced phase
VOI	Volume of interest

Introduction

Renal cell carcinoma (RCC) is the most common kidney malignancy and accounts for almost 90% of renal cancers [1–4]. Clear cell renal cell carcinoma (ccRCC) makes up about 70% of RCCs and has poorer prognosis than other subtypes [1–4]. Nuclear grade is established as an independent histological prognostic factor and is significant for the clinical management of ccRCC [5]. Numerous grading systems have been applied in the pathological nuclear grading of ccRCC. Among them, the Fuhrman nuclear grading system is most widely used [6]. There are 4 nuclear grades (1–4) based on nuclear size, irregularity, and nucleolar prominence [6]. Grades 1 and 2 indicate low-grade tumors with better prognosis while grades 3 and 4 indicate high-grade tumors with poor prognosis [6, 7].

Preoperative prediction of nuclear grade is crucial for personalized treatments like partial nephrectomy, radical nephrectomy, or active surveillance [8–11]. Percutaneous renal tumor biopsy is an invasive procedure for preoperative determination of nuclear grade [12]. However, it is insensitive in 2.5–22% of the cases [5, 13] and may cause complications. Thus, novel noninvasive methods are needed to overcome this drawback.

Radiomics modeling is a promising tool for guiding clinical decisions through quantitative evaluation of medical images [14–16]. Various studies have sought to quantify radiomics features for stratification of low and high Fuhrman nuclear grades [17, 18]. However, this is hampered by a lack of standardized RCC lesion segmentation [19]. Various tumor delineation strategies, like adhering to the visible lesion edge, shrinking the margin to a certain amount to account for partial volume effect or volume averaging, or dilating the margin to include peritumor perirenal fat, have been proposed [20–22]. The impact brought by using different definitions of tumor volume of interest (VOI) for nuclear grading is still unclear. Here, we comprehensively

investigated the potential influence of different tumor VOI delineation methods on radiomics-based models to discriminate low- from high-grade ccRCC using dynamic contrast-enhanced CT (CECT).

Materials and methods

Patients

The study was approved by our local institutional review board, and the requirement for informed patient consent was waived due to the study's retrospective nature. Data were collected through an electronic search of the picture archiving and communication system (PACS) from January 2011 to January 2019. Inclusion criteria were (1) pathologically proven ccRCC with defined Fuhrman grade and (2) preoperative examination with four-phase CECT scans. Exclusion criteria were (1) cases of purely cystic ccRCC, (2) ccRCC without Fuhrman grade, and (3) prominent CT artifacts. The study workflow is shown on Fig. 1.

Fuhrman stage and image acquisition

To ensure reproducibility of pathological diagnosis and reduce intra/inter-observer variability, the traditional 4-tier Fuhrman grading system was re-categorized into a simplified Fuhrman grading system with low grade (grades 1 and 2) and high grade (grades 3 and 4). Fuhrman grading was done by a specialized genitourinary pathologist (W. S. Ding) with 9 years of experience.

Preoperative CECT images were acquired on Toshiba Aquilion One, Siemens Somatom Definition, GE HiSpeed 16, and Philips Brilliance 64. Acquisition parameters were as follows: 120–140-kV tube voltage, automated tube current modulation, and varied milliampere-second settings. All patients were injected with nonionic intravenous contrast material (1 mL/kg body weight, maximum volume = 150 mL) through the antecubital vein using mechanical power injectors. All patients underwent preoperative four-phase CT scans—phase 1: unenhanced (UP); phase 2: postcontrast corticomedullary phase (CMP); phase 3: post-contrast nephrographic phase (NP); and phase 4: postcontrast excretory phase (EP).

Segmentation

All retrieved CT images were stored in anonymized DICOM format. ITK-SNAP software (<http://www.itksnap.org>) was used to delineate target 3D VOI on the CT slice of the tumor in phases 1–4 for tumor segmentation. First, a contour-focused lesion VOI was manually delineated on the NP and then applied to the other 3 phases with slight adjustments

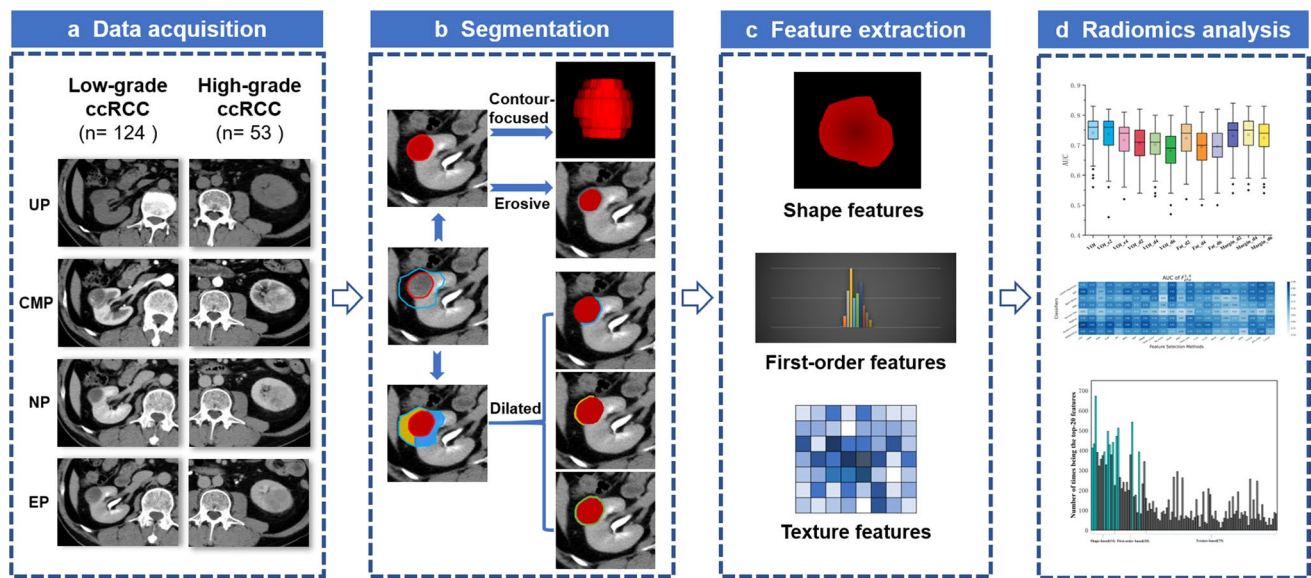


Fig. 1 Study workflow. UP, unenhanced phase; CMP, corticomedullary phase; NP, nephrographic phase; EP, excretory phase

tailoring VOIs to each phase. Next, a larger VOI containing perirenal fat and peritumoral renal parenchyma was generated by dilating the contour-focused original VOI by ~ 1 cm. This process was not entirely isotropic as the dilation would stop when encountering the bowel, liver, spleen, adrenal gland, vasculature, lymph nodes, adjacent visceral, or muscular tissue. Subtraction of the 2 VOIs yielded a loop VOI (VOI_{loop}), which was automatically divided into 2 parts of perirenal fat (VOI_{loop_fat}) and peritumoral renal parenchyma (VOI_{loop_margin}) using a predefined Hounsfield unit (HU) threshold of -20 . The VOI_{loop_fat} (< -20 HU) or the VOI_{loop_margin} (> -20 HU) was post-processed by removing isolated parts and filling small cavities. Manual segmentation was done by 2 investigators without prior knowledge of the lesions' pathology (S.W. Luo and R.L. Wei, with 4 and 5 years of experience in radiological diagnosis, respectively). Conformity of the delineated VOIs was measured using Dice similarity coefficient. For those CT slices with Dice indexes > 0.9 , the unanimous segmentation was the intersection of the two individual segmentations, while for those slices with Dice < 0.9 , discrepancies on lesion boundary were resolved by further discussions to reach consensus.

Different VOI delineation strategies

Based on the original contour-focused VOI and loop VOI, erosion and dilation procedures were done slice-by-slice on the aforementioned VOIs to simulate delineation uncertainties. The VOI was dilated by 2, 4, and 6 mm but still within the scope of VOI + VOI_{loop} (yielding the VOI_{d2}, VOI_{d4}, VOI_{d6}), VOI + VOI_{loop_fat} (yielding the

Fat_{d2}, Fat_{d4}, Fat_{d6}), or VOI + VOI_{loop_margin} (yielding the Margin_{d2}, Margin_{d4}, Margin_{d6}). Tumor VOI was eroded by 2 and 4 mm to yield VOI_{e2} and VOI_{e4}. Together with the original contour-focused VOI, 12 types of VOI delineations were obtained for subsequent radiomics analysis.

Feature extraction and representation

Texture feature extraction was done on each of the 12 VOIs in each phase using Pyradiomics [23]. The extracted features on each phase included 107 candidate features that can be categorized into 3 subtypes, including the shape, first-order statistics (histogram analysis), and second-order statistics ("texture features"). Features extracted on each of the 4 phases included F_{pha}^1 , F_{pha}^2 , F_{pha}^3 , and F_{pha}^4 (each with 107 features) and were termed group 1. Concatenation of features of any 2 phases were termed as group 2, including $F_{pha}^{1,2}$, $F_{pha}^{1,3}$, $F_{pha}^{1,4}$, $F_{pha}^{2,3}$, $F_{pha}^{2,4}$, and $F_{pha}^{3,4}$ (each with 214 features). Concatenation of features of any 3 phases were termed group 3, including $F_{pha}^{1,2,3}$, $F_{pha}^{1,2,4}$, $F_{pha}^{1,3,4}$, and $F_{pha}^{2,3,4}$ (each with 321 features). Concatenated features of all 4 phases were termed as group 4, i.e., $F_{pha}^{1,2,3,4}$ (with 428 features). Discriminative capabilities were respectively compared using the above 180 (12×15) types of features as input for a specific discrimination model.

Modeling and comparisons

We used 22 feature selection methods and 8 classification algorithms to build a total of 176 (22×8) discrimination

models. The above 180 types of features were fed into each of the 176 discriminative models, resulting in 31,680 (180×176) combinations for comparison. We evaluated each of these models with five fold cross-validation, in each of which an optimal subset of features (20, 40, 60, and 80 features for groups 1–4, respectively) was first estimated by a specific feature selection method. The prescreened features were then fed into a classifier for discrimination modeling. To ease data imbalance of the patient cohort, the synthetic minority oversampling technique (SMOTE) [24] was used to oversample the minority high-grade ccRCC group by introducing synthetic feature samples. The discrimination powers of the models were quantified using area under the receiver operating characteristic (ROC) curve (AUC), accuracy (ACC), sensitivity (SEN), and specificity (SPE).

Statistical analysis

Continuous variables are reported as mean \pm SD. Categorical variables are reported as numbers and proportions. Normality of the data distribution was assessed by the Kolmogorov–Smirnov test. Comparisons between groups were done

using the chi-square test for categorical variables, the independent *t*-test for normally distributed continuous variables, and the Mann–Whitney *U* test for non-normally distributed continuous variables. Discriminative comparisons between the 15 types of features were done using the independent samples Kruskal–Wallis test with Bonferroni correction to adjust significance level in pairwise comparisons. All statistical analyses were done on SPSS version 20 (IBM). Two-tailed $p < 0.05$ was considered statistically significant.

Results

Demographics

The study cohort comprised of 124 low-grade (17 were grade 1 (9.6%) and 107 were grade 2 (60.5%)) and 53 high-grade (40 were grade 3 (22.6%) and 13 were grade 4 (7.3%)) ccRCC patients who met the inclusion criteria (Table 1). The groups did not differ significantly with regard to age, sex, or lesion diameter ($p > 0.05$). Imaging and histological results from two representative patients are provided in Fig. 2.

Table 1 Demographics and characteristics of the study cohort

Characteristic	Low-grade ($n = 124$)	High-grade ($n = 53$)	<i>p</i> value
Patient age (mean \pm SD, year)	53.1 \pm 12.4	56.3 \pm 12.4	0.109 ^a
Sex			
Male ($n/\%$)	93 (75)	39 (73.6)	0.843 ^b
Female ($n/\%$)	31 (25)	14 (26.4)	
Lesion diameter (mean \pm SD, cm)	5.00 \pm 2.90	5.95 \pm 3.24	0.054 ^a

Data are presented as mean \pm SD or frequency (%).^a Independent *t*-test; ^b chi-square test. $p < 0.05$ was considered statistically significant

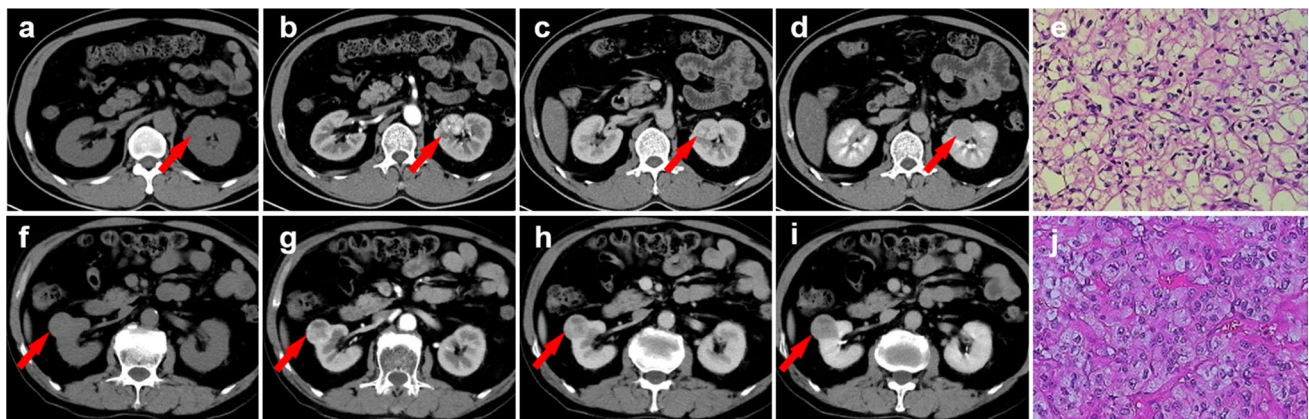


Fig. 2 Representative examples of clear cell renal cell carcinoma (ccRCC). **a–e** Low-grade (Fuhrman nuclear grade 2) ccRCC in a 49-year-old man. **a–d** Unenhanced phase (UP), corticomedullary phase (CMP), nephrographic phase (NP), and excretory phase (EP) CT images (red arrows point to the tumor in the left kidney). **e** Histo-

logic photomicrograph (hematoxylin–eosin, H & E stain). **f–j** High-grade (Fuhrman nuclear grade 3) ccRCC in a 69-year-old man. **f–i** UP, CMP, NP, and EP CT images (red arrows point to the tumor in the right kidney). **j** H & E stain

Discriminative capabilities of different feature types

The 12*15 feature types were compared by being fed to each of the 176 discrimination models. The running time of the established models ranged between 0.02 and approximately 90 s, with a mean time of ~8.9 s. Figure 3a shows the boxplot of the AUC distributions achieved by the 176 discriminative models for 15 phase-based feature types based on contour-focused VOI. Statistical comparisons revealed that F_{pha}^1 outperformed other single phases. The

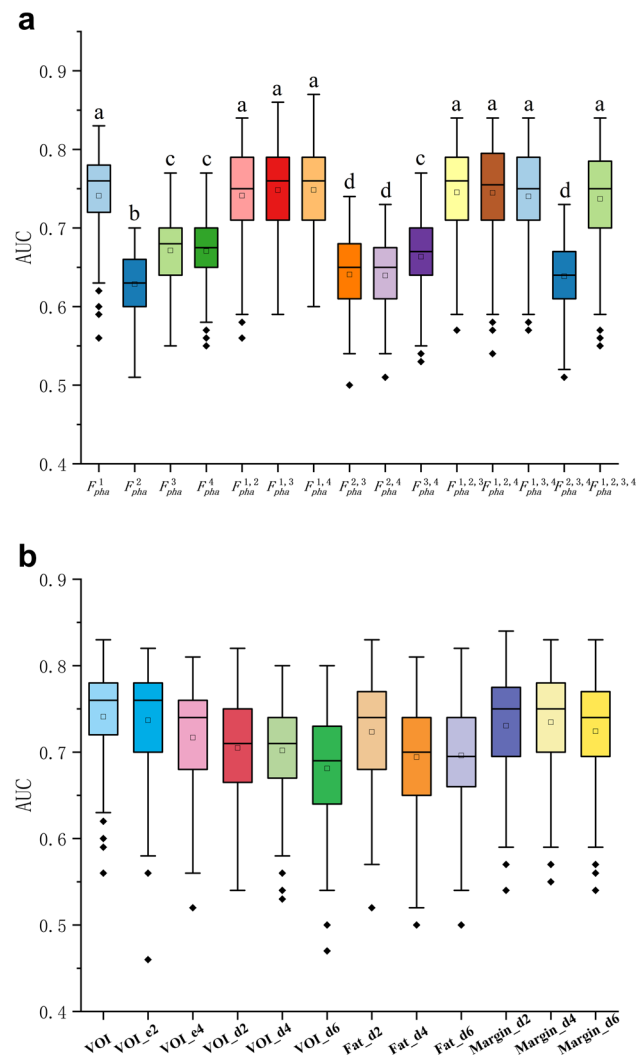


Fig. 3 Boxplots of the AUC distributions achieved by the 176 discriminative models, categorized by **a** 15 phases (features extracted on VOI) or **b** 12 VOIs (features extracted on F_{pha}^1). The boxes run from the 25th to 75th percentile; the two ends of the whiskers represent the 5% and 95% percentiles of the data; the horizontal line and the square in the box represent the median and mean values, respectively. Diamonds represent outliers. Letters above each box in **a** indicate statistical significance (Kruskal–Wallis test with the Bonferroni correction) between any two features types. No common letters indicate that the two feature types are significantly different

phase combinations including phase 1, e.g., $F_{pha}^{1,2}$, $F_{pha}^{1,3}$, $F_{pha}^{1,4}$, $F_{pha}^{1,2,3}$, $F_{pha}^{1,3,4}$, $F_{pha}^{1,2,4}$, and $F_{pha}^{1,2,3,4}$ showed superior performance than those without phase 1, e.g., $F_{pha}^{2,3}$, $F_{pha}^{2,4}$, $F_{pha}^{3,4}$, and $F_{pha}^{2,3,4}$ (Table 2). The highest AUC (0.87) was obtained using $F_{pha}^{1,4}$, with the discriminative models of combination of “Random Forest” and “CIFE”. Furthermore, $F_{pha}^{1,4}$ was the most frequent (36 times) feature type and was ranked as the best feature, followed by $F_{pha}^{1,2,4}$ (27 times) and $F_{pha}^{1,2,3,4}$ (27 times). Figure 3b shows the boxplot of the AUC distributions for 12 VOI delineations based on F_{pha}^1 . There were no significant differences between the original contour-focused VOI and VOI_e2, Margin_2, and Margin_4 (Table 3). Significant inferior performances were seen in other VOIs relative to the original VOI. A summary of the highest performance within each of the 12 VOI types, including the specific phase the feature extracted from, and the classifier and feature selection method used is shown on Table 3. The model built with “Random Forest” and “CIFE” based on $F_{pha}^{1,4}$ and VOI performed best (ACC 81%, SEN 67%, SPE 87%, AUC 0.87), followed by the model built with “Random Forest” and “MIFS” based on $F_{pha}^{1,4}$ and Margin_d2 (ACC 78%, SEN 61%, SPE 86%, AUC 0.87).

Key feature analysis

AUC values obtained by all 176 discriminative models using features from F_{pha}^1 and $F_{pha}^{1,4}$ as feature input were visualized on a heatmap (Fig. 4). The highest AUCs for F_{pha}^1 and $F_{pha}^{1,4}$ were 0.83 and 0.87, respectively. For all discriminative models with AUC > 0.80, we counted the number of times each feature in F_{pha}^1 had been selected in the top-20 features in the fivefold cross-validation (Fig. 5). The top-10 most frequently selected features in F_{pha}^1 are highlighted as blue in Fig. 5 (two features ranked 10th) and summarized in Table 4, including 7 shape features and 4 first-order statistics features. No texture features were included. Of the top-10 features, all 4 first-order statistical features (2 with $p < 10^{-3}$ and 2 with $p < 10^{-5}$) and 3 shape features (least axis length with $p = 0.0074$, surface volume ratio with $p = 0.0046$, elongation with $p = 0.0249$) were statistically significant features.

We estimated the capability of using the mean of the mean feature values of the 2 groups (i.e., “M” in Table 4) as threshold to differentiate the 2 groups. It was observed that the first-order statistical features, i.e., the median, the 90th percentile, the 10th percentile, and root mean squared, demonstrated good discriminative capabilities in which about 70% of the high-grade group had larger feature values and about 65% of the low-grade group had smaller feature values.

Table 2 The best performances achieved in 15 phases (features extracted on the contour-focused VOI)

	Feature type	Best performance			Number of times being ranked as the best feature type
		Highest AUC (ACC/SEN/SPE)	Classifier	Feature selection method	
Group 1	F_{pha}^1	0.83 (0.74/0.59/0.81)	Random Forest	CIFE	23
	F_{pha}^2	0.70 (0.75/0.49/0.85)	Random Forest	NDFS	0
	F_{pha}^3	0.77 (0.75/0.62/0.80)	AdaBoosting	CMIM	0
	F_{pha}^4	0.77 (0.75/0.62/0.80)	AdaBoosting	ICAP	0
Group 2	$F_{pha}^{1,2}$	0.84 (0.75/0.67/0.78)	Random Forest	MIFS	23
	$F_{pha}^{1,3}$	0.86 (0.80/0.68/0.86)	Random Forest	CIFE	7
	$F_{pha}^{1,4}$	0.87 (0.81/0.67/0.87)	Random Forest	CIFE	36
	$F_{pha}^{2,3}$	0.74 (0.78/0.56/0.87)	Random Forest	ls_l21	0
	$F_{pha}^{2,4}$	0.73 (0.72/0.43/0.84)	Random Forest	NDFS	1
	$F_{pha}^{3,4}$	0.77 (0.75/0.62/0.80)	AdaBoosting	CMIM	0
Group 3	$F_{pha}^{1,2,3}$	0.84 (0.79/0.67/0.84)	Random Forest	MRMR	8
	$F_{pha}^{1,2,4}$	0.84 (0.77/0.57/0.85)	Random Forest	ICAP	27
	$F_{pha}^{1,3,4}$	0.84 (0.76/0.63/0.81)	Random Forest	CMIM	24
	$F_{pha}^{2,3,4}$	0.73 (0.76/0.56/0.84)	SVM	MCFS	0
Group 4	$F_{pha}^{1,2,3,4}$	0.84 (0.79/0.59/0.87)	Random Forest	CMIM	27

Discussion

In this retrospective study, we explored the influence of different VOI delineation strategies on radiomics modeling for the discrimination of low and high nuclear grade ccRCC with dynamic CECT. Experimental results demonstrated that the $F_{pha}^{1,4}$ -based model achieved the best performance compared with other phasic combinations. The discrimination model based on “Random Forest” and “CIFE” yielded satisfactory performance (ACC 81%, SEN 67%, SPE 87%, AUC = 0.87) with radiomics features extracted from $F_{pha}^{1,4}$ based on tumor contour-focused VOI. Furthermore, VOI, VOI_e2, Margin_d2, and Margin_d4 exhibited similar performances and could act as references for tumor segmentation (manual or automatic) to minimize the influence of segmentation uncertainties on performance variations in nuclear grade stratifications of ccRCC.

Currently, radiomics analysis is used to help distinguish high-grade from low-grade ccRCC. However, most studies

have analyzed features from a single phase or incomplete contrast-enhanced phases. For example, Betkas et al. [17] evaluated the performance of 2D portal-phase CT texture features combined with different ML-based classification schemes in discriminating low and high nuclear grade ccRCCs. The best model was created using SVM with overall ACC, SEN, SPE, and AUC (for detecting high-grade ccRCCs) of 85.1%, 91.3%, 80.6%, and 0.860, respectively. Lin et al. [25] established machine learning models based on single- or three-phases (pre-contrast phase, corticomedullary phase and nephrographic phase) CT images to differentiate low- and high-grade ccRCC. The best diagnostic performance was observed when using all three-phase CT images (AUC = 0.87), followed by single-phase NP (AUC = 0.84), CMP (AUC = 0.80), and PCP (AUC = 0.82) images. However, some valuable information might have been missed since unenhanced phase or excretory phase was not considered in that study. In our study, EP (F_{pha}^4) exhibited good performance that was only slightly inferior to the UP (F_{pha}^1)

Table 3 Best performances in each of the 12 VOIs (achieved with the associated phases, classifier, and feature selection), and the number of times they ranked as best feature. $p < 0.05$ was considered statistically significant (Mann–Whitney U test)

	Best feature type	Best performance			p value	Number of times being ranked as the best feature type
		Highest AUC (ACC/SEN/SPE)	Classifier	Feature selection method		
VOI	$F_{pha}^{1,4}$	0.87 (0.81/0.67/0.87)	Random Forest	CIFE		36
VOI_e2	$F_{pha}^{1,4}$	0.85 (0.79/0.66/0.84)	Random Forest	Gini_index	0.8243	37
VOI_e4	$F_{pha}^{1,3}$	0.84 (0.80/0.77/0.81)	Logistic Regression	RFS	$< 10^{-5}$	12
VOI_d2	$F_{pha}^{1,2}$	0.84 (0.77/0.65/0.82)	AdaBoosting	Gini_index	$< 10^{-9}$	27
VOI_d4	$F_{pha}^{1,2,3,4}$	0.83 (0.79/0.64/0.85)	Bagging	CIFE	$< 10^{-12}$	35
VOI_d6	$F_{pha}^{1,2,4}$	0.84 (0.79/0.57/0.89)	Bagging	T_score	$< 10^{-20}$	24
Fat_d2	$F_{pha}^{1,2}$	0.86 (0.77/0.63/0.84)	Random Forest	MIFS	0.0029	21
Fat_d4	$F_{pha}^{1,3}$	0.84 (0.80/0.65/0.86)	Random Forest	MRMR	$< 10^{-13}$	13
Fat_d6	$F_{pha}^{1,2,4}$	0.86 (0.79/0.60/0.87)	AdaBoosting	LI_l21	$< 10^{-12}$	21
Margin_d2	$F_{pha}^{1,3}$	0.86 (0.82/0.68/0.88)	Random Forest	CIFE	0.0861	14
Margin_d4	$F_{pha}^{1,4}$	0.87 (0.78/0.61/0.86)	Random Forest	MIFS	0.2132	33
Margin_d6	$F_{pha}^{1,4}$	0.84 (0.77/0.57/0.86)	Random Forest	CIFE	0.0013	18

The Mann–Whitney U test is used for all p values. A p value less than 0.05 was considered statistically significant

among single phases, and their combination into $F_{pha}^{1,4}$ achieved the best performance over all phasic combinations. The findings by Kocak et al. [18] on the role of unenhanced CT in differentiating nuclear grade are consistent with ours. The possible explanation might be that the Fuhrman nuclear grade (nuclear size, irregularity, and nucleolar prominence) correlated more with heterogeneity of the tumor itself rather than tumor perfusion or vascularity.

Lesion segmentation is a critical procedure that might have substantial impact on the performance of radiomics analysis. [19] Kocak et al. [26] used 47 cases to assess the influence of segmentation margin on radiomics performance and found that contour-focused segmentation (AUC = 0.865–0.984) performed better than models with a lesion margin shrinkage of 2 mm (AUC: 0.745–0.887, $p < 0.05$). The finding was inspiring but this analysis was conducted only on the CMP. Gill et al. [22] assessed if juxtatumoral perinephric fat (JPF) may aid in machine learning-based nuclear grading of RCC. The CT-based texture analysis of ccRCC showed statistically significant differences in JPF adjacent to high- versus low-grade ccRCC. Here, we carried out a more thorough investigation to

determine the impact of different VOI delineation strategies on pathological nuclear grading, including dilation and erosion of contour-focused VOI. For a better understanding of the role of the peritumor components in nuclear grading, we divided peritumor components into peritumor renal parenchyma and peritumor perirenal fat. Our results showed no significant differences between the original VOI and the 2-mm erosive VOI, as well as the 2- or 4-mm margin dilated VOI. In particular, the 2-mm erosive VOI exhibited similar performance with the contour-focused VOI, which differs from past findings [26]. The peritumor renal parenchyma did not interfere with the discriminative performance within the range of 4-mm extension from original VOI, probably because higher grade ccRCCs tend to invade adjacent renal parenchyma and tumor-renal parenchyma interface within a certain range may reflect the biological behavior of tumors. However, relative to original VOI, VOI delineations inclusive of the perirenal fat showed inferior performance, suggesting that peritumor perirenal fat might not provide additional information compared to the tumor itself and conversely may weaken a model's differentiation capability. These findings may serve as references for determining RCC lesion segmentation tolerance for radiomics analysis.

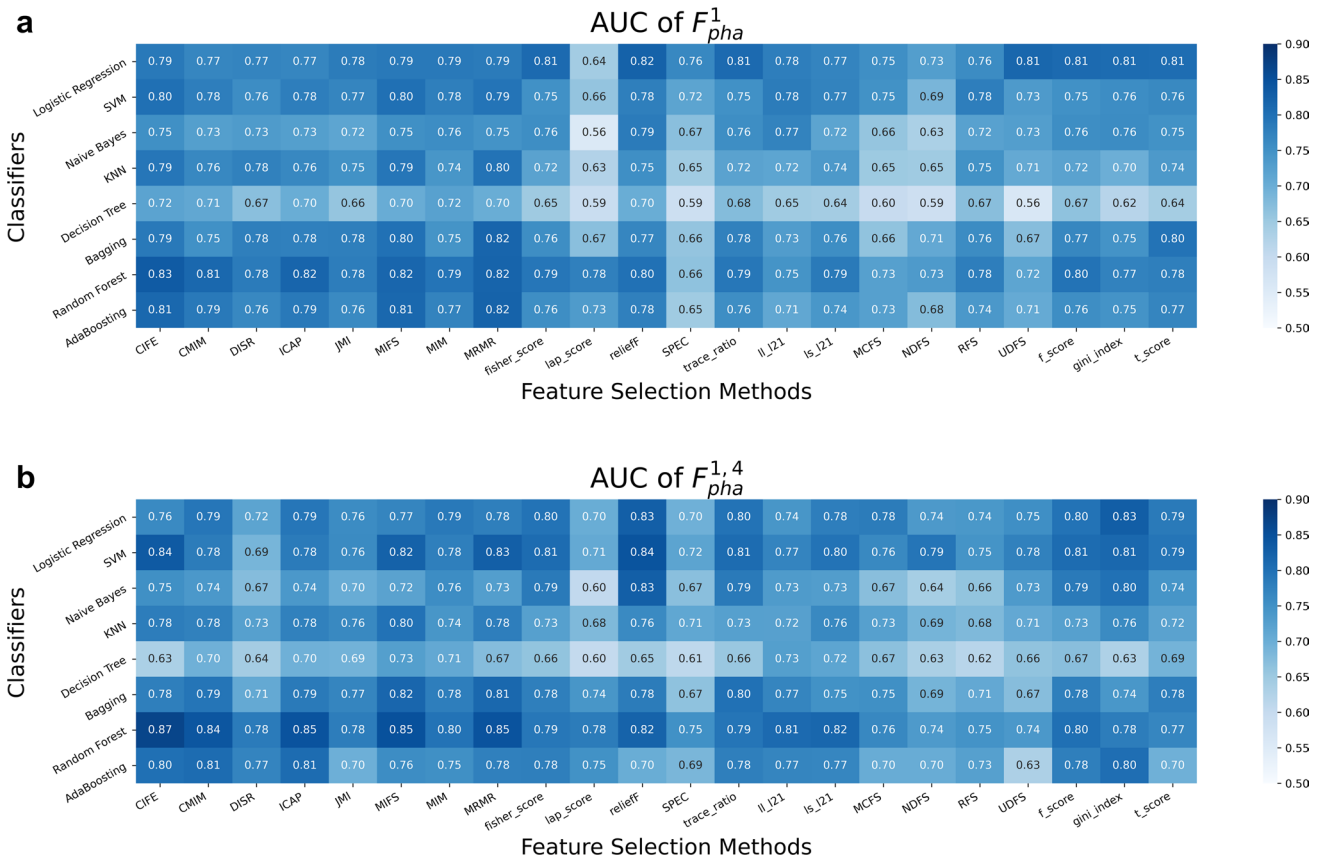


Fig. 4 Heatmap of the AUC values obtained by the 176 discriminative models (a, F^1_{pha} as feature input; b, $F^{1,4}_{pha}$ as feature input) built with different combinations of classifiers and feature selection methods (features extracted on VOI)

Fig. 5 Histogram showing the number of times of each feature in F^1_{pha} being selected as the top 20 features (features extracted on VOI)

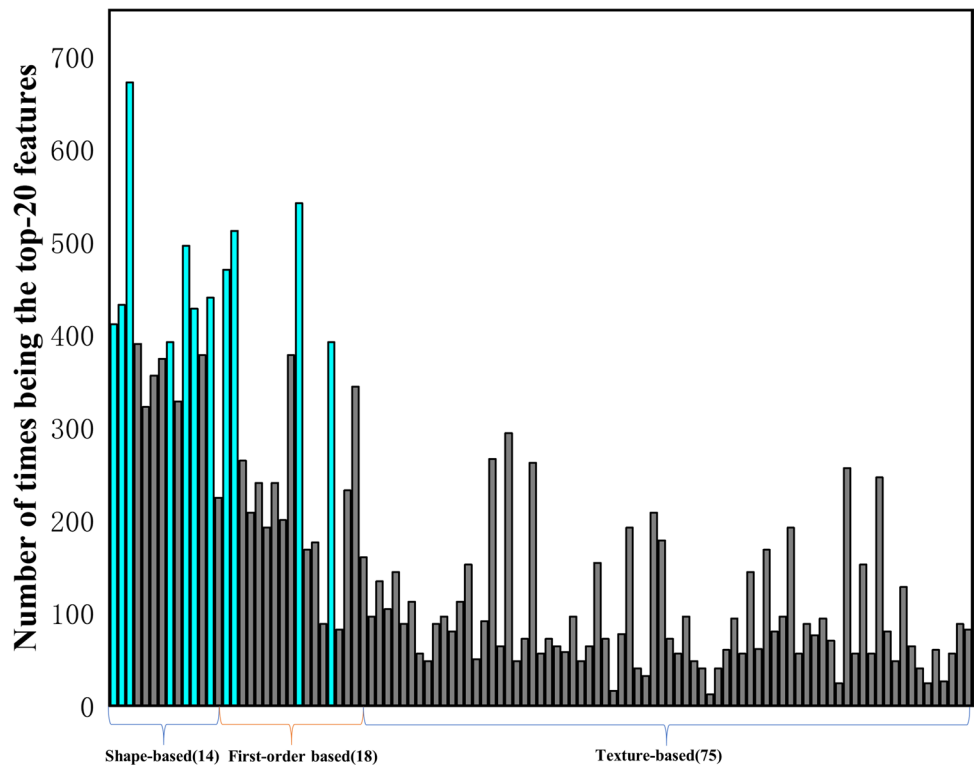


Table 4 The top-10 most frequently selected features in F_{pha}^1 . Their corresponding mean (\pm SD) values (or median (IQR)) and p values between low and high Fuhrman grades are shown

Feature category	Top-ranked features	High grade	Low grade	p value	$M^{\dagger 4}$	(< M > M) ^{†5}
Shape-based ($n=7$)	Least axis length (1st)	24.37 (2.96, 78.71)	16.18 (1.94, 35.87)	0.0074 ^b	20.27	High (49.1% 50.9%) Low (68.5% 31.5%)
	Minor axis length (4th)	55.52 (16.02, 152.18)	54.03 (15.03, 152.92)	0.4411 ^b	54.78	High (60.4% 39.6%) Low (58.9% 41.1%)
	Surface volume ratio (6th)	0.30 (0.09, 0.98)	0.35 (0.15, 1.24)	0.0046 ^b	0.33	High (69.8% 30.2%) Low (57.3% 42.7%)
	Flatness (7th)	0.45 (0.07, 0.90)	0.39 (0.05, 0.91)	0.5354 ^b	0.42	High (45.3% 54.7%) Low (57.3% 42.7%)
	Sphericity (8th)	0.59 (0.29, 0.79)	0.56 (0.26, 0.81)	0.8489 ^b	0.58	High (45.3% 54.7%) Low (57.3% 42.7%)
	Elongation (9th)	0.82 (0.60, 0.98)	0.85 (0.60, 0.99)	0.0249 ^b	0.84	High (56.6% 43.4%) Low (34.7% 65.3%)
	Maximum 3D diameter (10th)	81.21 (25.00, 229.85)	76.34 (23.62, 242.09)	0.2755 ^b	78.78	High (60.4% 39.6%) Low (63.7% 36.3%)
First-order-based ($n=4$)	Median (2nd)	34.11 \pm 3.72	28.47 \pm 6.52	< 10 ⁻⁵ . ^a	31.29	High (22.6% 77.4%) Low (67.0% 33.0%)
	90th percentile (3rd)	48.32 (41.00, 62.00)	42.49 (24.00, 61.00)	< 10 ⁻³ . ^b	45.41	High (32.1% 67.9%) Low (69.4% 30.6%)
	10th percentile (5th)	15.72 (2.00, 29.00)	10.30 (-19.00, 29.00)	< 10 ⁻³ . ^b	13.01	High (30.2% 69.8%) Low (58.1% 41.9%)
	Root mean squared (10th)	35.79 \pm 3.47	30.91 \pm 5.89	< 10 ⁻⁵ . ^a	33.36	High (20.7% 79.2%) Low (68.5% 31.5%)

^{†4}M, the mean of the mean feature value of the high-grade group and the mean feature value of the low-grade group

^{†5}M, percentage of patients in high-grade and low-grade groups with feature values less than or larger than the “M” value

^aIndependent t -test

^bMann-Whitney U test

The values in italics indicate satisfactory discriminative capability such that about 70% of the high group had smaller feature values and comparatively, about 70% of the low-grade group had larger feature values

It is interesting to find that the top-10 features were shape-based features ($n=7$) and first-order features ($n=4$) and that no texture features were significant. Of the 7 shape-based features, the least axis length, the surface volume ratio, and the elongation exhibited statistically significant differences ($p < 0.05$) between the 2 groups. The low-grade group had lower values of least axis length but higher values of surface volume ratio and elongation relative to the high-grade group, indicating that low-grade tumors are more inclined to be round. Yap et al.[27] compared the relative contributions of shape and texture metrics in differentiating benign from malignant renal masses and found that shape metrics alone have high prediction performance (AUC 0.64–0.68). Our data also reveal a positive role of shape features in ccRCC nuclear grading. The four top first-order features (median, 90th percentile, 10th percentile, and root mean squared) showed significantly higher values in the high-grade group ($p < 10^{-3}$). This is attributable to larger nuclear sizes and more prominent nucleolar appearances in high-grade tumors.

This study is limited by the relatively small sample size and lack of an external validation set. This is

attributable to our strict inclusion criteria that required that all 4 phases of contrast-enhanced CT to be available for all patients. However, we have to admit that radiomics applications continue to suffer from poor external validation due to inter-institutional variations in CT protocolling and workflow—both of which have shown strong implications in affecting the overall generalizability. In our study, we employed patient data from four different CT scanners, and such data heterogeneity, to some extent, has compensated for the limitation of no external validation. Another limitation is the use of Fuhrman nuclear grading system instead of the latest WHO/ISUP grading system [28–33]. This is because the included cases date back from 2011 when the Fuhrman grading system was widely used. Furthermore, the use of manual segmentation is time-consuming and automatic segmentation is necessary.

In conclusion, machine learning–based radiomics analysis on UP and NP outperformed other phases. There were no statistically significant performance differences between tumor VOI and VOI eroded by 2 mm or dilated by 2 or 4 mm within peritumor renal parenchyma, which may act

as reference for manual or automatic segmentation tolerance for radiomics-based modeling of ccRCC nuclear grading.

Funding This study has received funding from the National Natural Science Foundation of China [grant numbers 81971574, 81874216], the Natural Science Foundation of Guangdong Province, P.R. China [grant number 2021A1515011350], the Science and Technology Project of Guangzhou, P.R. China [grant numbers 201904010422, 202002030268, 202102010025], the Guangzhou Key Laboratory of Molecular Imaging and Clinical Translational Medicine and the Construction of High-level Key Clinical Specialty (Medical Imaging) in Guangzhou.

Declarations

Guarantor The scientific guarantor of this publication is Dr. Ruimeng Yang.

Conflict of interest The authors of this manuscript 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.

Methodology

- retrospective
- observational
- multicenter study

References

1. Siegel RL, Miller KD, Jemal A (2020) Cancer statistics, 2020. *CA Cancer J Clin* 70:7–30
2. Capitanio U, Bensalah K, Bex A et al (2019) Epidemiology of renal cell carcinoma. *Eur Urol* 75:74–84
3. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F (2015) International variations and trends in renal cell carcinoma incidence and mortality. *Eur Urol* 67:519–530
4. Leibovich BC, Lohse CM, Crispen PL et al (2010) Histological subtype is an independent predictor of outcome for patients with renal cell carcinoma. *J Urol* 183:1309–1315
5. Ljungberg B, Albiges L, Abu-Ghanem Y et al (2019) European Association of Urology Guidelines on Renal Cell Carcinoma: the 2019 update. *Eur Urol* 75:799–810
6. Fuhrman SA, Lasky LC, Limas C (1982) Prognostic significance of morphologic parameters in renal cell carcinoma. *Am J Surg Pathol* 6:655–663
7. Delahunt B (2009) Advances and controversies in grading and staging of renal cell carcinoma. *Mod Pathol* 22(Suppl 2):S24–36
8. Jewett MA, Mattar K, Basiuk J et al (2011) Active surveillance of small renal masses: progression patterns of early stage kidney cancer. *Eur Urol* 60:39–44
9. Kunkle DA, Egleston BL, Uzzo RG (2008) Excise, ablate or observe: the small renal mass dilemma—a meta-analysis and review. *J Urol* 179:1227–1233
10. Zini L, Perrotte P, Capitanio U et al (2009) Radical versus partial nephrectomy: effect on overall and noncancer mortality. *Cancer* 115:1465–1471
11. Campbell N, Rosenkrantz AB, Pedrosa I (2014) MRI phenotype in renal cancer: is it clinically relevant? *Top Magn Reson Imaging* 23:95–115
12. Marconi L, Dabestani S, Lam TB et al (2016) Systematic review and meta-analysis of diagnostic accuracy of percutaneous renal tumour biopsy. *Eur Urol* 69:660–673
13. Kutikov A, Smaldone MC, Uzzo RG, Haifler M, Bratslavsky G, Leibovich BC (2016) Renal mass biopsy: always, sometimes, or never? *Eur Urol* 70:403–406
14. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278:563–577
15. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762
16. Erichson BJ, Korfiatis P, Akkus Z, Kline TL (2017) Machine learning for medical imaging. *Radiographics* 37:505–515
17. Bektas CT, Kocak B, Yardimci AH et al (2019) Clear cell renal cell carcinoma: machine learning-based quantitative computed tomography texture analysis for prediction of fuhrman nuclear grade. *Eur Radiol* 29:1153–1163
18. Kocak B, Durmaz ES, Ates E, Kaya OK, Kilickesmez O (2019) Unenhanced CT texture analysis of clear cell renal cell carcinomas: a machine learning-based study for predicting histopathologic nuclear grade. *AJR Am J Roentgenol* 212:W132–W139
19. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B (2020) Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights Imaging* 11:91
20. He X, Wei Y, Zhang H et al (2020) Grading of clear cell renal cell carcinomas by using machine learning based on artificial neural networks and radiomic signatures extracted from multidetector computed tomography images. *Acad Radiol* 27:157–168
21. Ding J, Xing Z, Jiang Z et al (2018) CT-based radiomic model predicts high grade of clear cell renal cell carcinoma. *Eur J Radiol* 103:51–56
22. Gill TS, Varghese BA, Hwang DH et al (2019) Juxtatumoral perinephric fat analysis in clear cell renal cell carcinoma. *Abdom Radiol (NY)* 44:1470–1480
23. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107
24. Chawla N, Bowyer K, Hall L, Kegelmeyer W (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357
25. Lin F, Cui EM, Lei Y, Luo LP (2019) CT-based machine learning model to predict the Fuhrman nuclear grade of clear cell renal cell carcinoma. *Abdom Radiol (NY)* 44:2528–2534
26. Kocak B, Ates E, Durmaz ES, Uluhan MB, Kilickesmez O (2019) Influence of segmentation margin on machine learning-based high-dimensional quantitative CT texture analysis: a reproducibility study on renal clear cell carcinomas. *Eur Radiol* 29:4765–4775
27. Yap FY, Varghese BA, Cen SY et al (2021) Shape and texture-based radiomics signature on CT effectively discriminates benign from malignant renal masses. *Eur Radiol* 31:1011–1021
28. Humphrey PA (2014) Grading renal cell carcinoma: the International Society of Urological Pathology grading system. *J Urol* 191:798–799

29. Delahunt B, Cheville JC, Martignoni G et al (2013) The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. *Am J Surg Pathol* 37:1490–1504
30. Shu J, Wen D, Xi Y et al (2019) Clear cell renal cell carcinoma: machine learning-based computed tomography radiomics analysis for the prediction of WHO/ISUP grade. *Eur J Radiol* 121:108738
31. Dagher J, Delahunt B, Rioux-Leclercq N et al (2017) Clear cell renal cell carcinoma: validation of World Health Organization/International Society of Urological Pathology grading. *Histopathology* 71:918–925
32. Zhou H, Mao H, Dong D et al (2020) Development and external validation of radiomics approach for nuclear grading in clear cell renal cell carcinoma. *Ann Surg Oncol* 27:4057–4065
33. Cui E, Li Z, Ma C et al (2020) Predicting the ISUP grade of clear cell renal cell carcinoma with multiparametric MR and multiphase CT radiomics. *Eur Radiol* 30:2912–2921

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.